SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(107th Edition)

NOVEMBER 2021

BY: CENTRAL LIBRARY

DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY *DELHI COLLEGE OF ENGINEERING)*
GOVT. OF N.C.T. OF DELHI
SHAHBAD DAULATPUR, MAIN BAWANA ROAD
DELHI 110042

PREFACE

This is the **One Hundred Seventh** Issue of Current Awareness Bulletin started by Delhi Technological University, Central Library. The aim of the bulletin is to compile, preserve and disseminate information published by the faculty, students and alumni for mutual benefits. The bulletin also aims to propagate the intellectual contribution of Delhi Technological University (DTU) as a whole to the academia.

The bulletin contains information resources available in the internet in the form of articles, reports, presentations published in international journals, websites, etc. by the faculty and students of DTU. The publications of faculty and student which are not covered in this bulletin may be because of the reason that the full text either was not accessible or could not be searched by the search engine used by the library for this purpose.

The learned faculty and students are requested to provide their uncovered publications to the library either through email or in CD, etc. to make the bulletin more comprehensive.

This issue contains the information published during **November, 2021**. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. A new intelligent multilayer framework for insider threat detection, Mohammed Nasser Al-Mhiqani, Rabiah Ahmad, Z. Zainal Abidin, Karrar Hameed Abdul kareem, Mazin Abed Mohammed, ***7.1.Deepak Gupta***, K. Shankar, CSE, DTU

2. An analysis of customer's purchase intention towards the green products and its effect on manufacturing: A statistical approach, ***3.Pravin Kumar***, Manoj Kumar Singh and ***3.Saurabh Agrawal***, Mechanical and DSM, DTU

3. An assessment of green practices in Small and Medium-sized Enterprises (SMEs) using Graph- Theoretic Matrix Approach, Manoj Kumar Singh, ***3.Pravin Kumar*** and ***3.Saurabh Agrawal***, Mechanical and DSM, DTU

4. An intelligent model based on integrated inverse document frequency and multinomial Naive Bayes for current affairs news categorization, Sachin Kumar, Aditya Sharma, B Kartheek Reddy, Shreyas Sachan, Vaibhav Jain and ***3.Jagvinder Singh***, DSM, DTU

5. Assessment and modeling of sewer network development utilizing Arc GIS and SewerGEMS in Kabul city of Afghanistan, , ***6.Ali Reza Noori*** and ***S.K.Singh***, Environmental and Civil, DTU

6. Automatic Recognition of License Plates, ***7.Tanushri Bhagat*** and ***3.Rahul Thakur***, Electronics, DTU

7. BNDNN: Batch Normalization based Deep Neural Network for predicting flood in urban areas, ***6.Vinay Dubey*** and ***3.Rahul Katarya***, CSE, DTU

8. Carbon Nanotubes in Emerging Photovoltaics: Progress and Limitations, Sakshi Sharma and ***3.Vinod Singh***, Applied Physics, DTU

9. Causality Constraint on Circuit Complexity from COSMOEFT, Sayantan Choudhury, Arghya Mukherjee, ***8.Nilesh Pandey*** and Abhishek Roy, Applied Physics, DTU

10. Comparison of Coronary Heart Disease Prediction models using various Machine Learning Algorithms, ***6.Sunil Kr. Tiwari*** & ***3.Suresh Kumar Garg***, Mechanical, DTU

11. Constraints on holographic dark energy model with matter creation in Brans–Dicke theory and thermodynamic analysis, ***6.Simran Kaur*** and ***3.C.P. Singh***, Applied Mathematics, DTU

12. CVD Diamond Coated Tungsten Carbide (WC) Tool Inserts, ***6.Ramakant Rana***, Lucky Krishnia, R.S. Walia and ***3.Qasim Murtaza***, Mechanical, DTU

13. Digitally Controlled Power Management Circuit with Dual-functioned Single-stage Power Converter for Vibration Energy Harvesting, Hyunjun Jung, ***8.1.Yamini Sharma***, and Lei Zuo, Electronics, DTU

14. Does Gender Diversity on Board Promote Corporate Social Responsibility? An Empirical Analysis of Sustainable Development Goals, Amit Kumar Singh, Hima Bindu Kota, Varda Sardana and ***6.Shubham Singhania***, USME, DTU

15. Effect of asymmetries of stiffness and mass on parametric instabilities of hollow rotor system, Sandeep Chauhan, Chandan Kumar, Ashish Gupta and ***3.Vikas Rastogi***, Mechanical, DTU

16. Effect of critical success factors (CSFs) on the organizational performance, ***3.R. S. Mishra*** and ***6.Rakesh Kumar***, Mechanical, DTU

17. Effect of Process Parameter on Tensile Strength and Hardness of SS304 Processed by Microwave Radiation, ***6.Uma Gautam*** and ***3.Vipin***, Mechanical, DTU

18. Temperature dependency and linearity assessment of dual-metal gate stack junctionless accumulation-mode cylindrical surrounding gate (DMGS-JAM-CSG) MOSFET, ***6.Sumedha Gupta***, ***3.Neeta Pandey*** and R S Gupta, Electronics, DTU

19. Evaluation of tribological, oxidation and corrosion behavior of HVOF sprayed sustainable temperature-dependent carbon coating, ***6.Ankit Tyagi***, ***6.Shubhangi Chourasia***, ***3.Qasim Murtaza*** and RS Walia, Mechanical, DTU

20. Exergy and energy analysis of sensible heat storage based double pass hybrid solar air heater, Om Prakash, ***3.Anil Kumar***, ***3.Samsher***, Kumaresh Dey and Ankish Aman, Mechanical and Environment, DTU

21. EXPLORING THE RELATIONSHIP BETWEEN ETHICS AND KNOWLEDGE CULTURE: A CONCEPTUAL FRAMEWORK FOR SUCCESSFUL ORGANIZATIONS, ***6.Nishant Gaur*** and ***3.Vikas Gupta***, DSM, DTU

22. Fractional Oscillator - Harmonic Oscillator with Memory Effects, Vishwamittar, ***6.Yashika Taneja*** and ***7.Nipun Ahuja***, Mechanical, DTU

23. Fuzzy k-plane clustering method with local spatial information for segmentation of human brain MRI image, Puneet Kumar, ***3.Dhirendra Kumar*** and Ramesh Kumar Agrawal, Applied Mathematics, DTU

24. Highly Compact Triple Ring Slotted Circular Polarized Implantable Antenna for Bio-medical Applications, Rajiv Kumar Nehra and ***3.N S Raghava***, Electronics, DTU

25. Influence of strongly coupled plasma environment on photoionization of H-like O7þ ion, Dishu Dawra, Mayank Dimri, A. K. Singh, Alok K. S. Jha, Rakesh Kumar Pandey, ***3.Rinku Sharma*** and Man Mohan, Applied Physics, DTU

26. Influence of suspension parameter for derailment analysis of a full railway vehicle model cruising on a curved track, ***6.Yamika Patel***, ***3.Vikas Rastogi*** and ***3.Wolfgang Borutzky***, Mechanical, DTU

27. Is Dynamic Rumor Detection on social media Viable? An Unsupervised Perspective, ***7.Chahat Raj*** and ***3.Priyanka Meel***, IT, DTU

28. Junctionless Accumulation Mode Ferroelectric FET (JAM-FE-FET) for high frequency Digital and Analog applications, ***6.Snehlata Yadav***, ***3.Sonam Rewari*** and ***3.Rajeshwari Pandey***, Electronics, DTU

29. Landslide hazard assessment using analytic hierarchy process (AHP): A case study of National Highway 5 in India, ***6.Sandeep Panchal*** and ***3.Amit Kr. Shrivastava***, DTU

30. Leaky-momentum control algorithm for voltage and frequency control of three-phase SEIG feeding isolated load, Duli Chand Meena, Madhusudan Singh and ***6.Ashutosh K. Giri***, Electrical, DTU

31. Low Power Pre-charge Voltage Level and Low Swing Logic Based 8T SRAM Cell for High Speed CMOS Circuits, ***3.R. M. Rawat*** and ***3.V. Kumar***, CSE, DTU

32. Modeling and Simulation of Molecular Communication Based Nanonetwork Using Finite Shaped Spherical Receiver, ***6.Garima Singh*** & ***3.Gurjit Kaur***, Electronics, DTU

33. Modelling Of Wind Turbine Magnet for Magnetic Abrasive Finishing and Magnetic Field Assisted Abrasive Flow Machining Process, ***int Bhardwaj***, ***3.Krovvidi Srinivas*** and ***3.Rajiv Chaudhary***, Mechanical, DTU

34. Modified Current Control for Tracking Global Peak under fast changing Partial Shading Conditions, Raghavendra Rao P, Vignesh Kumar V, Balasubramanian Venkatesaperumal and ***3.Vanjari Venkata Ramana***, Electrical, DTU

35. Multi-Band Polarization Insensitive Ultra-Thin THz Metamaterial Absorber for Imaging and EMI Shielding Applications, Gaurav Saxena, Yogita Khanna, Y K Awasthi and ***3.Priyanka Jain***, Electronics, DTU

36. Multi-purpose hybrid optimization (DA-ALO) algorithm for efficient job shop scheduling issues, Vineet Kumar and ***3.Radhey Shyam Mishra***, Mechanical, DTU

37. Multiview Running and Walking Gait Analysis using the First and Third Person Data, ***6.Nikita Malik*** and ***3.Sudipta Majumdar***, Electronics, DTU

38. NATURAL LANGUAGE PROCESSING AND SENTIMENT ANALYSIS, ***8.Aryaman Sharma***, CSE, DTU

39. Neural network approach for faster optical properties predictions for different PCF designs, ***8.Hardik Kumar***, ***8.Tanya Jain***, ***8.Mritunjay Sharma*** and ***3.Kamal Kishor***, Applied Physics, DTU

40. Nonlinear adaptive normalized Huber control algorithm for 5-level distribution static compensator, ***6.Praveen Bansal*** and ***3.Alka Singh***, Electrical, DTU

41. Optimizing the Machining performance of CNC tools inserts coated with Diamond like Carbon Coatings under the dry cutting environment, Ramakant Rana, Lucky Krishnia, ***3.Qasim Murtaza*** and R.S. Walia, Mechanical, DTU

42. People Lie, Actions Don't! Modeling Infodemic Proliferation Predictors among Social Media Users, ***7.Chahat Raj*** and ***3.Priyanka Meel***, IT, DTU

43. Performance evaluation of NH3H2O vapour absorption refrigeration system with three cascaded vapour compression refrigeration systems using HFO refrigerants for ultra-low temperature applications, ***3.R. S. Mishra***, Mechanical, DTU

44. Performance evaluation of single effect Li/Br-H2O vapour absorption refrigeration system with three cascaded vapour compression refrigeration systems using HFO refrigerants for ultra-low temperature applications, ***3.R. S. Mishra***, Mechanical, DTU

45. Performance optimization for UWB communication network under IEEE 802.15.4a channel conditions, Abhishek Sharma, ***8.1.Amik Garg***, Sanjay Kumar Sharma, Vibhav Kumar Sachan and Parvin Kumar, Electronics, DTU

46. REVERSE ENGINEERING, ***8.Omkaar Mazumdar***, ***8.Sumeet Padhi*** and ***3.Prof. A.K. Madan***, Mechanical, DTU

47. Reviewing the Analeptic Activity of Calcium Citrate Malate, Bharat Kwatra*, Chelsea Rumao, Hiya Abrol, Ishika Gulati, Pankaj Pradeep & Srashti Bajpai, DTU

48. SmartSplit: Latency-Energy-Memory Optimisation for CNN Splitting on Smartphone Environment, ***8.Ishan Prakash***, ***8.Aniruddh Bansal***, Rohit Verma and Rajeev Shorey, Electrical, DTU

49. Strategies for assessment and implementation of sustainable manufacturing, ***8.1.Deepak Sharma***, Mechanical, DTU

50. TCAD investigation of ferroelectric based substrate MOSFET for digital application, ***6.Rashi Mann*** and ***3.Rishu Chaujar***, Applied Physics, DTU

51. The Merger of Topology Optimisation in Additive Manufacturing, ***3.Prof. A. K. Madan*** and ***7.Srashti Saxena***, Mechanical, DTU

52. Thermodynamic exergy analysis of two-stage vapour compression refrigeration systems using HFO refrigerants for replacing R134a, ***3.R. S. Mishra***, Mechanical, DTU

53. Trajectory Tracking of a 2-DOF Helicopter System using Fuzzy Controller Approach, ***3.Abhishek Chaudhary*** and ***3.Bharat Bhushan***, Electrical, DTU

54. Tribological Analysis of novel Apricot oil based Biolubricant against 15W40 oil tested on High Temperature Tribometer, ***6.Anshul Kumar***, ***3.Rajiv Chaudhary*** and ***3.Ramesh Chand Singh***, Mechanical, DTU

55. Wear and Residual Stress Analysis of Waste Sea Shell and B4C Particles Reinforced Green Hybrid Aluminium Metal Composite, Vipin Kumar Sharma, Sri Kant Rana, ***3.Roop Lal*** and Ramakant Rana, Mechanical, DTU

1. *Vice Chancellor*
2. *Pro Vice Chancellor*
3. *Faculty*
4. *Teaching-cum-Research Fellow*
5. *Asst. Librarian*
6. *Research Scholar*
7. *PG Scholar*
8. *Undergraduate Student*

1.1. *Ex Vice chancellor*
2.1. *Ex Pro Vice Chancellor*
3.1. *Ex Faculty*
4.1. *Alumni*
5.1 *Others*
6.1. *Ex Research Scholar*
7.1. *Ex PG Scholar*
8.1. *Ex Undergraduate Student*

Computers and Electrical Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

A new intelligent multilayer framework for insider threat detection

Mohammed Nasser Al-Mhiqani [a,*], Rabiah Ahmad [a], Z. Zainal Abidin [a], Karrar Hameed Abdulkareem [b], Mazin Abed Mohammed [c], Deepak Gupta [d], K. Shankar [e]

[a] Center for Advanced Computing Technology, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia
[b] College of Agriculture, Al-Muthanna University, Samawah, Al Muthanna Province, Iraq
[c] College of Computer Science and Information Technology, University of Anbar, Anbar, Iraq
[d] Maharaja Agrasen Institute of Technology, Sector-22, Rohini, Delhi, India
[e] Department of Computer Applications, Alagappa University, Karaikudi, India

ARTICLE INFO

Keywords:
Insider threat
Cybersecurity
Classification
Entropy
VIKOR

ABSTRACT

In several earlier studies, machine learning (ML) has been widely used for building insider threat detection systems. However, the selection of the most appropriate ML classification model for insider threats detection remains a challenge. Despite the prominence of ML in the domain of insider threat detection, none of the previous works have utilized ML techniques for building a hybrid solution that can take advantage of the misuse and anomaly insider threat detection. In this study, a new multilayer framework has been proposed for insider threat detection. The first layer of the framework is used for selecting the best insider threat detection classification model among many based on the multi-criteria decision making techniques. The selection procedure has been developed based on the integration of the entropy-VIKOR methods. For the second layer, a hybrid insider threat detection method has been proposed, where the Misuse Insider Threat Detection (MITD) model has been created using the random forest algorithm. Subsequently, using the K-Nearest Neighbors algorithm, an anomaly insider threat detection algorithm has been developed. The proposed multilayer framework for insider threat detection has been evaluated by using the CERT r4.2 dataset. Results of the experiment demonstrate that the validity of the results produced by the selection framework is proven by the validation procedure obtained from previous research. The proposed hybrid detection method is observed to exhibit an overall accuracy of 99% and a false positive rate of 0.29% for known insider threats, whereas it exhibits 97% accuracy and 2.88% false-positive rate for unknown insider threats.

1. Introduction

Information technology (IT) systems have significantly changed and improved human lives in various ways [1] However, the

This paper is for special section VSI-mlsec. Reviews were processed by Guest Editor Dr. Suyel Namasudra and recommended for publication.

* Corresponding author.
E-mail addresses: almohaiqny@gmail.com (M.N. Al-Mhiqani), rabiah@utem.edu.my (R. Ahmad).

https://doi.org/10.1016/j.compeleceng.2021.107597
Received 16 January 2021; Received in revised form 5 July 2021; Accepted 4 November 2021

security threats in IT are related to all the components of the current computing infrastructure world. One of these threats is the malicious insider threat. The problem of insider threats has been a critical source of threats to any organization owing to a large number of opportunities and accessibilities that an insider has, which can cause massive damage to the organization [2]. Unlike an outsider, a malicious insider has legitimate access to the information on various facilities and is also well-informed about the organization and its critical facilities. As a result, it becomes easy for the malicious insider to carry out malicious activities and hide their hacking trail [3].

In recent years, well-known incidents of insider threats have made their way in the public media, one of which was the high-profile insider case involving Edward Snowden, and another was the data leakage cases by Chelsea Manning and Kim [4]. In the insider threat cybersecurity report released in 2018, more than half of the studied threats (around 53%) in the past 12 months have originated from within organizations. Furthermore, 27% percent of the respondents have stated that frequent insider threats have been experienced by their organizations in the past 12 months [5]. However, in the recent cybersecurity report, the number of respondents who believe that an insider attack has been frequent has almost doubled, with 63% of the respondents believing that insider attacks have become more frequent in the past 12 months. Additionally, 53% of the respondents believe that detecting insider attacks is becoming significantly difficult [6]. Compared to outside intruders, there is a wide consensus that insiders are more capable of conducting an attack even without possessing advanced technical skills or experience. The detection of insider attacks is considerably difficult compared to the attacks initiating by outsiders whose hacking trails are difficult to remain hidden [4]. This is due to the fact that an insider is an authorized user of the system, which makes it considerably difficult to draw a line between malicious and legitimate activities, and the insider threat attacks manifest in various ways, such as a traitor insider planting a logic bomb for disrupting the origination systems or stealing classified data. The variety in insider attacks increases the complexity of detecting such insider threats. Lastly, insider attacks are often carried out by malicious or careless insiders within the daily working time, thus scattering the abnormal activities of the insider in large amounts of benign employees' behavior logs. However, analyzing such logs of activities would be considerably difficult for any analyst because of the sheer volume of the daily activities [7].

The prominence of machine learning (ML) is currently increasing in the domain of insider threats. In many previous studies, ML algorithms were employed for constructing a classifier that is capable of detecting insider threats. Nevertheless, the choice of the most appropriate ML classification model remains a "hard nut to crack", and it is important to have the correct model for accurately detecting insider threats [8]. The difficulty in selecting the most appropriate ML model is due to the existence of several detection performance evaluation criteria such as false positive rate, accuracy, recall, etc. Therefore, no framework is available at present that covers all aspects of the insider threat detection system evaluation [9]. Therefore, the selection of the best detection method for insider threats is still a difficult and challenging decision-making task. Another challenge with the studies done to date is that most of the existing solutions are based on anomaly or misuse detection only. Misuse detection methods detect insider attacks based on the existing insider threat scenarios [10]. They are effective in detecting the known insider attacks with low errors. However, they cannot detect unknown insider attacks [8]. In contrast, anomaly insider threat detection is based on the hypothesis that the behavior of a malicious insider differs from that of a benign user. Anomaly detection methods are useful for detecting new insider attacks, but they are not as effective as the misuse detection models in detecting known insider attacks. However, a robust and good detection method should use a combination of different independent approaches [10].

To overcome the abovementioned challenges, a multilayer framework has been proposed in this study. Firstly, a special framework has been proposed for ranking and selecting the existing ML classification models used in insider threat detection. In this study, an adopted decision matrix (DM) has been built by combining nine classification models with the multi-evaluation criteria. The classification models are processed through the ranking and selection process using the multi-criteria decision making (MCDM) algorithm. The first part in MCDM is the evaluation criteria that are calculated using entropy. Another technique, called VIKOR, has been used for benchmarking and ranking the classification models. Secondly, based on the results of the first layer, the best two classification models are selected to build the second layer, the misuse insider threats detection is built based on the random forest classification algorithm, and based on the k-nearest neighbors method, the anomaly insider threat detection model is built. The following are the contributions of this study:

- A new intelligent multilayer framework has been proposed for selecting the best insider threat detection model and as a hybrid detection model for insider threat detection.
- A new hybrid insider threat detection system based on a combination of misuse and anomaly insider threat detection has been proposed for improving the detection of insider threats.
- Using the CERT benchmark datasets from Carnegie Mellon University, the proposed multilayer framework has been evaluated in terms of all the frequently used classification metrics such as accuracy, recall, precision, F-score, area under the curve (AUC), false positive rate (FPR), true negative rate (TNR), false negative rate (FNR), and computation time.

The remainder of this paper has been organized as follows: Section 2 discusses the related works. The background of the study is described in Section 3. Section 4 describes the proposed multilayer framework. Security analysis of the proposed multilayer framework have been discussed in Section 5. Performance analysis is illustrated in Section 6, and finally, the conclusions and future works are described in Section 7 and Section 8 respectively.

2. Related works

In terms of selecting the best ML technique, there is no framework available at present for evaluating and benchmarking the

Table 1
Critical review analysis of various studies in the evaluation and detection of insider threats using different ML classification models.

Study	Study Scheme	Key Finding	ML model	Evaluation Criteria Used	Remarks
[13]	Evaluation and model selection	Evaluated six ML classifiers as detection tools with the aim of showing the viability of the evaluated tools for detecting new malicious scenarios by using command and injection of data.	NB, RF, OneR, J48, NNge, and SVM	Recall and Precision	It is one of the first works that evaluated ML techniques and considered the insider threat problem. However, only recall and precision was used for evaluation.
[14]	Evaluation and model selection	Evaluated multiple learning algorithms using the Weka toolkit and CERT insider threat datasets.	Neural Network, NB, SVM, RF, DT, and LR	Accuracy and ROC	Only accuracy and ROC were used for evaluating the performance and only a few classifiers were evaluated.
[11]	Evaluation and model selection	Evaluated several ML classifiers to choose the best technique for detecting insider threats that require new anticipatory analytics for capturing the risky behavior before losing data.	Multiple ML classifiers	Accuracy and time	It is the first study that evaluated more than 10 ML classifiers using the CERT dataset. However, accuracy and time were the only evaluation metrics used.
[15]	Evaluation and model selection	By focusing on the extremely imbalanced dataset of insider threats, this study evaluated three ML classifiers.	J48, NB, SVM, and RF	Accuracy, Time, TPR, FPR, Recall, Precision, and F-score	One of the few works that cover several performance measurement criteria. However, only three classifiers were evaluated and all the evaluation criteria were not covered.
[17]	Detection	Introduced a surveillance method for detecting insider threats, operates in two phases, namely, anomaly and misuse detection	KNN and Dempster–Shafer theory	FPR	This work did not present a performance evaluation. Based on the analysis, the focus was on detecting the outsider and the masquerader more than the malicious insider.
[18]	Detection	Two methods were proposed in this work for suspicious e-mails. Pattern matching using the regular expression algorithm and k-means clustering was used in the first stage and supervised ML was used in the second stage.	Regular expression pattern matching algorithm and k-means clustering	Accuracy and Precision	This study used a combination of several independent approaches. However, the article focused on insider IT sabotage using only e-mail and only accuracy and precision were used as the measurement performance metrics.
[7]	Detection	A framework was proposed to model the problem of malicious-insider-based behavioral and psychological observation. The proposed framework provided the analyst with the ability of reasoning based on the data on a user and building hypothesis trees that describe the potential insider threats.	Principle component analysis and tree structure	Precision and Recall	A human analyst was still required to follow up on the false and true positive outcomes and to mark them as such for refining the parameters in the system and giving feedback to the system to continue the learning with a human analyst in the loop. Because of the different data formats, the identified anomaly metrics presented might not always be suitable and new insider anomalies generated from the users' logs will need to be defined.

different insider threat detection systems [10]. Thus, the selection of the best insider threat detection approaches is still difficult and a challenging decision-making task, owing to the availability of multiple detection evaluation metrics, such as recall, precision, accuracy, FPR, computation time, etc., for evaluating the models. Noever [11] compared 16 major families of ML models using the median predictive accuracy and relative execution time to evaluate the classifiers. They validated the performance of the models using the CERT insider threat dataset [12] by partitioning the training and testing data, along with repeated cross-fold validation in the under-sampled and over-sampled cases. Beaver et al. [13] evaluated six ML algorithms as intrusion detection tools using the Modbus RTU dataset with the aim of showing the viability of the evaluated tools. The authors used recall and precision as the evaluation metrics for comparing the results of the studied classifiers. Hall et al. [14] evaluated multiple learning algorithms using the Weka toolkit. The performance of the models was evaluated using the receiver operating characteristic (ROC) curve and accuracy as the evaluation metrics. The classification models used in this study include RF, naïve Bayesian network, neural network, DT, logistic regression (LR), and support vector machine (SVM). Similar to the previous work, Sheykhkanloo and Hall [15] evaluated five ML classifiers and gave a score to the classifiers based on the obtained results.

Many studies have been conducted to date on the detection methodology of insider threats. However, only a few hybrid solutions have been proposed. For detecting the network data exfiltration, Maloof and Stephens [16] proposed the ELICIT system which stand for exploit latent information to counter insider threats centered on 76 binary detectors that investigate the abnormal patterns and volumetric anomalies. The outputs of the detectors were transferred to the Bayesian networks (BN) for evaluating the threats. Taking the cyber-observables improved by psychological measurements into account, Legg et al. [7] suggested a hybrid approach based on known attacks, policy violations, thresholds, and abnormalities. Because of the different formats of the data, the identified anomaly metrics presented in this work might not always be suitable, and new insider anomalies generated from the users' logs will be required

to be defined. Punithavathani et al. [17] analyzed the network traffic for detecting insider threats by using the KNN classification model and the Dempster-Shafer theory. Nevertheless, the evaluation of their proposed network was implemented for a smaller size sample, and thus the results cannot be generalized. In addition, the article did not present any performance measures. A two stages approach was proposed by Michael and Eloff [18], where a regular expression pattern matching algorithm and k-means clustering were applied in the first stage, and three different supervised ML techniques were used for classifying the output of the first stage.

From Table 1, it can be seen that the insider threat evaluation methods have suffered from two main challenges: (i) Based on the discussed articles, it can be seen that the selection of the best insider threat detection classification model is still difficult and challenging task. Only a few works have evaluated the different ML classification models in the domain of insider threats and validated their performance using standard insider threat datasets. However, these works had many limitations, such as depending on the evaluation and benchmarking of the accuracy and one or two more metrics only. In addition, evaluation of insider threat detection should not depend on using only accuracy [19]. (ii) In terms of the detection methodology, it can be seen that the major proposed solutions are based on a single model [17,18,7]. However, these works have limitations, as discussed in Table 1.

Despite the considerable efforts made in the area of selecting and evaluating insider threat detection classification tasks, none of the previous works have considered all the main evaluation metrics for assessing the classification models in order to help researchers and practitioners in the field of insider threat detection to select a suitable classification model. Furthermore, a majority of the proposed solutions are based on anomaly or misuse detection. Despite the prominence of ML in the domain of insider threat detection, none of the previous works have investigated the use of ML techniques for building a hybrid solution that can take advantage of the misuse and anomaly detection to tackle the well-known and zero-day insider attacks.

Thus, the present study has been conducted to fill in the identified research gap in the area of selecting the classification models for insider threat detection in the first layer and the hybridization of the detection methodology using ML in the second layer to detect the insider attacks.

3. Background studies

The problem of insider threats has become an extensively reported problem and is considered as one of the big challenges in cybersecurity [20]. This indicates that insider threats require special detection methods, systems, and tools, which entail the capability to facilitate fast and accurate insider threat detection. Many studies have been conducted on insider threat detection and the related domain to deal with this challenging issue.

3.1. Multi-criteria decision making

Keeney and Raiffa [21] stated that the MCDM method is an extended version of the theory of decision that covers any multi-objective decision. MCDM involves the evaluation of alternatives that mostly have conflicting evaluation criteria. MCDM can also be used for consolidating the contradictory criteria in a separate universal assessment.

Weighting methods are categorized in terms of two contexts in MCDM, namely, subjective and objective methods. One of the objective methods is entropy weighting, which is used for determining the criteria weight [22].

When determining the weights of the indices, subjective fixed weight methods, such as the expert survey, analytic hierarchy process (AHP), and Delphi methods, are generally used. These methods can lead to deviation of the index weights due to subjective factors. The objective fixed weighting method is based on the inherent information of the indices for determining their weights, which can help in eliminating man-made disturbances and enable the results to be in accordance with the facts [23].

If dependable subjective weights are not produced, the objective weighting approach is well suited. From another point of view, the weights calculated by the subjective strategies reflect the opinion of the decision-maker or subjective judgment. However, due to the lack of knowledge or experience, the alternative classification methods might be significantly impacted by the weights given by the decision-maker [24].

In objective (non-subjective) methods, the weights are automatically defined by mathematical procedures, such as multi-target programming and entropy methods, without considering the preferences of the decision-makers. However, the benchmarking and selection result will be influenced by the type of the weighting approach. Hence, the selection of the most appropriate weighting method will have a significant impact on the results.

Shannon introduced the concept of entropy in 1948 to measure the uncertainty in the formulated information. Different fields of study, such as management, engineering, and computer science, have witnessed the implementation of the entropy model. In accordance with the ideal theory of entropy, the amount of information or quality of the process of decision-making is determined by the reliability and accuracy of the decision-making problem. Hence, the entropy method can be employed in assessing the problems associated with the diverse decision-making processes. Furthermore, entropy can also be used for analyzing the information quantity provided by the data [25].

In the domain of insider threat detection, the MCDM method can be used for performing an extensive evaluation of multiple criteria to ensure reliability of the results. The weight of each criterion can be determined using the entropy method depending on the extent of variation of its value and providing a base for a comprehensive evaluation [22]. Notably, the entropy method increases monotonically as a function of uncertainty. A smaller uncertainty results in a smaller entropy and vice versa. Therefore, the degree of dispersion of each criterion can be judged from the entropy value calculation. An increase in the degree of dispersion of the index can have an additional effect on the overall evaluation, and therefore the weight of an index should be large. Table 2 shows the comparison of the entropy method with other feasible weighting techniques.

Table 2
Summary of the different MCDM weighting methods.

Technique	Advantages	Disadvantages
AHP	(i) Scalable(ii) Easy to use (iii) Not data intensive (iv) Can fit several sized problems because of its hierarchy structure (v) Compared to the other techniques, the computation process of AHP is quite simple [26].	(i) Interdependence between alternatives and criteria (ii) Might result an inconsistency between the judgment and the ranking criteria (iii) Reversal of rank. The more the number of decision makers involved, the more complex is the weight assignment process [26].
WSM	(i) Computation process is very simple (ii) Suitable for managing problems of a single-dimension [26]	(i) Multiple preferences cannot be integrated with WSM (ii) Only one dimension can be evaluated [26]
BWM	(i) Has the ability of eliminating any element to be measured (ii) It can use proportional values instead of real (actual) values. (iii) Compared to AHP, BWM is always consistent [27]	If the number of criteria increase, the number of pairwise comparison will significantly increase in BWM, which greatly complicates the mathematical formulation of the model [27].
Entropy	(i) Simplifies understanding of the problem (ii) Entropy methods automatically define the weight by solving mathematical procedures without considering the preferences of the decision makers	If the raw of the data changes, the precision of weight obtained for the evaluation criteria might be also changed.

Nevertheless, the VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR) method and the technique for order of preference by similarity to ideal solution (TOPSIS) method are very popular among the many alternatives corresponding to various criteria. Both models are applicable in objective and quantitative data provision. The key principle of TOPSIS is based on the distance between the positive and negative solutions. Conversely, it does not take into account the importance of comparing these distances [28]. The VIKOR process fits with the discrete alternative, taking into account the most realistic approach in dealing with such problems in the real world. The key benefit of VIKOR is that the best alternative can indeed be swiftly pinpointed. In situations with many alternatives and attributes, VIKOR is therefore acceptable [28] and has been used in this work to rank the different classification models. VIKOR employs a methodology in which the priority for various response optimizations is compromised[29]. This method simply ranks the different classification models depending on their closeness to the optimal solution and finds the best model easily and accurately [29]. Lately, the adaption style in the VIKOR method has been changed by its integration with another MCDM method. Earlier studies have shown that integration of VIKOR with entropy has several advantages that ensure reliable and vastly coherent objective weights [30]. Therefore, increasingly robust methods have been proposed. Shemshadi et al. [30] discussed the main advantages of combining these two applied methods in terms of tackling uncertainty issues linked to a problem. Finally, integration of entropy and VIKOR is required for evaluating and benchmarking the insider threat detection classifiers. The foundation for the integration is shaped through the assignment of weight for each criterion according to their entropy values. However, the VIKOR method is suggested for ranking the different insider threat detection classifiers.

3.2. Machine learning techniques

3.2.1. Random forest

The random forest (RF) classifier is an ensemble classification technique, which uses multiple decision tree (DT) models to achieve a superior classification performance. It generates several sub-trees, which are used for training each tree from the training dataset. This approach considers only the random subset of the variable so that any node has a split. For classification, every tree in the RF that votes for one class is fed with the input vector. Subsequently, the RF selects the class with the most votes. It can accommodate larger datasets in contrast to other methods [31].

The RF method enhances bagged trees by de-correlating the trees, thus reducing the disparity in the average of the trees. When DTs (in parallel) are generated, a random selection of m predictors is picked as a subset of the split applicants from a full collection of predictors each time a split in the tree has been considered. Therefore, a new set of m predictors created for each split, and the choice of the map implies that at each split, (m) the number of predictors is approximately equal to the square root of the total predictors p.

Suppose that the training set is $S = \{S^1, S^2, \ldots, S^n\}$ consisting of n training samples. Let B be the number of predefined parameter of the bootstrap trees in RF, and k_{RF} be a predefined parameter of a fixed number of variables for selecting the subset (i.e., feature subspace size). In many implementations, the default value is $k_{RF} = \int^m$ in which m denotes the number of features, but $k_{RF} = \log_2 m$ is used as well. The normalization of RF is performed as follows:

- **Step One:** Bagging is applied repeatedly by RF for a total of B times in order to select B samples of the random bootstrap for B bootstrap trees. Thus, for every sample of the bootstrap, S_b, $1 \leq b \leq$ B is chosen, with the replacement from the set of m-sized original training.
- **Step Two:** In every tree, T_b, $1 \leq b \leq B$ and the main root node, r_b, is built of the entire bootstrap sample S_b.
- **Step Three:** In every node, $n_{b,i}$, firstly, the RF selects k_{RF} random features (i.e., k_{RF} feature subset size), and secondly, searches through the entire subset of the features to decide the best node to be split into two child nodes. In contrast to the traditional tree

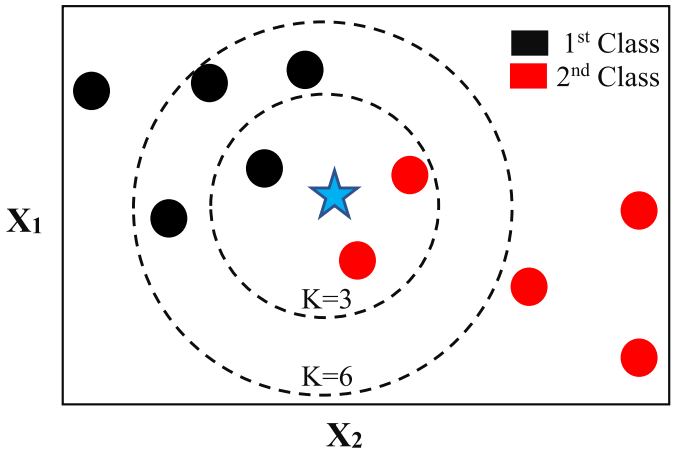


Fig. 1. Schematic showing the search process in the KNN method (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.).

pruning process, the trees of RF are growing all the way down and are not pruned. For achieving high memory efficiency, a minimum gain or a maximum depth (i.e., improvement) in all the implementations is set to stop the deep growth of the trees.

- **Step Four:** For the new instance classification in S_t at the session slot, t, firstly, for every tree, T_b, in RF, a unit is voting for the class label, and secondly, the label $y_t^{'}$ of the predicted class is selected for the S_t that has the majority of B votes.

3.2.2. *k-nearest neighbors*

The k-nearest neighbors (KNN) technique is a non-parametric method, which has been used for classification problems. This technique customizes the information of its neighbors' points for the output label classification. This method has been used widely in the security domain, such as in intrusion detection and spam and malware detection. Typically, the KNN algorithm selects the nearest samples based on the measurement of the distance between all the samples of the training set and the new samples. It has the ability to predict the target class in a simpler manner and with high precision. This technique belongs to the lazy-learning algorithm type. Its function is only locally estimated, and the calculation is delayed until the classification process [32].

KNN is an efficient approach that exhibits significant strength for noisy training data or a massive dataset.

Fig. 1 shows the KNN search scenario. If the value of K is set to 3, there are two observations of the 2nd class and one observation of the 1st class. Thus, the new instance will be classified with the red circle to the 2nd class. If the value of K is set to 6, there are two observations of the 2nd class and four observations of the 1st class. Thus, the new instance will be classified with the black circle to the 1st class.

The algorithm typically selects the nearest instances based on the distance measurement between all the training samples and the new samples. For real-valued input instances, the most widely used distance measurement is the Euclidean distance.

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{1}$$

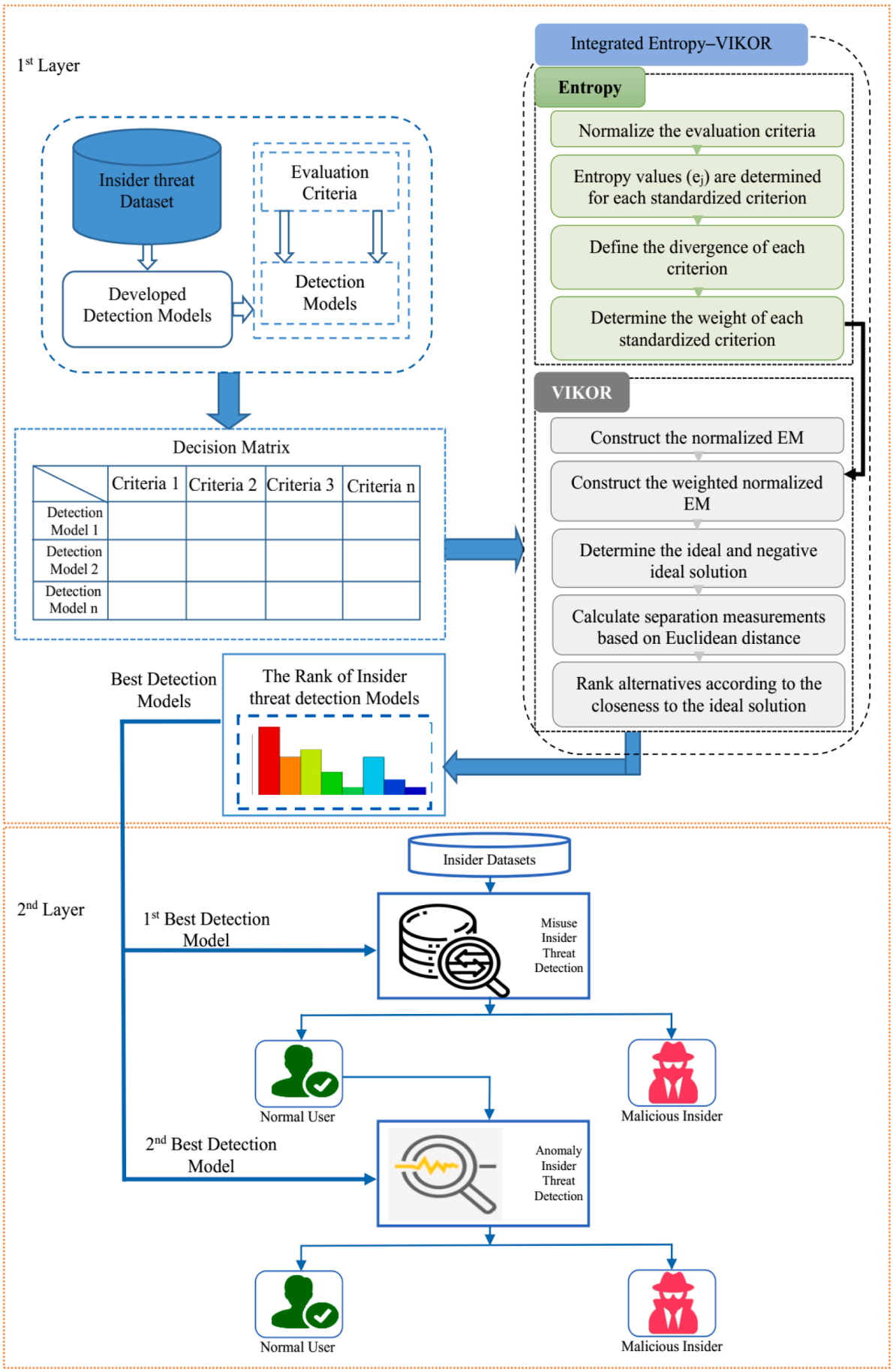
The Euclidean distance is defined as the square root of the sum of the squared differences between a new sample point and the current points in all attributes of the input.

4. Proposed multilayer framework

This section describes the working principle of the proposed multilayer framework developed for selecting and detecting insider threats using the best classification model. This multilayer framework outputs a ranking of the various classification models based on multiple criteria using an integration of the entropy method for weighting and VIKOR method for ranking in the first layer and a hybrid insider threat detection in the second layer. Fig. 2 illustrates the different components of the proposed multilayer framework. As can be seen from the figure, two major steps are involved in developing the multilayer framework: (i) Construction of the adopted decision matrix and (ii) Calculation of weights of the classification models for insider threat detection using the entropy method and ranking the models using the VIKOR method. Subsequently, based on the result of the first layer, the hybrid insider threat detection system is built.

4.1. *Selection layer*

The MCDM techniques, discussed in Section 3.1, have been analyzed to choose a suitable method for ranking the classification



(caption on next page)

Fig. 2. Proposed multilayer insider threat detection.

Table 3
List of notations.

Symbol	Name
d_j	Degree of Diversification
e_j	Entropy Value
f_i^-	Worst Value
f_i*	Best Value
m	Number of Alternatives
n	Number of Criteria
Q_j	Final Ranking
R^-	Worst Value of R_i
$R*$	Best Value of R_i
r_{ij}	Normalized Matrix
R_j	Individual Regret
S^-	Worst Value of S_i
$S*$	Best value of S_i
S_j	Unity Measure
v	The Maximum Group Utility
w_j	Weight Vector
WM	Weighted Matrix

models. Therefore, the next subsection discusses the components and the selected MCDM techniques used for building the selection layer of the multilayer framework. The notations used in this article are given in Table 3.

4.1.1. Insider threat ML models

Numerous ML classification techniques have been employed in modeling insider threat behavior. Nine classification techniques that have been widely used in detecting insider threats and these classification models are considered as the alternatives in the DM. These classification models include the following: Naïve Bayes (NB), LR, RF, SVM, KNN, DT, long short-term memory (LSTM), gated recurrent units (GRU), and multilayer perceptron (MLP).

From the abovementioned classification models, the SVM technique has the highest frequency in terms of usage for insider threat detection. However, even with the popularity of the SVM model, there is no study that has identified the level of importance for any particular model, and this raises a significant challenge in terms of selecting the most optimal model.

4.1.2. Evaluation criteria

Several classification evaluation metrics have been used for evaluating the insider threat detection models, and some of them are known by multiple names, for instance, some researchers use the term sensitivity, while others use true positive rate (TPR) or recall and so on.

For binary classification, a confusion matrix is formed using the four predicted outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

TP represents the number of malicious insider activities that are classified correctly, TN represents the number of benign user activities that are classified correctly, FP represents the number of benign user activities that are classified incorrectly, and FN represents the number of malicious insider activities that are classified incorrectly. Table 4 presents the most frequently used performance evaluation metrics that have also been used as the criteria for the classification models.

From the above table, the accuracy criterion is the most widely used, whereas TNR is the least used criterion. However, all studies have a significant variation in terms of the usage of all nine criteria. Furthermore, there are no studies that have identified the importance level of these individual criteria for insider threat detection. Nevertheless, in contrast to all studies, the present research includes all the abovementioned criteria in order to provide a comprehensive evaluation and benchmarking scenario.

4.1.3. Construction of the decision matrix

Generally, in terms of evaluation and benchmarking for ML techniques, only two studies have provided new DMs [33,34]. On the one hand, in binary classification, Yas et al. [33] have provided a new DM based on the crossover of skin detection models and the identified criteria. On the other hand, for multiclass classification, Alsalem et al. [34] have presented a new DM based on the crossover of automated acute leukemia detection models and the identified criteria. However, given that our study is based on binary classification, we have adopted the DM proposed by Yas et al. [33], but within the selected alternatives and attributes in the insider threat case study. Nonetheless, nine ML classification models, which have been used for building an insider threat detection model, have been used in this study. The CERT dataset, described in Section 5.2 was split into two main parts to enable the building of the classification models. The classification models were trained using the first part, whereas the other part was utilized for testing them. The DM contains m of alternatives and n of criteria that need to be formed. The alternative and criteria intersection are given as x_{ij}. Hence, this work will have a matrix (x_{ij}) _ (m*n), which can be described as follows:

$$A = \begin{matrix} A_1 \\ A_2 \\ \cdots \\ A_m \end{matrix} \begin{bmatrix} C_1 & C_2 & \cdots & C_n \\ X_{11} & X_{12} & \ldots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \cdots & \ldots & \cdots & \ldots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix} \quad (2)$$

where A_1, A_2, …, A_m are potential alternatives to be scored by decision makers (i.e., classification model), C_1, C_2, …, C_n are the provided criteria that are used in measuring the performance of each alternative in terms of accuracy and recall, x_{ij} is the score of alternative A_i along with criterion C_j, and W_j represents the weight value of criterion C_j. The alternatives can be ranked using different kinds of processes, such as the addition of weights, maximization index, normalization, and other processes based on the technique.

The establishment of DM depends on the crossover between the selected evaluation criteria, namely, accuracy, precision, recall, AUC/ROC, F-score, FPR, TNR, FNR, and computation time, and the nine classification models.

The DM structure, which is based on the insider threat evaluation requirements, consists of the nine evaluation criteria, which are considered as attributes, and the classification models that have been developed for insider threat detection, which are considered as alternatives in the DM. The values in the DM indicate the evaluation results of the different insider threat classification models corresponding to the nine evaluation criteria. Based on these evaluation criteria, the evaluation of each classification model for insider threat detection was done.

4.1.4. Integrated entropy and VIKOR

In order to develop a procedure for insider threat detection based on DM for selecting the best classification model, a hybrid entropy-VIKOR method has been proposed, where the criterion weight from the entropy method is linked with the second step of the VIKOR method as shown in Fig. 2. The VIKOR approach is applied to rank and solve (1) the multi evaluation-criteria, (2) trade-off issues faced by the proposed DM, and (3) the importance of the criteria of the proposed DM. The entropy method also determines the weights of the criteria that are to be solved.

4.1.4.1. Entropy weights for criteria. The entropy method was used for determining the weights of various criteria as follows:

Step 1: Normalize the evaluation criteria as shown in Eq. (3):

$$r_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}. \quad (3)$$

A multi-criteria problem DM with m alternatives and n criteria, where x_{ij} ($i = 1, 2, \ldots, m; j = 1, 2, \ldots, n$) represents the i^{th} alternative performance value for the j^{th} criteria.

Step 2: Calculate the measure of entropy, e_j, for each criterion:

$$e_j = -h \sum_{j=1}^{m} r_{ij} \cdot \ln r_{ij}, \quad j = 1, 2, \ldots..n, \quad (4)$$

where h is the constant of the entropy equal to$(\ln m)^{-1}$, and $r_{ij}.\ln r_{ij}$is defined as 0 if$r_{ij} = 0$ [35].

Step 3: Define the degree of diversification dj of each criterion using.

$$d_j = 1 - e_j \; j = 1, 2, \ldots..n., \quad (5)$$

If the value of $d_j$is high, then the relevance of the j^{th} criterion is high.

Step 4: Determine the normalized weights of all criteria as follows:

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j}, \; j = 1, 2, \ldots..n, \quad (6)$$

where $w_j$is the degree of importance of the criterion j.

If the entropy value is small, the entropy-based weight is also small, and the additional information is given by the specific criterion.

This criterion becomes more important than the other criteria considered in the decision-making process [35].

4.1.4.2. VIKOR method for ranking the classification models. The weighted matrix is typically the baseline for ranking the alternatives in the decision-making process. The previous section describes the process of obtaining the weights that are applied for each criterion in the DM, thereby producing the weighted DM. During this process, the classification models are evaluated and eventually ranked based on the weighted DM. The ranking process is described as follows:

Step 1: Calculate best value f_i^* and worst value f_i^- values for all criterion functions:

$$f_i^* = max_j f_{ij},\ f_i^- = min_j f_{ij} \tag{7}$$

where f_{ij} is the value of i^{th} criterion function for the alternative x_i. An ideal positive solution expands the benefit criteria and minimizes the cost criteria, whereas the cost criteria are maximized by the ideal negative solution which also minimizes the benefit criteria [36].

Step 2: Based on the entropy method, calculate the weights of every criterion. A weightset $w = w_1, w_2, w_3, \cdots, w_j, \cdots, w_n$ from a decision-maker is accommodated in the DM. This set is equal to 1. The resulting matrix can be calculated as follows:

$$WM = wi * \frac{f_i^* - f_{ij}}{\mathrm{f}_i^* - \mathrm{f}_i^-}. \tag{8}$$

The outcome of this step is a weighted DM.

Step 3: Calculate the values of Sj and Rj ($j = 1, 2, 3, \ldots,$ m and $i = 1, 2, 3, \ldots,$ n) using the following equations:

$$S_j = \sum_{i=1}^{n} w_i * \frac{f_i^* - f_{ij}}{\mathrm{f}_i^* - \mathrm{f}_i^-}, \tag{9}$$

$$R_j = \max_i w_i\ * \frac{f_i^* - f_{ij}}{\mathrm{f}_i^* - \mathrm{f}_i^-}, \tag{10}$$

where S_j and R_j indicate the measure of the utility and regret for the alternative f_i. w_i specifies the relative weights of the criterion.

Step 4: Determine the values Qj ($j = (1, 2, \cdots, J)$) using the following relation:

$$Q_j = \frac{v(S_j - S^*)}{S^- - S^*} + \frac{(1-v)(R_j - R^*)}{R^- - R^*}, \tag{11}$$

where $S^* = \min_j S_j$, $S^- = \max_j S_j$, $R^* = \min_j R_j$, $R^- = \max_j R_j$, and v is defined as the strategy weight of "the majority of criteria" (or "the maximum group utility") and $v = 0.5$.

Step 5: In this step, the alternatives can be run by sorting the values, thus obtaining the values of *S, R*, and *Q* in ascending order. The ideal performance is indicated by the smallest value.
Step 6: The alternative (a') is presented as the alternative solution and is ranked as the best solution depending on the minimum *Q* measure.

4.2. Detection layer

Based on the result from the first layer, two best classification models are used for building the second layer of the proposed hybrid solution, as shown in Fig. 2, for detecting insider threats. The new data are checked by the first model, namely, the misuse insider threat detection (MITD) model, and are classified as either malicious or benign users. If the MITD model detects a known sample, an alarm is issued. However, if the new data are classified as normal by the MITD model, it will be transferred to the anomaly insider threat detection (AITD) model, the second stage of the framework, for double-checking. Both these models are discussed below:

4.2.1. Misuse insider threat detection

MITD is applied in the first stage owing to its efficient performance in detecting familiar insider threats and low probability of

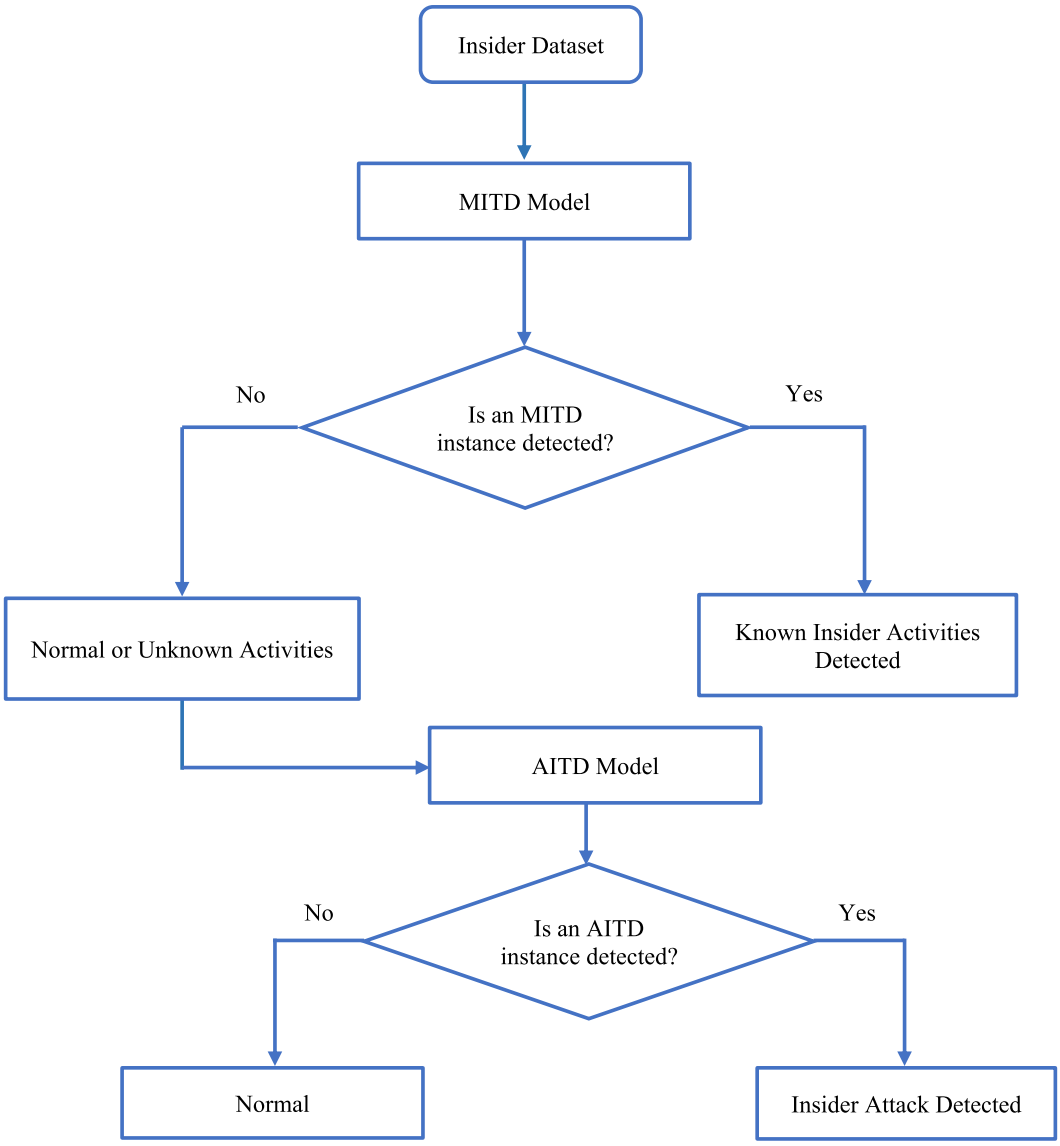


Fig. 3. Architecture of the HITD model.

inducing false alarms. This implies that the false positive outcomes are small since the database contains all the known signatures on malicious insider samples. Consequently, a high accuracy detection rate for attacks is achieved, and false positive rate is reduced.

4.2.2. Anomaly insider threat detection

To effectively detect new insider threats, the classified normal instances from the MITD model should be inputted into the AITD model in order to detect unknown new insider threats. AITD should be built based on the insiders' normal activities, where these activities are used for the training process. Subsequently, the malicious insiders are detected based on the measured state of the insider's profile, which is compared to the normal insider profile (determined on the basis of AITD). When this profile differs from that of the normal insider profile, it is considered a malicious insider.

4.2.3. Hybrid insider threat detection

MITD and AITD models have their unique strengths and weaknesses, and in this work, we have established a hybrid method incorporating both techniques. The working of the proposed hybrid insider threat detection (HITD) method is as follows: First, the MITD model is implemented using the training data. It is well-known that a model for misuse detection is capable of detecting known insider attacks with a very small FPR, but it cannot detect unknown attacks accurately. Since the MITD model has low false positive outcomes, the effect of the attack detection of this model is observed.

Second, for any normal training set decomposed by the MITD model, the AITD model is trained. The main reason for decomposing a normal training set is that in earlier developed hybrid threat detection approaches, the anomaly detection models have tried to use a single outlier detection pattern to profile the normal patterns. Fig. 3 shows the overall architecture of the hybrid insider threats detection.

5. Security analysis

5.1. Experiments

When conducting experiments, due to the size of the dataset, two environments were used: (i) For processing the dataset and preparing the dataset for the ML model, a local server using ubuntu 16.04 64-bit operating system running on a machine with an AMD Opteron(tm) Processor 6376 and 315 GB RAM was used. (ii) For ML, Google Colaboratory Notebook was used for training and testing the models. Google Colaboratory is an online cloud service, which is free and provides a Jupyter Notebook environment for ML research and education. The Jupyter Notebook, provided with Python 3.0, is preconfigured with all the essential ML libraries, such as Pandas, Scikit-learn, Keras, and Matplotlib, and these libraries do not require installation on a local machine in order to use them [37].

5.2. Details of datasets

The data that have been used in this study have been extracted from the CERT dataset, which is a combination of multiple sources of datasets on synthetic insider threats created by the CERT Division at Carnegie Mellon University and other partners [12]. In this study, CERT v4.2 has been used owing to the higher number of insider threat instances contained in this version as compared to the other versions of the CERT datasets and since it is the only comprehensive dataset available publicly [38,39]. For a period of 17 months, 32, 770,227 lines of logs have been generated by one thousand users, and there are 7323 malicious activities among these logs, which were manually added by the experts for representing malicious insider threat scenarios. The dataset has been created based on various scenarios that include masquerade attacks and traitor activities. The CERT dataset includes the activity logs of logon activity data, browsing history or hypertext transfer protocol (HTTP), file access logs, e-mail data, device usage, lightweight directory access protocol (LDAP), and psychometric data. Therefore, this study has been performed using the data corresponding to the cyber behavior of users in order to detect the different types of insider attacks that can be traced based on the available data. The values of the data were gathered as they were, however, the date features were split and encoded into two main features, namely, time and day. ML techniques understand the input only in the integer number format. Therefore, the date and time were converted into integer numbers and the same idea was applied to other features such as the activity feature. The initial values of the activity feature corresponded to the actions of the user, such as logging into the system, logging out, connecting or disconnecting a thumb drive, sending an e-mail, processing a file, or accessing the internet. Since the activity features have seven different user actions, each action was replaced with an integer number, i.e., login action (1), logout action (2), connect action (3), disconnect action (4), HTTP action (5), file action (6), and e-mail action (7).

Similar to the work of Nicolaou et al. [39], Table 5 presents the selected features along with their possible value space.

5.3. Insider attacks scenarios

Table 6 describes the three scenarios of insider threat attacks and the malicious users who conducted these attacks.

Based on the above insider threat scenarios, attack records could be classified into known insider attacks and unknown insider attacks in the CERT dataset. In addition, if the attack labels are similar, the attack scenarios might be different. Therefore, to discriminate the adequately known attacks from unknown attacks, the CERT datasets were organized as follows:

Firstly, for testing the proposed model on known attacks, the target dataset was separated into two sets, namely, a training set and a testing set. The training set contained the normal user and malicious activities from all three different insider attack scenarios that are defined in the CERT dataset, whereas the testing set consisted of normal and malicious insider activities that are known in the training sets. Secondly, for the unknown insider attack scenarios, the CERT dataset was divided into four different sets; 90% of the data from the first two sets were used for training, and 10% were used for testing. From the remaining two sets, 80% of the data were used for training, and 20% were used for testing, where the testing set contained random unknown malicious activities.

The training and testing process adopted for the CERT dataset in different stages is as follows:

Step 1: The CERT dataset was divided into two subsets, the first subset was used for training, and the second was used for testing.
Step 2: The MITD model was trained using the entire dataset to learn the known attacks and then tested based on the testing data subset.
Step 3: The AITD model was trained using the training dataset.
Step 4: The MITD model was tested using the CERT testing data. If the MITD model classified an instance as an insider, it labeled it as a malicious insider. However, if an instance was classified as benign, it was inputted in the AITD model for further analysis.
Step 5: The AITD model checked the samples that have been labeled as normal by the MITD model.

6. Performance analysis

This section presents the performance analysis results of the multilayer framework. Section 6.1 describes the selection framework results by presenting the results of weighting based on the entropy method and ranking based on the VIKOR method. The results of the second layer of the HITD model are presented in Section 6.2.

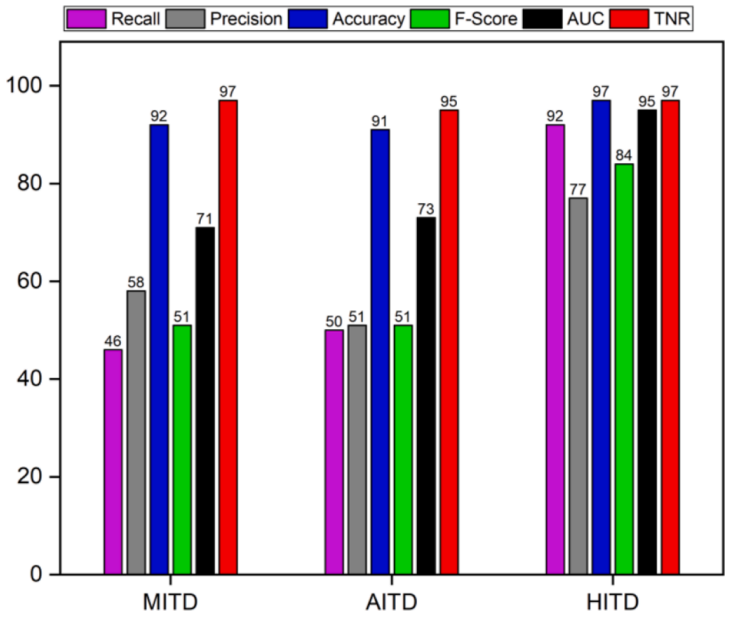


Fig. 4. Overall performance of the three models when applied to the first two datasets (90% training data and 10% testing data).

6.1. *Results of the first layer (selection)*

6.1.1. *Weighting results based on entropy*

Based on the DM data, the data normalization has been presented in Table 7. The criterion weighting has been done objectively by considering nine criteria for every model.

Table 8 shows the values and weights of the entropy determined for the nine criteria using the entropy method. The maximum weight of entropy is achieved by FPR, whereas the minimum weight of entropy is achieved by TNR. Based on the results, the highest entropy weight criteria are considered to be the most important criteria, whereas the lowest entropy weight criteria are considered to be the least important.

6.1.2. *Results of ranking based on the VIKOR method*

The results of the ranking of the different insider threat detection classification models based on the weighted DM are discussed in this subsection. The weighted DM can be obtained using the VIKOR ranking process. It is important to note that we have considered recall, precision, accuracy, F-score, AUC, and TNR as the benefit criteria, whereas FPR, FNR, and computation time have been considered as the cost criteria. Subsequently, the distance of the alternatives from the positive and negative ideal solutions was calculated. Finally, the Q_i values for the nine classification models for insider threat detection were estimated. Therefore, the classification models were ranked based on the VIKOR principle, and the results thus obtained are listed in Table 9.

As can be seen from the table, RF, which has the minimum values of S_i, R_i, and Q_i outranks the other eight classification models. Therefore, RF was selected as the best classification model for insider threat detection. Nevertheless, the ranks based on the above table were considered as the final results of the ranking and taken to be the basis of the validation processes.

6.2. *Results of the second layer (detection)*

Based on the result from the first layer, two best classification models were used for building the second layer of the proposed multilayer framework for detecting insider threats. According to the results given in Table 9, the RF and KNN classification methods were observed to outperform the other classification models. Therefore, the architecture of the hybrid detection system consists of two main components: (i) a MITD component based on RF, and (ii) an AITD component based on KNN. In the MITD component, a large number of individual DTs that operate as an ensemble in the RF are constructed. Each tree in the RF generates a class output, and the class with the most votes in the proposed hybrid detection could be known insider activities. This is detected in the first stage. Further, the unknown activities are double-checked by the AITD component, and the new data is treated by KNN while assuming that every data point in close proximity to each other corresponds to the same class. In other words, it classifies a new data point based on similarity.

6.2.1. *Known insider attacks*

Firstly, the performance of the proposed hybrid method was evaluated using the CERT dataset, where the model was trained using the known insider attacks and normal user activities. This implies that the models were already trained, and the patterns of the insider threat scenarios were already learned by the models. The results of the proposed model are given in Table 10.

Table 10 shows the performance of the three models in the case of the known attack detection. The highest values of the benefit

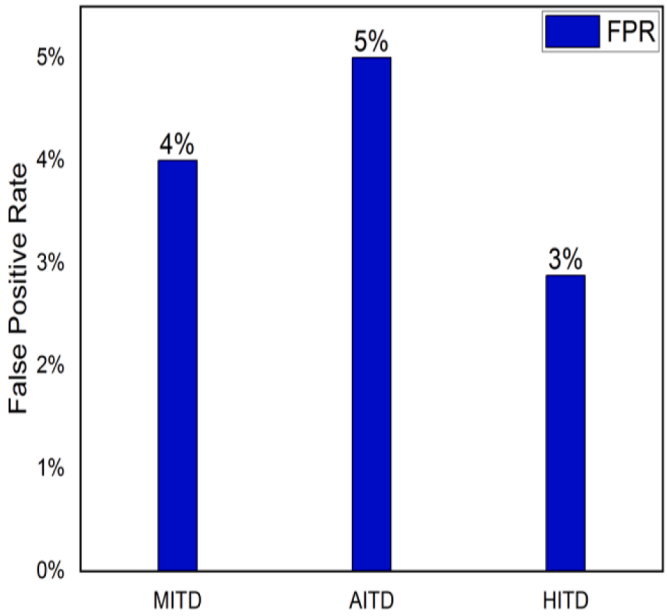


Fig. 5. FPR obtained using the three models using the first two datasets (90% training data and 10% testing data).

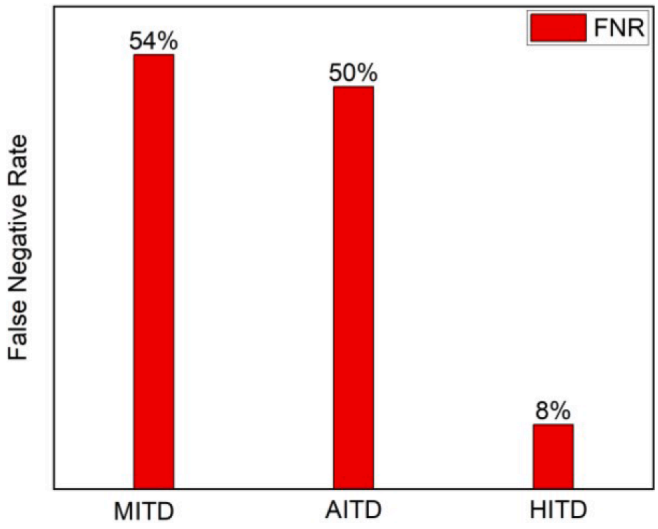


Fig. 6. FNR obtained using the three models using the first two datasets (90% training data and 10% testing data).

criteria, namely, recall, accuracy, precision, AUC, F-score, and TNR, indicate the best performance, whereas their lowest values indicate the worst performance. The lowest values of the cost criteria, namely, FNR, FPR, and computation time, indicate the best performance, whereas their highest values indicate the worst performance. For the known insider threat attacks, it can be seen that the MITD and HITD models provide a desirable result since these models have already been trained and have learned the attack patterns of the given insider threat scenarios. However, the MITD model was observed to exhibit a slightly higher precision, accuracy, F-score, and TNR, with values of 99%, 99.7%, 98.2%, and 99.9%, respectively. On the other hand, the HITD model exhibited high accuracy and AUC with low FNR. It is beneficial to detect the known insider attacks; however, as is commonly observed, the superior performance of the MITD model is observed only when detecting known insider threats. It is unable to detect attacks with new patterns. Detection of such patterns requires constant and quick updating of the model in order to keep up with the new, evolving threats, which imposes an extra workload on the security experts. New attacks can cause enormous damage to the system before a security expert generates a signature for it.

6.2.2. Unknown insider attacks

In order to evaluate the proposed HITD model in the case of unknown insider threat attacks, the CERT dataset was randomly divided into four different sets for performing the experiment. In the first two sets, 90% dataset was taken as the training set and 10% was taken as the testing set, whereas in the second two sets, 80% dataset was taken as the training set and 20% was taken as the testing set. Thereafter, the testing set was used to test the three models. The results of the MIDT, AITD, and HITD model obtained using the first

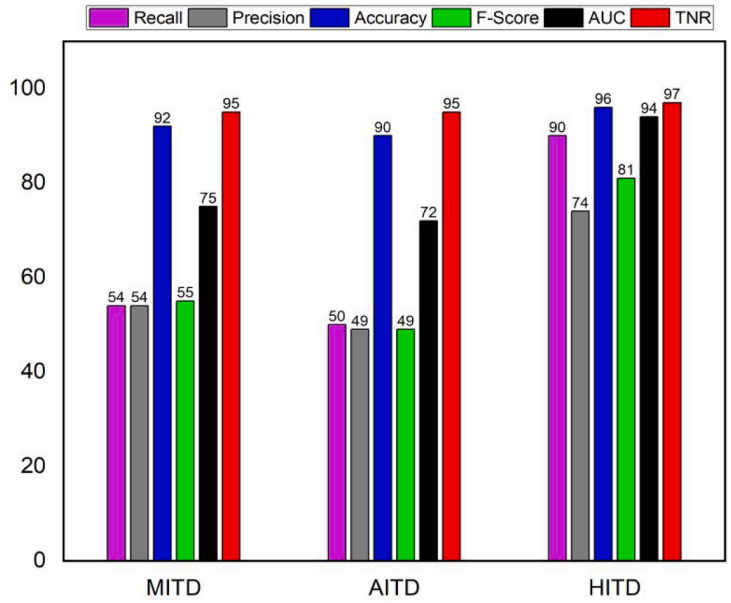


Fig. 7. Overall performance of the three models for the second two datasets (80% training data and 20% testing data).

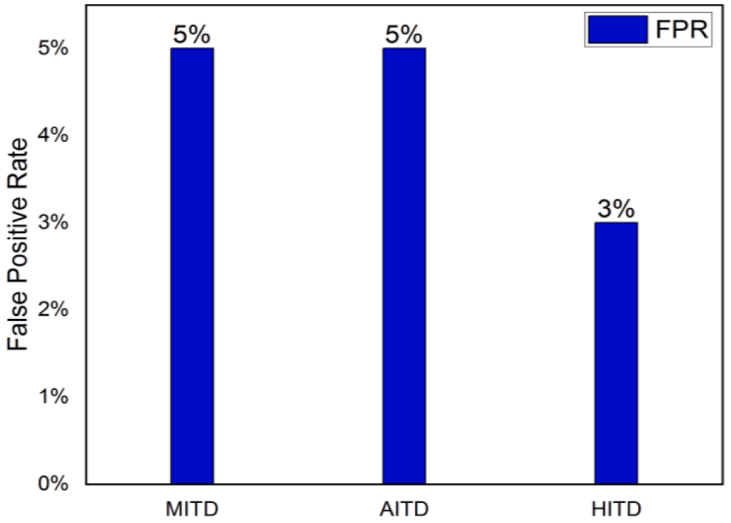


Fig. 8. FPR obtained from the three models using the second two datasets (80% training data and 20% testing data).

two datasets (90% training data and 10% testing data) are summarized in Table 11 and visualized in Fig. 4, Fig. 5, and Fig. 6, respectively. All the classification evaluation metrics that have been presented in Table 4 have been used for evaluating the proposed detection model.

Fig. 4 shows the results obtained from the three detection models in the case of unknown insider threat attacks. The values of the benefit criteria such as recall, accuracy, precision, AUC, F-score, and TNR, where the highest value (100%) indicates the best performance, and the lowest (0%) value indicates the worst performance have been plotted in the figure. The values of the performance metrics obtained for the HITD model are better compared to the single detection models, as can be seen from Fig. 4. The results indicate that the overall performance of the HITD model is better than the other two models. This is because of the two stages through which the data is processed and the power of the RF model in the first stage and KNN in the second stage of the HITD model. The RF model produces many classification trees and is used for training each tree from the training dataset and deals with the large dataset provided in the CERT standard datasets. Further, the KNN in the second stage efficiently detects the malicious insider class in a simple and precise manner. The values of the nine evaluation metrics obtained from the three models are given in Table 11. The accuracy of the MITD, AITD, and HITD models is observed to be 92%, 91%, and 97%, respectively, which clearly shows the effectiveness of the HITD

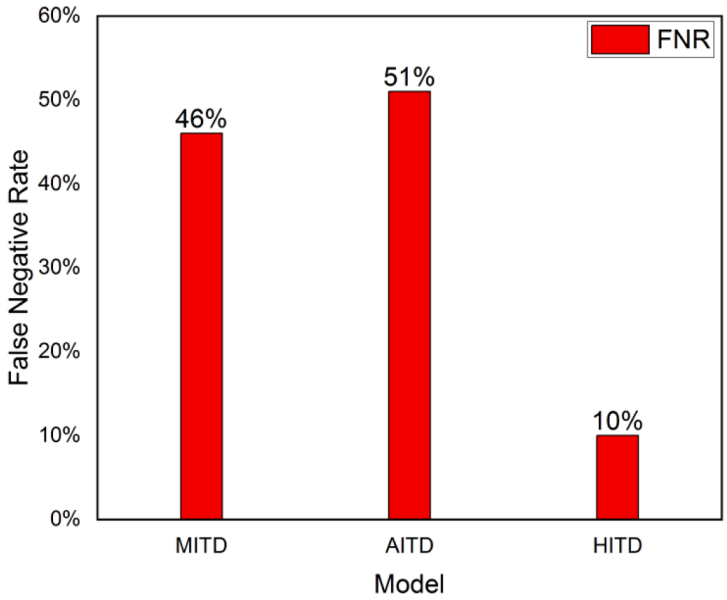


Fig. 9. FNR obtained from the three models using the second two datasets (80% training data and 20% testing data).

model in terms of accuracy as the results are approximately 5 to 6% higher than the single models. Similarly, in terms of the recall metric, the HITD model exhibits the highest value of 92%, which is 46% and 42% higher than that of MITD and AITD models, respectively. However, in terms of the precision and F-score metrics, the three models exhibit average results. This is due to the average of the false rates, especially with the single models. However, the HITD model exhibits 77% precision and 84% F-score, which are higher than the MITD model that exhibits 51% precision and 51% F-score, the AITD model, which exhibits 58% precision and 51% F-score. In addition, the HITD model outperforms the MITD and AITD models in terms of TNR (97.12%) and AUC (95%).

FPR is the frequency of a user who is incorrectly identified as a malicious insider. In other words, if a user is benign and the insider threat detection model considers them as a malicious insider, the FPR value will be high. FPR is one of the cost metrics, where the lowest value represents the best performance and the highest represents the worst performance. Fig. 5 shows a comparison of the FPR obtained for the insider threat detection using the single detection models MITD and AITD with our proposed HITD model. It can be observed from the figure that the HITD model minimizes the FPR by 2% as compared to the AITD model and by 1% compared to the MITD model.

FNR is the frequency of not detecting a malicious insider among the users. Fig. 6 shows a comparison of the FNR for insider threat detection obtained using the single insider threat detection models with our proposed HITD model. As can be seen from the figure, MITD and AITD insider threat single detection models alone exhibit higher FNR values whereas the proposed HITD model exhibits lower FNR. Results in Fig. 6 shows that HITD exhibits a low FNR of 7.6% compared to the MITD and AITD models, whereas a very high FNR of 51% for the AITD model and 46% for MITD model are observed. This implies that many malicious insiders were not detected when only the single detection models were employed for the detection.

Further, results obtained by applying the three models to the second two datasets, namely, the dataset divided into 80% training data and 20% testing data, are summarized in Table 12 and have been visualized in Fig. 7, Fig. 8, and Fig. 9, respectively.

Fig. 7 represents the recall, precision, accuracy, F-score, AUC, and TNR for all three models obtained by applying the three models on the second two datasets (80% training data and 20% testing data). In general, the six metrics are observed to be higher for the HITD model as compared to the MITD and AITD single models. Therefore, in terms of the benefit criteria, the HITD model still outperforms the single models when the training set reduces from 90% to 80%. The HITD achieves higher performance compared to the MITD and AITD models, in terms of the accuracy, precision, and recall values (96%, 74%, and 90%, respectively). Furthermore, the HITD also exhibits good F-score, TNR, and AUC values (81%, 97%, and 94%, respectively). The overall result in Fig. 7, corresponding to the use of the second two datasets, is slightly reduced compared to the result corresponding to the first two datasets owing to the nature of the ML classification models, where the larger the dataset used for training, the better are the results obtained by the model.

Similar to the first two datasets (90% training data and 10% testing data), the HITD model exhibits a low cost values in terms of FPR, compared to the single models, as shown in Fig. 8. In this case, the HITD model exhibits an FPR of 3.22%, whereas the MITD as well as AITD single models exhibit an FPR of 5%. This small increment in the cost value of FPR can be linked with the fact that the more the model trains, the less is the false alarm generated.

Fig. 9 present a comparison of the FNR obtained from the three models. The average FNR is observed to be higher for all three models for the case of using the second two datasets (80% training and 20% testing data) as compared to the case of using the first two

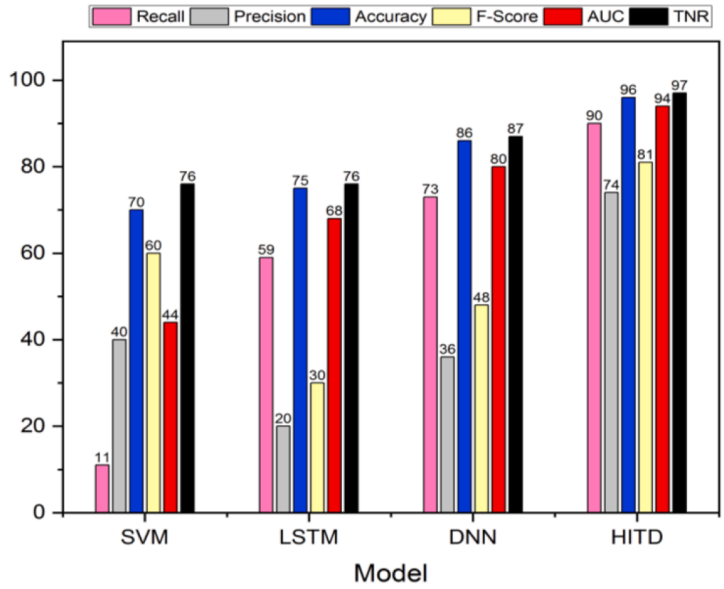


Fig. 10. Comparison of the overall performance of the HITD model with the SVM, LSTM, and DNN classification models.

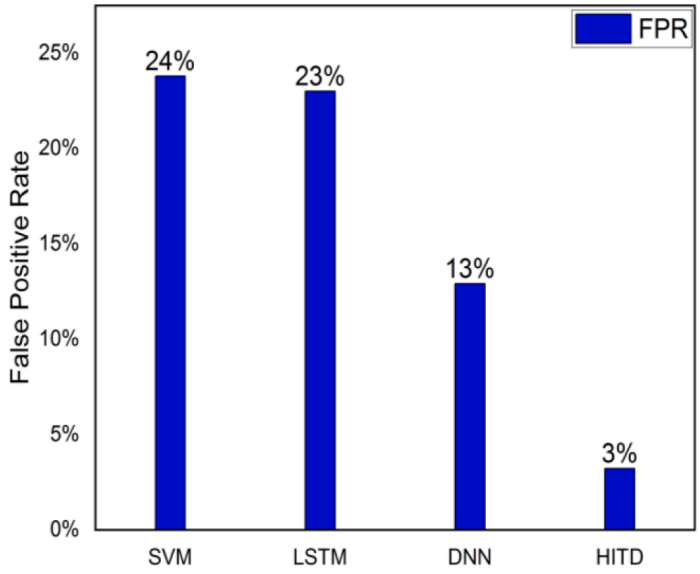


Fig. 11. Comparison of the FPR values corresponding to the different classifiers.

datasets (90% training data and 10% testing data), implying a high cost value in the former case. However, when considering the results obtained using only the second two datasets, the HITD model is still able to achieve better results compared to the single models.

6.3. Performance comparison

In order to assess the HITD model, its performance was compared with three other well-recognized and currently used classification models, namely, SVM, LSTM, and deep neural network (DNN). The nine metrics mentioned in Section 4 were used for the comparative analysis. Table 13 lists the results obtained from HITD and the other three models.

Fig. 10 plots the values of the different performance metrics corresponding to the proposed HITD model and three other popular ML classification models. The HITD model is observed to exhibit superior performance compared to the single ML classification models. This is because the former involves the integration of the two ML classification models for building a hybrid detection, which gains the advantage of the misuse as well anomaly detection techniques through the integration and is thus able to tackle the well-known as well

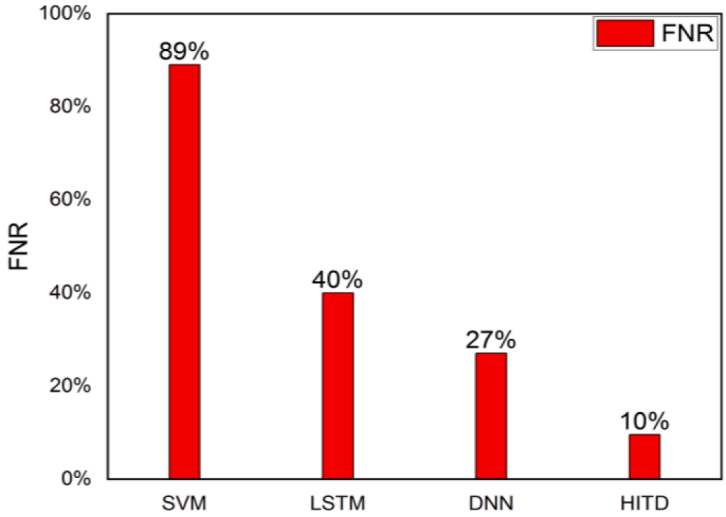


Fig. 12. Comparison of the FNR values corresponding to the different classifiers.

Table 4
Classification evaluation metrics [10].

Formula	Description
$Precision = \frac{TP}{TP+FP}$	A fraction of data entries of insider threats is labeled as a malicious insider, which indicates a truly malicious insider.
$TPR = \frac{TP}{TP+FN}$	TPR, which is also called Recall or Sensitivity, is defined as the fraction of insider threat data on malicious entries that are correctly classified.
$TNR = \frac{TN}{TN+FP}$	Also known as Specificity, TNR pertains to the ability of the detector to correctly identify data without malicious entries.
$FPR = \frac{FP}{FP+TN}$	It is calculated as the ratio between the number of benign insiders that are incorrectly classified as malicious insiders and the total number of benign insiders.
$FNR = \frac{FN}{FN+TP}$	False negative indicates the fraction of the missed malicious insider that the model failed to identify but classified as a benign insider.
$F-score = 2.\frac{Precision.Recall}{Precision+Recall}$	Also known as F-measure, it is the harmonic mean between sensitivity and precision.
$Accuracy = \frac{TN+TP}{TN+TP+FN+FP}$	It is the fraction of all negative and positive entries that are correctly classified. It can be considered the overall effectiveness of a classifier.
ROC or AUC	The ROC curve is a graphical visualization and used to plot the TPR against the FPR. AUC is computed by taking the area under the ROC curve as the ranges of the FPR from 0 to 1.
Computation Time	It pertains to the time required to complete the classification task of such a classifier.

Table 5
Encoding of the CERT features.

Feature	Possible Values
Activity Features	1, 2,3,4,5,6,7
Day Features	0, 2,3,4,5, 6
Time Features	1, 2,3,4 (…), 24
PC Features	unique PC ID
User_Id Features	1, 2,3,4, (…), 1000
User_Role Features	1, 2,3,4 (…), 42
User_Functional_Unit Features	1, 2,3,4,5,6
User_Department Features	1, 2,3,4,5,6,7

as unknown insider attacks efficiently. The RF model has been very effective in a wide variety of applications and can produce good solutions for very low false alarm outcomes. Similarly, the KNN model is an efficient approach that has significant strength for processing a massive dataset that normal working users produce during their daily work. The results shown in Fig. 10 shows that the HITD model exhibits the highest accuracy of 96% and the highest recall and precision of 90%, and 74% respectively. Furthermore, the HITD

Table 6
CERT insider scenarios.

Scenario	Description	Malicious Insiders
1	"User who did not previously use removable drives or work after hours begins logging in after hours, using a removable drive, and uploading data to wikileaks.org and leaves the organization shortly thereafter" [38].	""AJR0932", AAM0658", "BIH0745", "BDV0168", "BTL0226", "BLS0678", "DCH0843", "CAH0936", "EHD0584", "EHB0824", "FTM0406", "FMG0527", "HJB0742", "GHL0460", "JRG0207", "JMB0308", "KPC0073", "KLH0596", "LQC0479", "LJR0523", "MAS0025","MAR0955","MYD0978", "MCF0600", "RAB0589", "PPF0435", "RKD0604", "RGG0064", "WDD0366","TAP0551".
2	"User begins surfing job websites and soliciting employment from a competitor. Before leaving the company, they use a thumb drive (at markedly higher rates than their previous activity) to steal data" [38].	"ABC0174", "AAF0535", "CCL0068", "AKR0057", "CQW0652", "CEJ0109", "DRR0162", "DIB0285", "EGD0132", "EDB0714", "HBO0413", "FSC0601", "IJM0776", "HXL0968", "IUB0565", "IKR0401", "KRL0501", "JJM0203", "MDH0580", "LCC0819", NWT0098","MOS0047","PSF0133", "PNL0301", "RHL0992","RAR0725","TNM0961", "RMW0542", "XHW0498","VSS0154".
3	"System administrator becomes disgruntled and downloads a keylogger and uses a thumb drive to transfer it to his supervisor's machine. The next day, he uses the collected keylogs to log in as his supervisor and send out an alarming mass email, causing panic in the organization. He leaves the organization immediately" [38].	"BSS0369", "BBS0039", "CSC0217", "CCA0046", "JGT0221", "GTD0219", "JTM0223", "JLM0364", "MPM0220", "MSO0222"

Table 7
Normalized DM.

Classifier	Avg-Recall	Avg-Precision	Avg-Accuracy	Avg-F-Score	Avg-AUC	Avg-FPR	Avg-FNR	Avg-TNR	Computation Time
Logistic Regression	0.1298	0.1239	0.1147	0.1284	0.1264	0.0001	0.0054	0.1122	0.0643
Naïve Bayes	0.0503	0.0476	0.1020	0.0493	0.0843	0.0184	0.2893	0.1054	0.0000
Random Forests	0.1311	0.1239	0.1147	0.1284	0.1264	0.0000	0.0001	0.1122	0.0015
Decision Tree	0.1311	0.1239	0.1147	0.1284	0.1264	0.9648	0.0005	0.1122	0.0001
Support Vector Machines	0.0556	0.1051	0.1078	0.0726	0.0894	0.0024	0.2710	0.1122	0.6296
K-Nearest Neighbor	0.1232	0.1176	0.1136	0.1219	0.1226	0.0015	0.0301	0.1122	0.0012
Long short-term memory	0.1245	0.1176	0.1089	0.1219	0.0983	0.0052	0.2060	0.1111	0.1246
Gated Recurrent Units	0.1245	0.1164	0.1089	0.1206	0.0996	0.0076	0.1924	0.1100	0.1144
A multilayer perceptron	0.1298	0.1239	0.1147	0.1284	0.1264	0.0000	0.0053	0.1122	0.0642

Table 8
Entropy values and weights.

	Avg-Recall	Avg-Precision	Avg-Accuracy	Avg-F-Score	Avg-AUC	Avg-FPR	Avg-FNR	Avg-TNR	Computation Time
Entropy Value	0.9788	0.9878	0.9997	0.9837	0.9946	0.0900	0.6924	0.9999	0.5330
Entropy weight	0.0122	0.0070	0.0002	0.0094	0.0031	0.5230	0.1768	0.0000	0.2684

Table 9
Ranking results obtained using the VIKOR method.

Classifier	**S**	**R**	**Q**	**Rank**
Random Forests	0.01	0.0094	0	1
k-nearest neighbor	0.0303	0.0183	0.0280	2
A multilayer perceptron	0.0401	0.0273	0.0462	3
Logistic Regression	0.0402	0.0274	0.0464	4
Gated Recurrent Units	0.1826	0.1175	0.2703	5
Long short-term memory	0.1942	0.1259	0.2895	6
Naïve Bayes	0.2092	0.1768	0.3535	7
Support Vector Machines	0.4539	0.2684	0.6768	8
Decision Tree	0.5326	0.523	1	9

Table 10
Evaluation metrics obtained by applying the three models on the dataset in the case of the known attack scenario.

Model	Avg-Recall	Avg-Precision	Avg-Accuracy	Avg-F-Score	Avg-AUC	Avg-FPR	Avg-FNR	Avg-TNR	Computation Time
MITD	97%	**99.0%**	99.3%	**98.2%**	98.3%	**0.01%**	3.36%	**99.9%**	**3.663**
AITD	96%	97.1%	99.4%	96.0%	98.0%	0.3%	3.39%	**99%**	15. 081
HITD	**99%**	97.2%	**99.5%**	98.1%	**99.0%**	0.3%	**1%**	99.7%	3.828

Table 11
Evaluation metrics obtained by applying the three models to the first two datasets (90% training data and 10% testing data) in the case of the unknown attack scenario.

Model	Avg-Recall	Avg-Precision	Avg-Accuracy	Avg-F-Score	Avg-AUC	Avg-FPR	Avg-FNR	Avg-TNR	Computation Time
MITD	46%	58%	92%	51%	71%	3.89%	54.02%	96.61%	16. 93
AITD	50%	51%	91%	51%	73%	4.84%	49.72%	95.16%	12.808
HITD	**92%**	**77%**	**97%**	**84%**	**95%**	**2.88%**	**7.62%**	**97.12%**	**6.402**

Table 12
Evaluation metrics obtained by applying the three models to the second two datasets (80% training data and 20% testing data) in the case of the unknown attack scenario.

Model	Avg-Recall	Avg-Precision	Avg-Accuracy	Avg-F-Score	Avg-AUC	Avg-FPR	Avg-FNR	Avg-TNR	Computation Time
MITD	54%	54%	92%	55%	75%	4.64%	46.45%	95.36%	14. 457
AITD	50%	49%	90%	49%	72%	5.07%	50.63%	94.93%	12.677
HITD	**90%**	**74%**	**96%**	**81%**	**94%**	**3.22%**	**9.56%**	**96.78%**	**10.859**

Table 13
Results of the performance metrics obtained for the proposed HITD model and the other three popular classification models.

Model	Avg-Recall	Avg-Precision	Avg-Accuracy	Avg-F-Score	Avg-AUC	Avg-FPR	Avg-FNR	Avg-TNR	Computation Time
SVM	11%	40%%	70%	60%	44%	23.8%	89.10%	76%	**0.192**
LSTM	59%	20%	75%	30%	68	23.6%	40%	76%	19,354.965s
DNN	73%	36%	86%	48%	80%	12.9%	27%	87%	39.197s
HITD	**90%**	**74%**	**96%**	**81%**	**94%**	**3.22%**	**9.56%**	**97%**	10.859

model outperforms the SVM, LSTM, and DNN models in terms of the F-score, TNR, and AUC with 81%, 97%, and 94% respectively.

In the case of the cost criteria, similar to the benefit criteria, the HITD model exhibits good performance, with an FPR of 3%, compared to the other three classification models. Fig. 11 presents a plot of the FPR values for the HITD and the other three classification models.

Fig. 12 shows a comparison of the FNR value obtained from the HITD model with the corresponding values obtained from the widely used three classification models for insider threat detection. The proposed HITD has the least FNR of 9.6% compared to the other three classifiers.

All the above comparison results indicate that the overall performance of the HITD model is better than the single models. Thus, it can be concluded that the use of the HITD model can enable an accurate insider threat detection.

6.4. Comparison with other state-of-the-art methods

This section describes the performance comparison of the proposed HITD model with other recently developed state-of-the-art methods. The comparison results have been presented in terms of the nine evaluation metrics defined in Table 4 and used in the evaluation of the HITD model, as described in previous sections. Table 14 shows a comparison of the features and the values of the evaluation metrics obtained for the proposed HITD model with other methods involving hybrid and multilayer insider threat detection models. As can be seen from the table, the comparison of the HITD model with the other hybrid models reported in the literature shows that the proposed HITD model performs insider attack detection with a higher performance as compared to the other methods.

7. Conclusion

In this work, a new multilayer framework has been established to implement an intelligent insider threat detection method that overcomes the limitations of the currently available insider threat detection methods, where the classification models are first selected and benchmarked, and the detection methods are then developed for known and unknown insider threats. The development of the selection procedure in the hybrid model in this work was based on the integration of the entropy method and the VIKOR method. The final weight results obtained from the entropy model were shown, presenting the importance of the classification model criteria. Subsequently, the VIKOR approach was used for ranking and selecting the best classification model for insider threat detection according to the quantitative information of the measured criteria. For the second layer of the framework, a combination of the MITD and AITD methods was introduced, which takes advantage of the unique strengths of both these methods. In the proposed HITD model, the MITD model was used. The RF method was used for establishing the MITD model in order to break down the normal training data into minor subsets. Subsequently, the KNN method was used to develop the AITD model in each subset region. We have shown that the

ARTICLE IN PRESS

Table 14
Comparison of the proposed HITD model with other hybrid models developed for insider threat detection.

Refs.	Method	Dataset	Features				Results								
			Behavior of Biometrics	Cyber Features	Psychosocial Behaviors	Other Behaviors	Recall	Precision	Accuracy	F-Score	AUC	FPR	FNR	TNR	Computation Time
[39]	Local Outlier Factor	SY	X	√	X	X	91.4%	X	X	X	X	45%	X	X	481
[40]	Unsupervised KNN	SY	X	√	X	X	44.2%	47.5%	54%	44.9%	67%	X	X	X	X
[38]	Hidden Markov method	SY	X	√	X	X	55.9%	64.1%	71.1%	61.7%	0.79%	X	X	X	X
[41]	a new deep learning method (Dempster-Shafer theory)	SY	X	√	X	X	95.79%	X	95.47%	X	X	4.67%	X	X	X
[42]	user-centered ML-based anomaly behavior	SY	X	√	X	X	70%	9%	97.31%	X	X	2.61%	X	X	X
HITD	Hybrid ML	X	X	√	X	X	**92%**	**77%**	**97%**	**84%**	**95%**	**2.88%**	**7.62%**	**97.1%**	6.40216

*Abbreviations: Simulation (SE), Synthetic Dataset (SY), Own Environment (OE), Case Study (CS), Real Data (RD), values not provided or type of behaviors not analyzed (×), and values provided or behaviors are analyzed (√).

HITD model, which combines the RF and KNN methods, can achieve significantly better results compared to other techniques in terms of the nine important evaluation metrics considered in this work.

Future works

The proposed new multilayer framework provides promising performance result in comparison to the single detection models and most of the existing works that worked based on the analysis of cyber behaviors. However, in future research, the following points should be considered: Firstly, other types of insider threat detection classification models, that has been recently used and other evaluation criteria can be included in the selection process. Secondly, the problem of imbalanced dataset in the insider threats scenario can be consider by applying some of the imbalanced learning techniques such as synthetic minority oversampling technique (SMOTE) or adaptive synthetic technique (ADASYN).

CRediT authorship contribution statement

Mohammed Nasser Al-Mhiqani: Conceptualization, Methodology, Writing – original draft. **Rabiah Ahmad:** Conceptualization, Methodology, Writing – original draft. **Z. Zainal Abidin:** Investigation. **Karrar Hameed Abdulkareem:** Investigation, Validation, Formal analysis, Data curation. **Mazin Abed Mohammed:** Investigation, Validation, Formal analysis, Data curation. **Deepak Gupta:** Writing – review & editing. **K. Shankar:** Writing – review & editing.

CRediT authorship contribution statement

Mohammed Nasser Al-Mhiqani: Conceptualization, Methodology, Writing – original draft. **Rabiah Ahmad:** Conceptualization, Methodology, Writing – original draft. **Z. Zainal Abidin:** Investigation. **Karrar Hameed Abdulkareem:** Investigation, Validation, Formal analysis, Data curation. **Mazin Abed Mohammed:** Investigation, Validation, Formal analysis, Data curation. **Deepak Gupta:** Writing – review & editing. **K. Shankar:** Writing – review & editing.

Declaration of Competing Interest

The authors declare no conflict of interest.

References

[1] Magklaras GB, Furnell SM. Insider threat prediction tool: evaluating the probability of IT misuse. Comput Secur 2001;21(1):62–73. https://doi.org/10.1016/S0167-4048(02)00109-8.
[2] Dahmane M, Foucher S. Combating insider threats by user profiling from activity logging data. In: Proceedings of the 1st international conference on data intelligence and security, ICDIS; 2018. p. 194–9. https://doi.org/10.1109/ICDIS.2018.00039.
[3] Hashem Y, Takabi H, GhasemiGol M, Dantu R. Towards insider threat detection using psychophysiological signals. In: Proceedings of the 7th ACM CCS international workshop on managing insider security threats (MIST'15); 2015. p. 71–4. https://doi.org/10.1145/2808783.2808792.
[4] Collins ML, et al. Common sense guide to mitigating insider threats. 5th ed. PA, US: CERT, Software Engineering Institute, Carnegie Mellon University; 2016.
[5] Insider threat 2018 Report, (2018). Retrieved April 9, 2021, from https://crowdresearchpartners.com/wp-content/uploads/2017/07/Insider-Threat-Report-2018.pdf.
[6] 2020 Insider threat report. (2020). Retrieved April 9, 2021, from https://www.cybersecurity-insiders.com/wp-content/uploads/2019/11/2020-Insider-Threat-Report-Gurucul.pdf.
[7] Legg PA, Buckley O, Goldsmith M, Creese S. Automated insider threat detection system using user and role-based profile assessment. IEEE Syst J 2017;11(2):503–12. https://doi.org/10.1109/JSYST.2015.2438442.
[8] Gheyas IA, Abdallah AE. Detection and prediction of insider threats to cyber security : a systematic literature review and meta-analysis. Big Data Anal 2016;1(6):1–29. https://doi.org/10.1186/s41044-016-0006-0.
[9] Ullah F, Edwards M, Ramdhany R, Chitchyan R, Babar MA, Rashid A. Data exfiltration: a review of external attack vectors and countermeasures. J Netw Comput Appl 2018;101:18–54. https://doi.org/10.1016/j.jnca.2017.10.016. Jan.
[10] Al-Mhiqani MN, et al. A review of insider threat detection: classification, machine learning techniques, datasets, open challenges, and recommendations. Appl Sci 2020;10(15). https://doi.org/10.3390/app10155208.
[11] Noever, D. Classifier suites for insider threat detection. 2019, *arXiv preprint arXiv:1901.10948*.
[12] Software Engineering Institute, Carnegie Mellon University. [n. d.]. Insider Threat Test Dataset. Retrieved January 27, 2021, from https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=508099.
[13] Beaver JM, Borges-Hink RC, Buckner MA. An Evaluation of Machine Learning Methods to Detect Malicious SCADA Communications. In: Proceedings of the 12th international conference on machine learning and applications. 2; 2013. p. 54–9. https://doi.org/10.1109/ICMLA.2013.105.
[14] Hall AJ, Pitropakis N, Buchanan WJ, Moradpoor N. Predicting malicious insider threat scenarios using organizational data and a heterogeneous stack-classifier. In: Proceedings of the IEEE international conference on big data; 2018. p. 5034–9. https://doi.org/10.1109/BigData.2018.8621922. Big Data 2018.
[15] Sheykhkanloo NM, Hall A. Insider threat detection using supervised machine learning algorithms on an extremely imbalanced dataset. Int J Cyber Warf. Terror. 2020;10(2):1–26. https://doi.org/10.4018/IJCWT.2020040101.
[16] Maloof MA, Stephens GD. Elicit : a system for detecting insiders who violate need-to-know. In: Proceedings of the international workshop on recent advances in intrusion detection; 2007. p. 146–66.
[17] Punithavathani DS, Sujatha K, Jain JM. Surveillance of anomaly and misuse in critical networks to counter insider threats using computational intelligence. Cluster Comput 2015;18(1):435–51. https://doi.org/10.1007/s10586-014-0403-y.
[18] Michael A, Eloff J. Discovering 'insider IT sabotage' based on human behaviour. Inf Comput Secur 2020;28(4):575–89. https://doi.org/10.1108/ICS-12-2019-0141.
[19] Ferreira P, Le DC, Zincir-Heywood N. Exploring feature normalization and temporal information for machine learning based insider threat detection. In: Proceedings of the 15th international conference on network and service management, CNSM; 2019. p. 1–7. https://doi.org/10.23919/CNSM46954.2019.9012708.
[20] Al-Mhiqani MN, et al. A new taxonomy of insider threats : an initial step in understanding authorised attack. Int J Inf Syst Manag 2018;1(4):343–59.

[21] Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value trade-offs. Cambridge University Press; 1993.
[22] Zou ZH SJ, Yun Y. Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment. J Environ Sci 2006;18(5):1020–3.
[23] Li X, Wang K, Liuz L, Xin J, Yang H, Gao C. Application of the entropy weight and TOPSIS method in safety evaluation of coal mines. Proced Eng 2011;26(2011): 2085–91. https://doi.org/10.1016/j.proeng.2011.11.2410.
[24] Pamučar D, Stević Ž, Sremac S. A new model for determiningweight coefficients of criteria in MCDM models: full Consistency Method (FUCOM. Symmetry 2018; 10(9):1–22. https://doi.org/10.3390/sym10090393 (Basel).
[25] Khan S, Gani A, Wahab AWA, Singh PK. Feature selection of denial-of-service attacks using entropy and granular computing. Arab J Sci Eng 2018;43(2): 499–508. https://doi.org/10.1007/s13369-017-2634-8.
[26] Siksnelyte-Butkiene I, Zavadskas EK, Streimikiene D. Multi-criteria decision-making (MCDM) for the assessment of renewable energy technologies in a household: a review. Energies 2020;13(5):1–22. https://doi.org/10.3390/en13051164.
[27] Alsalem MA, et al. Systematic review of an automated multiclass detection and classification system for acute leukaemia in terms of evaluation and benchmarking, open challenges, issues and methodological aspects. J Med Syst 2018;42(11). https://doi.org/10.1007/s10916-018-1064-9.
[28] Opricovic S. Compromise solution by MCDM methods : a comparative analysis of VIKOR and TOPSIS. Eur J Oper Res 2004;156(2):445–55. https://doi.org/10.1016/S0377-2217(03)00020-1.
[29] S O, Tzeng GH. Extended VIKOR method in comparison with outranking methods. Eur J Oper Res 2007;178(2):514–29. https://doi.org/10.1016/j.ejor.2006.01.020.
[30] Shemshadi A, Shirazi H, Toreihi M, Tarokh MJ. Expert systems with applications a fuzzy VIKOR method for supplier selection based on entropy measure for objective weighting. Expert Syst Appl 2011;38(10):12160–7. https://doi.org/10.1016/j.eswa.2011.03.027.
[31] Resende PAA, Drummond AC. A survey of random forest based methods for intrusion detection systems. ACM Comput Surv. 2018;51(3). https://doi.org/10.1145/3178582.
[32] Rajaguru H, Sannasi Chakravarthy SR. Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer. Asian Pac J Cancer Prev 2019;20(12):3777–81. https://doi.org/10.31557/APJCP.2019.20.12.3777.
[33] Yas QM, Zadain AA, Zaidan BB, Lakulu MB, Rahmatullah B. Towards on develop a framework for the evaluation and benchmarking of skin detectors based on artificial intelligent models using multi-criteria decision-making techniques. Int J Pattern Recognit Artif Intell 2017;31(3). https://doi.org/10.1142/S0218001417590029.
[34] Alsalem MA, et al. Multiclass benchmarking framework for automated acute leukaemia detection and classification based on BWM and group-VIKOR. J Med Syst 2019;43(212). https://doi.org/10.1007/s10916-019-1338-x.
[35] Wu J, Sun J, Liang L, Zha Y. Expert systems with applications determination of weights for ultimate cross efficiency using shannon entropy. Expert Syst Appl 2011;38(5):5162–5. https://doi.org/10.1016/j.eswa.2010.10.046.
[36] Triantaphyllou E, Baig K. The impact of aggregating benefit and cost criteria in four MCDA methods. IEEE Trans Eng Manag 2005;52(2):213–26.
[37] Bisong E. Google colaboratory bt-building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners. Springer; 2019.
[38] Lo O, Buchanan WJ, Griffiths P, Macfarlane R. Distance measurement methods for improved insider threat detection. Secur Commun Netw 2018;2018 (5906368):1–18. https://doi.org/10.1155/2018/5906368. no. Art.
[39] Nicolaou A, Shiaeles S, Savage N. Mitigating insider threats using bio-inspired models. Appl Sci 2020;10(15). https://doi.org/10.3390/app10155046.
[40] Böse B, Avasarala B, Tirthapura S, Chung YY, Steiner D. Detecting insider threats using RADISH: a system for real-time anomaly detection in heterogeneous data streams. IEEE Syst J 2017;11(2):471–82. https://doi.org/10.1109/JSYST.2016.2558507.
[41] Tian Z, et al. Deep learning and dempster-shafer theory based insider threat detection. Mob Netw Appl. 2020:1–10. https://doi.org/10.1007/s11036-020-01656-7.
[42] Le DC, Zincir-Heywood N. Exploring anomalous behaviour detection and classification for insider threat identification. Int J Netw Manag 2020;(Special Issue (March)):1–19. https://doi.org/10.1002/nem.2109.

Mohammed Nasser Al-Mhiqani received his BSc. In Computer Science (Computer Networking) in 2014, and MSc in Computer Science from Universiti Teknikal Malaysia Melaka (UTeM) in 2015. Currently he is a Ph.D. candidate at UTeM. His-research interests Cyber Security, Cyber Physical System Security, and Insider Threats in Cyber Security, Machine Learning, Deep learning, and image processing.

Rabiah Ahmad received BSc in Computer Science from Universiti Teknologi Malaysia. She obtained MSc in Information Security from Royal Holloway University of London, and her PhD in Information Studies at the University of Sheffield. She is a Professor of Information Security since 2015 at UTeM. Her research interest includes information security management, security architecture, healthcare system security and insider threats.

Z. Zainal Abidin received BSc of Information Technology from University of Canberra, 2002. She completed her MSc. in Quantitative Sciences (2004), MSc. in Computer Networking (2008) and PhD in I.T. and Quantitative Sciences (2016) from Universiti Teknologi MARA. She served as a lecturer at UTeM (2009 – present). Her research interests include computer networking, information security, biometrics, steganography and image processing.

Karrar Hameed Abdulkareem received the B.S. degree in computer science from the University of Technology, Iraq, in 2007, and the M.S. degree in computer science from UTeM, Malaysia, in 2016. He completed his Ph.D. in computer science and information technology from Universiti Tun Hussein Onn Malaysia (UTHM), 2021. His research interests include multi-criteria decision making, image dehazing, and computer security.

Mazin Abed Mohammed received his B.Sc. degree in Computer Science from the University of Anbar, Iraq in 2008. He obtained his M.Sc. degree in Information Technology from UNITEN, Malaysia in 2011. He obtained his PhD in Information Technology from UTeM, Malaysia in 2018. His research interests include Artificial Intelligence, Biomedical Computing, and Optimization.

Deepak Gupta received a B.Tech. degree in 2006 from the Guru Gobind Singh Indraprastha University, Delhi, India. He received M.E. degree in 2010 from Delhi Technological University, India and Ph. D. degree in 2017 from Dr. APJ Abdul Kalam Technical University (AKTU), Lucknow, India. His research interests include Intelligent Data Analysis, Nature-Inspired Computing, Machine Learning and Soft Computing.

K. Shankar received the M.C.A., M.Phil., and Ph.D. degrees in computer science from Alagappa University, Karaikudi, India. Currently a Postdoctoral Fellow with the Department of Computer Applications, Alagappa University, Karaikudi, India. His-current research interests include Healthcare applications, Secret Image Sharing Scheme, Digital Image Security, Cryptography, Internet of Things, and Optimization algorithms.

Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-021-01471-7

ORIGINAL ARTICLE

An intelligent model based on integrated inverse document frequency and multinomial Naive Bayes for current affairs news categorisation

Sachin Kumar[1] · Aditya Sharma[1] · B Kartheek Reddy[1] · Shreyas Sachan[1] · Vaibhav Jain[1] · Jagvinder Singh[2]

Received: 6 December 2020 / Revised: 20 September 2021 / Accepted: 19 October 2021
© The Author(s) under exclusive licence to The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2021

Abstract Digital technologies, their product and services have empowered the masses to generate information at a faster pace. Digital technologies based information sharing platforms such as news websites and social media platforms such as Facebook, Twitter, Instagram, What's app etc have flooded the information space due to the easy generation of information and dissemination to the masses instantly. Information classification has been an important task, especially in newspapers and media organisations. In another area also, information or text classification has an important role to play so that important and vital information can be classified based on the already predefined categories. In journalism, editors and resources persons were allocated the task to recognise and classify the news stories so that they can be placed in the predefined categories of economy and business news, political news, social news, editorial section, education and career, and sports information etc. Nowadays the process of classification and segregation of textual information has become challenging due to the flow of diverse, vast information. Additionally, the pace of information and its updates, access and competition among the media House have made it more challenging. Hence automated and intelligent tools which can classify the information and text accurately and efficiently is needed to reduces human efforts, time and increase productivity. This paper presents an intelligent, efficient and robust intelligent machine learning model based on Multinomial Naive Bayes(MNB) to classify the current affairs news stories. The proposed Inverse Document Frequency(IDF) integrated MNB model achieves classification accuracy of 87.22 per cent. The experiment results are also compared with other machine learning models such as Logistics Regression(LR), Support Vector Machine(SVM), K-Nearest Neighbours(KNN) and Random forest(RF). The results demonstrate that the presented model is better in term of accuracy and may be deployed in real world information classification and media domain to improve the productivity, efficiency of the current affairs news classification process.

Keywords News Articles · Classification · Intelligent Methods · Machine Learning · Support Vector Machine · Multinomial Naive Bayes · Inverse Document Frequency(IDF)

✉ Sachin Kumar
skumar@cic.du.ac.in

Aditya Sharma
aatom098@gmail.com

B Kartheek Reddy
bkartheekreddy@gmail.com

Shreyas Sachan
shreyas.sachan@gmail.com

Vaibhav Jain
vabsweb@gmail.com

Jagvinder Singh
jagvinder.singh@gmail.com

[1] Cluster Innovation Centre, University of University, Delhi, India

[2] Dept of Management, Delhi Technological University, Delhi, India

Published online: 07 November 2021

1 Introduction

Present-day technological advancements, especially digital technologies have transformed the way information is generated, transmitted and disseminated (Van Veldhoven and Vanthienen 2021). Digital technologies and software tools developed have immense power and scale to circulate information all over the world (Ashuri 2016). At present, most news and media organisations responsible for sharing information have maintained their online web and app platform to reach the masses in a very efficient and effective manner (Bail 2017). The news and information collection, processing, publication and dissemination has become very challenging with new technological tools being used and upgraded in each of the phases of the news life cycle (Canito et al. 2018). These phases of the news life cycle require automated and intelligent methods of information processing to classify the information and news stories. In the present day competitive world, it requires fast, efficient, effective, intelligent solutions to be deployed in each phase so that the news reaches the masses in a very fast manner and the classification process is automated (Hogenboom et al. 2016). The digital news articles platforms and the media industry is growing and there is a need to create an intelligent system that automates hectic and time-consuming tasks such as manual news categorisation after reading the articles at the primary level (Thomson et al. 2020). For example, news articles come for publication, media organisations need an expert to classify them after reading so that they can be published in the appropriate section of newspapers or websites. This is a time-consuming process. Artificial Intelligence with Natural Language Processing(NLP) research and development can reduce and accurately classify the information automatically (Marconi 2020). Such practical issues come under text processing and categorisation in traditional models of classification works. For example, classification of the news related text in natural languages such as English, Hindi or regional languages into a predefined set of categories such as politics, business, entertainment, health, sports etc (Medagoda 2016). The mentioned classification task requires a lot of human effort of reading the article or asking the author itself for the classification but accuracy needs to be ensured (Daud et al. 2017). The large volume of the articles, their large quality and the complexity of information and facts necessitates the efficient system automating the classification process for the news agencies and organisations. News article classification, or in broader terms the text classification, have changed from core mathematical models to statistical and machine learning (Hu et al. 2017). Most of the research work in the fields of text classification in news articles have been short of producing benchmark level accuracy of the text classification. Some of the work produced, used only headlines and tags of the news stories leaving the body of the news, which did not give the good accuracy of the news story classification (Kanan and Fox 2016). Such models suffered from the loss of complex information ingrained in the natural language and did not result in the best accuracy for the models (Salminen et al. 2019). Additionally, they suffered from over fitting and did not make them eligible to be deployed in a real-world environment in practical situations (Singh et al. 2019). This research work creates an intelligent model to facilitate news article categorisation which will aid the users, writers, editors and news organisations and other stakeholders to accurately label the news articles with minimal efforts in terms of human and computational resources Miller and (Oswalt 2017). The proposed model reads the whole article with title and body of news and tags and extracts meaningful relationships between words of the news articles to label the article into one of the following categories: Business and Economy, Education and Career, Entertainment, Food and Health, International, Politics and Governance, Science and Technology, and Sports. The research work of the paper proposes the following three improvements and findings.

1 It proposes a model which takes news titles and news stories for classification of the news articles which is an improvement over previous work.
2 It proposed a machine learning-based model which is very efficient and effective in news categorisation.
3 It performs classification and comparison on newly developed data set of news stories.

This paper is organised in the following manner. Section 1 introduces the background of the research problem and its context. Section 2 highlights the recent work done in the domain and research gaps. Section 3 describes the mathematical models, their foundations with evaluation criteria. Section 4 explains the methodology of the experimentation and proposed model. Section 5 discusses the results and provides their analyses in the comparison to the existing work. Section 6 concludes the paper.

2 Literature survey

Text processing is one of the important and challenging topics for the researchers. Getting insights from diverse texts after their processing have several useful applications such as sentiment analysis, information categorisation, user feedback, language generation, hate speech and abusive speech detection, misinformation campaign detection etc. (Ur-Rahman and Harding 2012). With such diverse applications, several methods of text processing and

classification have been developed. The developed models have their foundations on mathematical and statistical inferences. Recent developments in intelligent models based on Artificial Intelligence/machine learning have also been used. Machine learning based methods have been applied in several diverse fields such as energy consumption predictions (Kumar et al. 2019, 2018b), pattern recognition (Kiranyaz et al. 2014), occupancy detection (Kumar et al. 2020), health informatics and governance (Ravì et al. 2016; Kumar et al. 2018c), fake news detection (Conroy et al. 2015), education and learning Lykourentzou et al. (2009), object recognition (Ramík et al. 2014), financial predictions (Ahmed et al. 2016; Lin et al. 2011), intelligent transportation Kumar et al. 2018a, environment and weather prediction (Rasouli et al. 2012). Machine learning approaches involve basic tasks such as classification, regression and clustering (Witten et al. 2005). Several studies have been conducted in text classification and processing which have used a variety of methods (Fong et al. 2013). Some of research work have developed approaches based on rigid mathematical understanding and decision support systems (Pröllochs et al. 2016). The rigid mathematical models suffer from low accuracy in a dynamic environment. Other statistical approaches have their own limitations (Paragios et al. 2006). Machine learning methods have produced some of the relevant models of some of the natural language processing work. Boumans et. al. worked on content analysis for digital journalism (Boumans and Trilling 2016). The work provided an automated content analysis for online news and print news. Alhothali et al. (2015) applied affect control theory to analyze readers'reaction towards news articles. Text classification and processing have been applied in health domain as well for developing personalised health systems. Mukwazvure et al. (2015) even provided a hybrid approach to analyze sentiments of news stories and their comments. Some have used hierarchical classification approaches for the classification of online news articles (Li et al. 2016) with heading only. Valdivia et. al. (2017) applied machine learning methods for the sentiments analysis and text analysis of trip advisor website. Several techniques have been proposed for the improvement in the accuracy of the machine learning text processing methods (Carstens and Toni 2017; Onan et al. 2016). Some researchers have also applied machine learning approaches on the news data collection and processing. Wen-Lin Hsu et. al. (1999) proposed various intelligent and conducted experiments on NETNEWS data set. The proposed were based on vector space paradigm and categorised the news articles quite efficiently. Carlos et.al. (2014) presented a study of the new article's life cycle using data from large international news networks and social media. The data set was generated had more than 3,000,000 website visits and 200,000 social media reactions. They showed that it is quite possible to accurately predict and model the overall traffic articles received in their life cycle by observing the first 10–20 minutes of social media reactions. Michela et.al. (2016) demonstrated a quantitative cross-platform analysis of public discourse and news consumption on online social media on the Italian referendum. They found that users from well-separated communities tend to restrict their attention to a specific set of articles on Facebook and Twitter. From Hakim et al. (2014), proposed a frequency-inverse document frequency (TF-IDF) framework which was weighting for classification of news articles into sports, business and science and technology categories. In their approaches, for assigning weights to the words of articles, the TF-IDF scheme was used, for which each news article is transformed into a vector of weights. For the training of the system, a total of 300 categorized news articles were collected and were tested on 60 randomly extracted articles. In their work, Yaocheng Gui et. al. (2012). used a different approach to classify news articles. They used a hierarchical text classification approach to classify text data documents into a pre-generated hierarchy of news article categories. Their work utilizes named entities as features for the classification task of news articles into a pre-generated hierarchy of articles about international relations. The experiment results analysis reported improved performance of classification for news articles using the named entities. After going through the recent work of article classification, it has been observed that the works have majorly used only the title/headline of news stories and some of the work have used just tags and body of news article alone. In the present work, we propose models based on machine learning with integrated text processing techniques for news article classification in which the input data includes three dimensions such as title of news articles, news article body and their tags/keywords. Additionally in our work, the number of news articles categories to be classified have been raised from 3 to 8 with newly generated data set.

3 Foundational approaches

3.1 Machine learning approaches

News classification problem is text processing and classification problem and can be solved using supervised machine learning approaches. Supervised learning problems deal with labeled data sets paired with an input object and the desired output value (Sen et al. 2020). It is processed by the machine learning approaches with some other feature engineering approaches to do the text categorisation. Following are the mathematical foundations of

the machine learning methods used in the research work. Logistic Regression (LR) Kurt et al. (2008) is considered one of the simplest machine learning methods, LR tries to predict, quantitatively, probability of a classification outcome having only two values. The prediction is based on the use of one or more predictors. These predictors may be of type numerical and categorical. A logistic model (P) is represented by as follows

$$P = h(x) = 1/(1 + e(-z(x))) \quad (1)$$

where $z = b_0 + b_1 * x_1 + b_2 * x_2 + ... + b_p * x_p$. is a linear model. The function $1/(1 + e(-z))$ is often called the "logistic"or"sigmoid"function which produces a logistic curve, which is limited to values from 0 and 1 to be interpreted as the P probability. To classify the news articles, our goal is to search for a value of $z(x)$ in such as way that the probability $P(y = 1|x) = h(x)$ is large when x represents news articles belonging to the"1"class and small when x belongs to the"0" class. Multinomial Naive Bayes (MNB) (Kibriya et al. 2004) is fundamentally based on naive bayes approaches and based on bayes' theorem. It has the independence assumptions between the features in order to predict the category of a given category. Naive Bayes method is referred to as the model of posterior probability. This method updates value of prior belief of an event if new information comes. The result of the event is the probability of the /categories occurring given the new data Korb and Nicholson 2010).

$$P(class/attributes) = \frac{P(class) * P(attributes/class)}{P(attributes)}$$

where P(class/attributes) is called posterior probability, P(class) is called the class prior probability, P(attributes/class) is called the likelihood and P(attributes) is called the predictor prior probability. MNB comes under probabilistic classifiers and calculates the probability of each category of news articles using Bayes theorem (McCallum et al. 1998). The news stories category with the highest probability will be output. So, the model calculates the probability that news article d belongs to a class c which is given as follows.

$$P(c/d)P(c) * k = 1ndP(tk/c) \quad (2)$$

In the above equation, $P(tk/c)$ is the called the conditional probability of the term tk of class news article c. We can interpret the probability $P(tk/c)$ as a measurement which is , how much evidence tk contributes for the class determination that c class is the correct class of d news article. $P(c)$is the prior probability of a news article occurring in class c. The term MNB simply explains to us that each $p(fi/c)$ represents the multinomial distribution, rather than some other distribution. K-Nearest Neighbours (KNN) Zahid et al. (2001) very simple and essential algorithms. It is a non-parametric method. More importantly, it does not assume something about the distribution of news article data. It is widely applied in real-life scenarios. The algorithm locates clusters of similar type in the dataset. If any unclassified news story is given, KNN assigns it to a cluster or group by observing what group its nearest neighbors belong to. KNN has an important parameter like k which defines the number of neighbors considered. For example, fitting a KNN model with parameter value $k = 2$, the two closest neighbors are taken to smooth the estimate at a given news article. Support Vector Machine (SVM) (Suykens and Vandewalle 1999) is one of the most important and widely used ML methods. SVM can be applied for many types of problems such as classification and regression (Tong and Koller 2001). In SVM, each news article item is plotted as a point in an n-dimensional space. In the space n is the number of attributes/features the data set and in this case news article. They are plotted the value of each feature being the value of a particular coordinate on the space (Noble 2006). News article classification is performed by finding the hyperplane that differentiates the two news article classes very well and in an optimal way (Joachims 1998). It is also robust to outliers type of news articles. Support vectors are simply the coordinates of individual observation. SVM is a frontier that best segregates the two news article classes using hyper-planes and maximizing the distances between the nearest news article data point. This distance is called Margin. In case a linear hyperplane is not sufficient, SVM has a technique namely kernel trick. Kernel functions take low dimensional input space and transform it into a higher-dimensional space. it converts not separable problems to separable problems, these functions are called kernels. Following is the mathematical representation of SVM. A hyperplane in an N dimensional feature space represented by the following equation as

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{w} + b = \sum_{i=1}^{n} x_i w_i + b = 0$$

We divide the above equation by $||\mathbf{w}||$ and obtain the following equation.

$$\frac{\mathbf{x}^T\mathbf{w}}{||\mathbf{w}||} = P_{\mathbf{w}}(\mathbf{x}) = -\frac{b}{||\mathbf{w}||}$$

This equation indicates that the projection of any point $\mathbf{x}$ on plane onto the vector $\mathbf{w}$ is always $-b/||\mathbf{w}||$. More simply, $\mathbf{w}$ represents the normal direction of the hyperplane, and $|b|/||\mathbf{w}||$ is the distance from the origin point to the hyperplane in space. Here the important thing to note is that the hyperplane is not unique from the equation. $cf(\mathbf{x}) = 0$ can represent N number of planes for varying

values of c. The N-dimensional space is divided into two regions by the plane by a mapping function $y = sign(f(\mathbf{x})) \in \{1, -1\}$,

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{w} + b = \begin{cases} > 0, & y = sign(f(\mathbf{x})) = 1,\ \mathbf{x} \in P \\ < 0, & y = sign(f(\mathbf{x})) = -1,\ \mathbf{x} \in N \end{cases}$$

Here any point in hyperspace $\mathbf{x} \in P$ on the positive side of the plane is mapped to value 1, while any point in hyperspace $\mathbf{x} \in N$ on the negative side is mapped to -1. A point $\mathbf{x}$ of unknown class will be classified to P if $f(\mathbf{x}) > 0$, or N if $f(\mathbf{x}) < 0$. For a decision type of hyper-plane $\mathbf{x}^T\mathbf{w} + b = 0$ to separate the two classes of points representing the news articles $P = \{(\mathbf{x}_i, 1)\}$ and $N = \{(\mathbf{x}_i, -1)\}$, it has to satisfy

$$y_i(\mathbf{x}_i^T\mathbf{w} + b) \geq 0$$

for both $\mathbf{x}_i \in P$ and $\mathbf{x}_i \in N$. Among all such hyper planes satisfying this condition, there is need to find the optimal one H_0 that separates the two classes with the maximal margin. Hence, the optimal plane must be in the middle of the two different classes of classification problem. This makes the distance from the dividing plane to the closest point on either side of news articles or data points equal. Two additional separate hyper planes are defined and plotted to set the boundary to the maximum of the middle point with following conditions.

$$\mathbf{x}_i^T\mathbf{w} + b = y_i$$

And the following holds for all support vectors:

$$b = y_i - \mathbf{x}_i^T\mathbf{w} = y_i - \sum_{j=1}^{m} \alpha_j y_j (\mathbf{x}_i^T\mathbf{x}_j)\,.$$

Now the problem of finding the optimal decision plane in terms of $\mathbf{w}$ and b can be formulated as:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\mathbf{w}^T\mathbf{w} = \frac{1}{2}||\mathbf{w}||^2 \quad \text{(objective function)} \\ &\text{subject to } y_i(\mathbf{x}_i^T\mathbf{w} + b) \geq 1, \text{ or } 1 - y_i(\mathbf{x}_i^T\mathbf{w} + b) \leq 0, \quad (i = 1, \cdots, m) \end{aligned} \tag{3}$$

The solution of the above equation gives the maximum margin of the points so that the news articles can be determined to be categorized in one of the categories.

4 Research methodology

4.1 Data collection

The experiment is designed with the following phases. First the experiment has collected the news articles from the three sources namely The Hindu, NDTV and Indian Express website. There are 1800 articles. It is to be noted that majorly, previous research studies have mostly used the headlines/titles of the news articles for the classification. There was no big dataset of news articles, along with the categories/classes , tags/keywords and content of the news articles/body. It has been a disadvantage of using only the article's headline for classification . The reason is that the headlines are not always honest in their interpretation for the sake of catching attention. Hence, the experiment tried to remove that or reduce that loop by collecting the large amount of data of new articles tagged with their categories and body of the news articles and headlines for our categorization problem. Three diverse sources have made the data more practical and reduces the chances of overfitting. Collected articles belongs to the following categories:-

1 Business and Economy
2 Education and Career
3 Entertainment
4 Food and Health
5 International
6 Politics and Governance
7 Science and Technology
8 Sports

These 8 categories cover the majority of the articles that are read in our daily life and the articles that do not fall into any of these categories were removed. The data set can be accessed on the following URL:https://github.com/sachinblessed95/Current-Affairs-News-Categorisation-Data-Set.

4.2 Data processing

Obtained newspaper articles are rich in the content. But for better accuracy, more concise information and feature rich content is required. To get the relevant and feature rich information, data cleaning is required. Data cleaning is the first step where all the text is converted into lowercase for simplicity and uniformity followed by tokenization by breaking the sentences into words and removing the stop-words. Alphanumeric characters, numbers and punctuation are also removed followed by the redundancy of words (Chava et al. 2018). However, there are still some words that are left out after these steps of data processing which are redundant for the learning model for differentiating the news articles. For example, consider the sentence"He is playing Football"and"He plays Football", Both the words "playing"and"plays"have similar meanings and increasing the complexity of the model may even lead to decreased accuracy. Hence, lemmatization needs to be incorporated which means converting these two words into a single meaningful form Gui et al. (2009). After processing the data, it contains only the essential words required for classification. Words are mapped to numbers which are taken as features for the mathematical models. There can be many features but there is a need to have important

features selected through selection and generation of relevant features to improve the accuracy.

4.3 Feature engineering

Articles processed require the embedding approaches for prepossessing the text of the news articles. There are many widely used embedding techniques but a suitable one is the Count Vector and frequency'inverse document frequency(TF-IDF) encoding which will work as a feature preparation phase (Wang et al. 2010). Count Vector embedding helps in forming a vocabulary from the given corpus or the news articles by calculating the frequency of each word and returns a vector with these frequencies assigned to the respective words of the news articles. On the other hand, TF-IDF vectorizer embedding technique calculates the TF-IDF frequency of each word in each document. These frequencies generate sparse matrices which are further used as the feature sets for the model. The mathematical formula for Tf-IDF frequency is given as $Tf - Idf(t) = Tf(t) * Idf(t)$.

Here $Tf(t)$ represents the number of times the words t comes in a targeted document) divided by the total number of words in the targeted document. $Idf(t)$ is log_e of the total number of documents divided by the number of documents with the word t in them.

4.4 Proposed model

After preprocessing and feature engineering phase, a proposed model with the integration of the feature enhancement and categorisation module has been presented as follows. The proposed model is based on the penetration of feature enhancement and machine learning model MNB. This proposed model takes three contents of the news articles to learn and classify the news articles which includes the title of the news article, news articles body/content and tags of the news articles. Following is the proposed model 1.

where $A^{'}$ represents the article that is free from stop words and punctuation. Then Lemmatization using StanfordNLP python library for the Lemmatization process Al Omran et al. (2017) is conducted. Then the vectorization of the articles is required to assign weights to the words in the article which has been carried out using two different commonly used approaches-Count Vectorization and TF-IDF Vectorization. Finally, the encoded articles are given as input to the classifiers. The proposed model in pictorial form is depicted in the Fig. 1. In the proposed model, procedural, news articles are obtained from three sources. In the proposed model, procedurally, news articles are obtained from three sources. They are combined together to make a centralised data set. In the next phase, the news articles are moved to the data cleaning phase. After this phase, the data of news articles is processed for removal of insignificant words, removal of redundant words. This makes the processing of the articles efficient and these types of words do not contribute to disturbing the articles from other categories. After this phase, the feature engineering process starts with conversion of the words into the feature in numerical form through vectorization. There are two feature generation techniques for the news articles techniques. The proposed model uses two vectersiation techniques namely count vector and TF-IDF vectorization. They both are integrated for the feature generation and selection integrated with ML models. Data set with property feature engineering is divided into two sets of 80 percent and 20 percent for the training and testing of the proposed model. Now the proposed model TF-IDF+MNB is run and results are obtained. To ensure the robustness of the model, it uses the cross validation techniques to ensure that results don't suffer from overfitting. The model is evaluated on several criteria of classification problems such as accuracy, confusion matrix, f-score, recall, etc. Model is also compared with other machine learning models such as SVM, RF, LR and KNN on several criteria.

Algorithm 1: Proposed model(A)

Result: Category of the article
Let $A^{'}$ be a new string array;
while *token* $\in A$ **do**
  if *token* $\notin$ *(stopwords* $\cup$ *punctuation)* **then**
    $A^{'}$ $A^{'} \cup$ token;
  end
end
A_{lemma} lemmatize($A^{'}$);
A_{vec} vectorize(A_{lemma});
category classify(A_{vec});
return category;

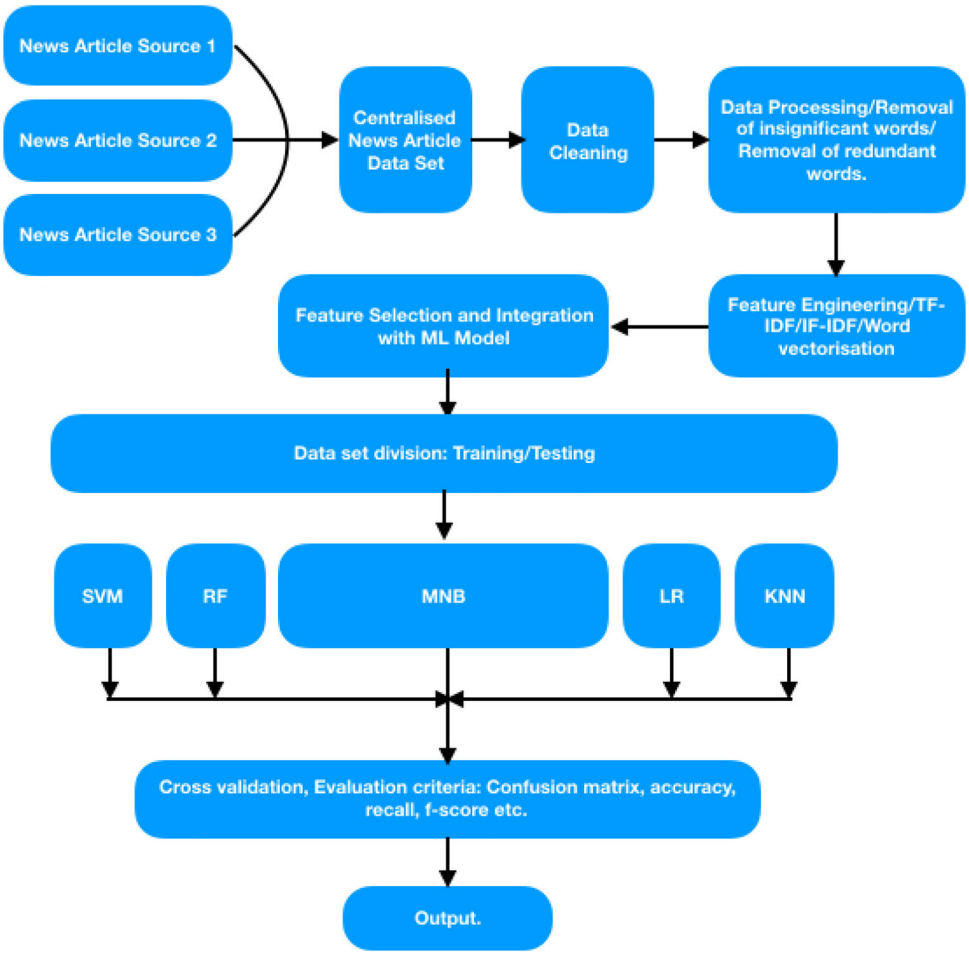


Fig. 1 Proposed model based on TF-IDF integrated MNB

4.5 Evaluation criteria

News article categorisation falls under the classification task and there are many standard evaluation matrices for the classification tasks such as confusion matrix, accuracy, precision and recall and f-measure. The proposed model solves multiclass classification problems of assigning news articles to pre-existing 8 categories of articles. The first performance metric is the confusion matrix that is described as follows. A confusion matrix provides a summary of classification results. Confusion matrix summarizes the number of correct and incorrect classifications with their count values by each class. Evaluation metric accuracy is considered as the most intuitive measure of performance (Sokolova and Lapalme 2009). It is simply a ratio of correctly predicted observation to total observations and defined as the $accuracy = (TP + TN)/(TP + TN + FP + FN)$. Other metrics are precision and recall that can be defined mathematically as follows. Let us define an experiment from P positive instances and N negative instances for some conditions. Then the precision and recall will be formulated as: $Precision(p) = TP/(TP + FP)$ and $Recall(r) = TP/(TP + FN)$. The last evaluation metric is F-measure, which is the harmonic mean of measurement precision and recall. F1 score reaches its best

Fig. 2 Experimental data with category, key words and article

	Category	Keywords	Article
0	2	\nBring on the laughs	dash stand comedy lot improvisation ton audien...
1	2	\nTamannaah Bhatia to play leading lady in Sun...	Actor Tamannaah Bhatia excited sign acclaim di...
2	2	\nSensation Rise 2018: Rise for the city and f...	say never forget first tricity music festival ...
3	2	\nThe Beatles in India\n	year 1968 Beatles spend time away limelight le...
4	2	\nNetflix renews Sacred Games for a second sea...	Sacred game Netflix first original indian show...

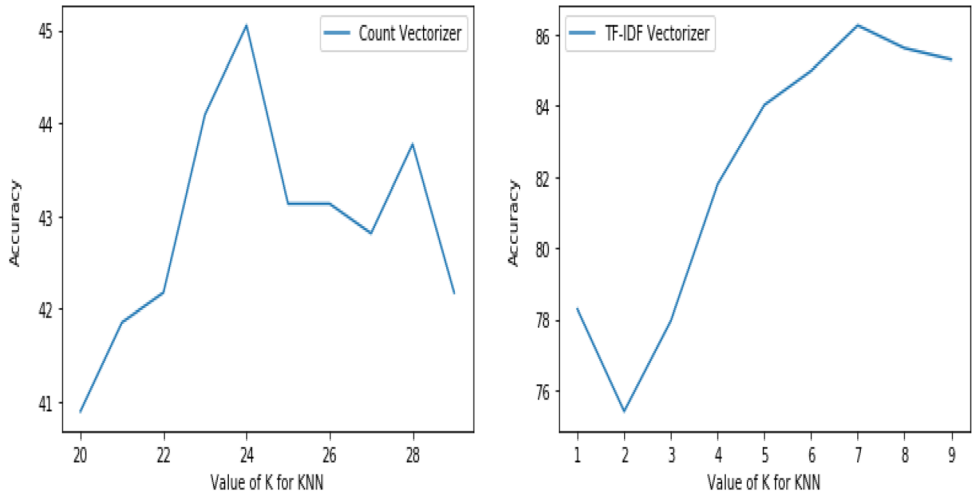


Fig. 3 Plot of accuracy versus the number of nearest neighbors using **a** Countvectorizer. **b** TF-IDF Vectorizer.

Table 1 Accuracy of all the classifiers using different vectorizers for categorization.

ML Models	Count Vectorizer	TF-IDF Vectorizer
Multinomial Naive Bayes(MNB)	**87.22**	**85.62**
Random Forest(RF)	79.55	81.15
Support Vector Machine(SVM)	82.42	85.62
K Nearest Neighbor(KNN)	45.04 (k=24)	86.26 (k=7)
Logistic Regression(LR)	84.02	85.26

value at 1 which is the perfect precision and recall and worst at 0. $F - measure = 2 * p * r/(p + r)$.

4.6 Experiment setup

The experiment has been conducted on Intel Core(TM) i5-4690 CPU@3.20 GHz with RAM 4 GB, 64 bit OS, x64 based processor. The work has been done using a Python programming language.

5 Experimental results and analysis

The articles were split into training and testing data in 80 percent and 20 percent respectively. The training dataset contained around 160 articles of each category and 40 articles of each category in the testing dataset. The sample dataset looks as given in Fig. 2. It contains the fields such as **Category**, **Keyword/Tags** and **Article**. The Category field contains a category id for each of the categories of the news articles, the keyword column is the about the headline of the article. The article column contains the article body which model would be processed to classify the news articles together with other features.

The results have been compiled in tables and have been put in pictorial representations through plots. Several models apart from the proposed model have been evaluated on evaluation criteria such as confusion matrix, accuracy,

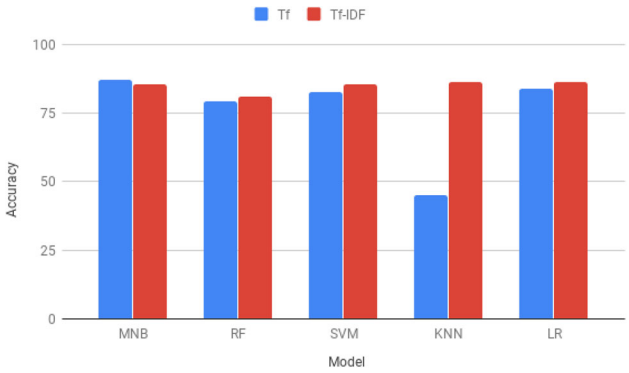


Fig. 4 Bar plot comparing accuracy of different algorithms

recall, f-score etc. The good results have been highlighted and compared with our proposed model on MNB. Let us take each ML model and compare it with the proposed model integrated with TF-IDF and MNB. KNN faces difficulty as it needs to decide and choose the correct value of the number of neighbours for good accuracy in the article categorization. This is achieved by testing the KNN algorithms for certain values of the number of neighbours. This gives the trend of how the change in the number of neighbours is changing the accuracy of the KNN model. The below Fig. 3 shows the trend of accuracy versus the number of neighbours, parameters of the model, for different vectorizers in the news article classification. It is observed that as the value of K increases, the value of

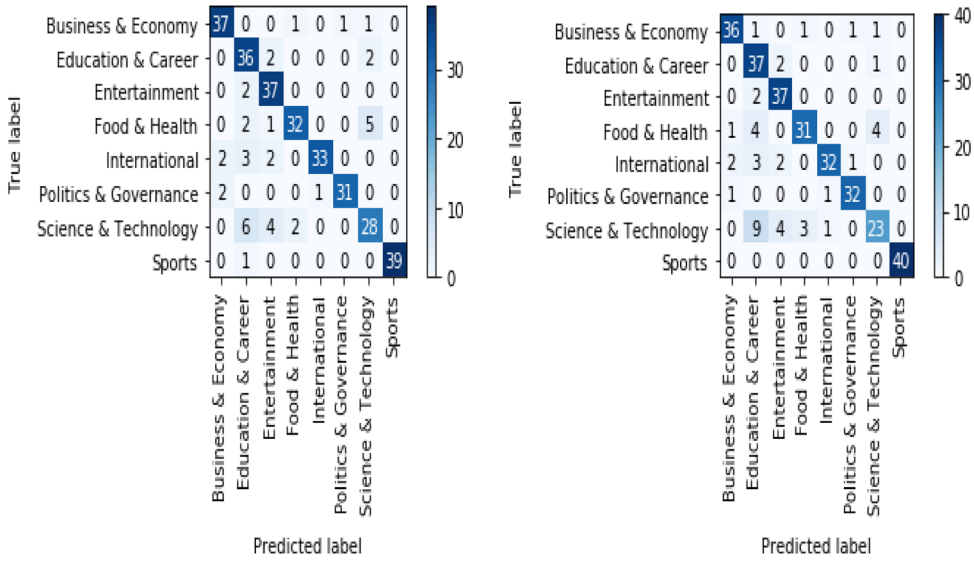


Fig. 5 Confusion matrix for Multinomial Naive Bayes **a** Using Count Vectorizer. **b** Using TF-IDF Vectorizer.

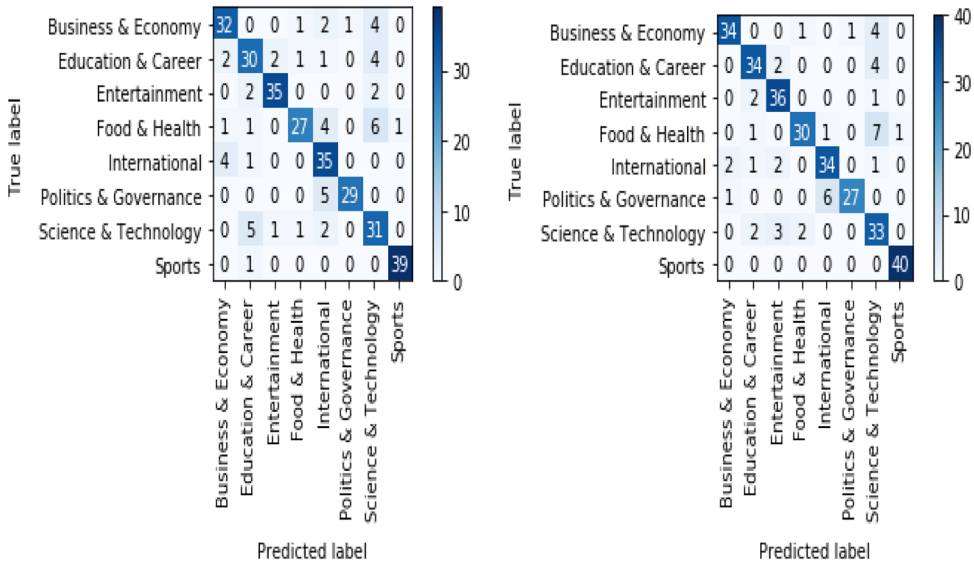


Fig. 6 Confusion matrix for Support Vector Machines **a** Using Count Vectorizer. **b** Using TF-IDF Vectorizer.

accuracy increases and on values of K equals to 4 and 8, KNN model shows the least accuracy. In the second Fig. 3 , with an increase in value of K, the accuracy of the KNN model increases.

The model got iterated in the range of 1 to 40 for the parameter value. The number of nearest neighbours of the model was quite costly in terms of efficiency. Hence the range of K has been restricted to 40 only and obtained good results with TF-IDF vectorizer. It needs a lot of storage to store all the training data. The other three algorithms were computationally less costly compared to KNN (Table1). Proposed model of TF-IDF + MNB produced 87.22 percent with Count Vectorizer as shown in Table1 and it was also computationally cheap as compared to other ML Models. Additionally, It scales relatively well to high dimensional data due to its properties. It is also depicted that the TF-IDF + KNN model produced good results with the value of K being less than 10 while other models of MNB produced good results with the value of K in range from 20 to 30. Specifically, KNN model with Count Vectorizer feature selection technique produced accuracy of 45.04 with the value of K equal to 24 while the KNN with TF-IDF produced the article classification accuracy of 86.26 with value of K equivalent to 7. Coming to the MNB based proposed model, Count Vecteriser + MNB model produced the best results of the accuracy 87.22 and TF-IDF + MNB produced 85.62. This is the only proposed model which has performed well for both feature selection techniques and closer to this accuracy comes the very simple model of LR. Count Vectoeriser + LR classified the news articles with accuracy of 84.02 and TF-IDF + LR produced the articles with 85.26 accuracy which is close to the

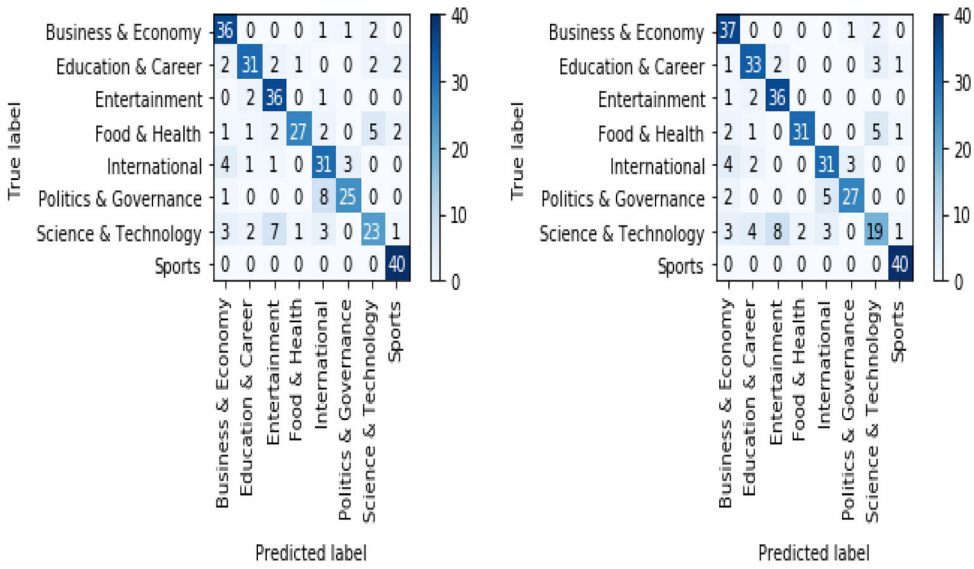


Fig. 7 Confusion matrix for Random Forest Neighbours **a** Using Count Vectorizer. **b** Using TF-IDF Vectorizer.

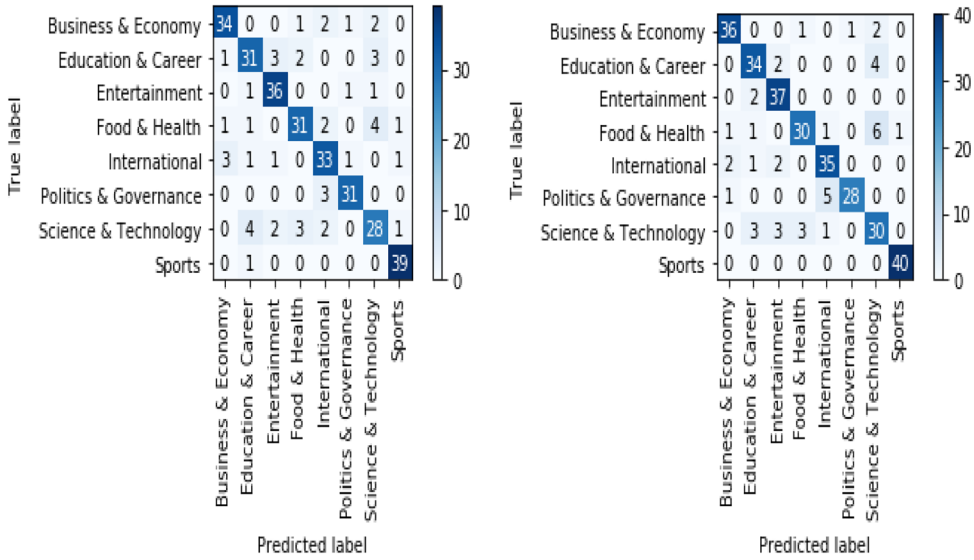


Fig. 8 Confusion matrix for Logistic Regression **a** Using Count Vectorizer. **b** Using TF-IDF Vectorizer.

proposed model. The count vectoriser + SVM and TF-IDF + SVM models produce the accuracy of 82.42 and 85.62. The parameter of the SVM used are $C = 1.0$, $kernel =' rbf'$, $degree = 3$, $gamma = 0.0$, $coef = 0.0$ The SVM produced very low accuracy compared to other ML models. It scales relatively well to high dimensional data. The more samples in the training set the more complex and slow the classification process. Table 1 lists the accuracy of the all ML models with counter vectorizer and TF-IDF. RF based integrated model with count vector and TF-IDF with 79.55 and 81.15 respectively. The results of RF have been obtained with parameter $estimator = 100$, $jobs = -1$.

Figure 4 shows the plot of the accuracies of all ML models used for the classification of news articles. The maximum accuracy is obtained with MNB followed by LR, KNN, SVM, and RF. Only MNB gives better accuracy with TF rather than TF-IDF processing of articles. The rest of the models perform better in processing with TF-IDF vectorisation. KNN accuracy is significantly low. The confusion matrix of the classification gives us a better visualization of how precisely the proposed machine learning models such as TF-IDF+MNB are able to classify each of the categories. The diagonal values of the confusion matrix represent the true positives. Figure 5 displays the confusion matrix of both count vectoriser+MNB and TF-IDF+MNB. It shows that MNB classifies articles with sports categories very accurately. TF-IDF+MNB faces difficulties in classifying the science and technology articles and performing the least accurately and confuses them with education and career and with food articles also. Figure 6 displays results of confusion matrix with SVM model. SVM faces problems in classifying articles on food

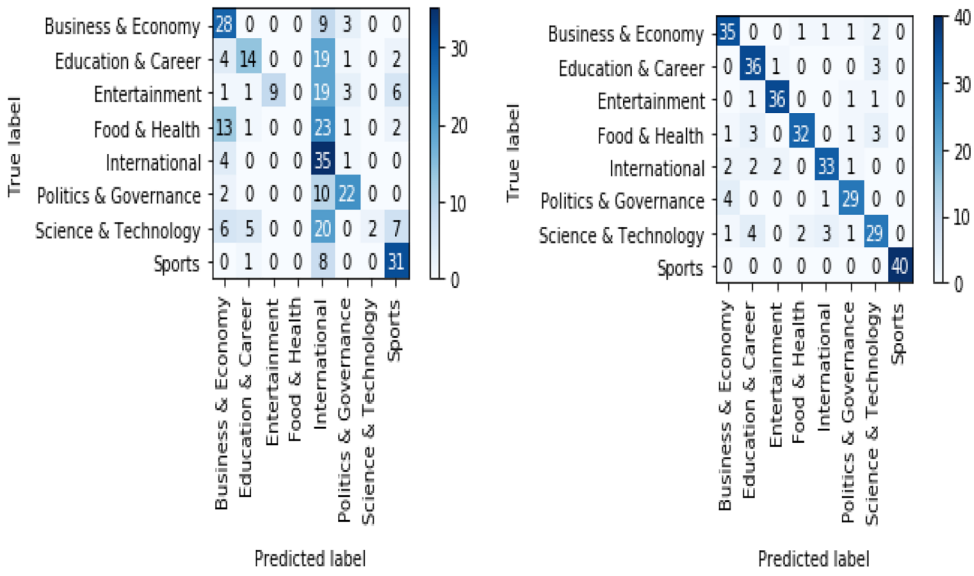


Fig. 9 Confusion matrix for KNN **a** Using Count Vectorizer. **b** Using TF-IDF Vectorizer.

and health and confuses them with science and technology and international news. Although international categories can overlap with health. SVM also performs better in the sports category. Figure 7 displays results with RF models and gives better accuracy with sports articles and less with science and technology. Similar results are obtained in the TF-IDF preprocessing phase. Figure 8 displays results of LR using a count vector and TF-IDF vectoriser. Here also similar results are obtained with highest classification accuracy with sports category and lowest in science and technology. Figure 9 displays results of KNN models with the best accuracy in the sports category but it gives a worse performance in science and technology articles and confuses them with business education. It is observed that the precision of the sports category for most of the algorithms is quite good which shows that the algorithms are good at recognising the articles belonging to the sports category. The confusion matrix of the Support Vector Machine for both TF and TF-IDF vectorizer is very disturbed which shows that the algorithm is not quite good at predicting the classes correctly which is also depicted by its accuracies.

The confusion matrix of the classification gives us a better visualization of how precisely the algorithms are able to classify each of the categories. The diagonal values of the matrix represent the true positives i.e., the classification label and the true label are the same. Figure 5 displays the confusion matrix of both CV and TF-IDF preprocessing articles methods and gives a clear idea of each category article classification and assessment. It shows that MNB classifies articles with sports category very accurately and science and technology articles least accurately and confuses them with education and career and with food articles. Figure 6 displays results of confusion matrix with SVM model. SVM faces problems in classifying articles on food and health and confuses them with science and technology and international news. Although international categories can overlap with health. SVM also performs better in the sports category. Figure 7 displays results with RF models and gives better accuracy with sports articles and less with science and technology. Similar results are obtained in the TF-IDF preprocessing phase. Figure 8 displays results of LR using a count vector and TF-IDF vectoriser. Here also similar results are obtained with highest classification accuracy with sports category and lowest in science and technology. Figure 9 displays results of KNN models with the best accuracy in the sports category but it gives a worse performance in science and technology articles and confuses them with business education. It is observed that the precision of the sports category for most of the algorithms is quite good which shows that the algorithms are good at recognising the articles belonging to the sports category. The confusion matrix of the SVM for both TF and TF-IDF vectorizer is very disturbed which shows that the algorithm is not quite good at predicting the classes correctly which is also depicted by its accuracies.

The Table 2 also lists the Precision, Recall and F-Measure. The precision metric shows how correctly the algorithms can classify into each category and recall shows out of all his categories how well can the algorithm recall a particular category. According to the results, the proposed model with TF-IDF+MNB, in the sports category, recall, f-measure and precision is 1.0 which is highest. Other machine learning models don't produces that level of accuracy and recall and f-score as depicted in the Table 2. The least results are obtained in health and food and science and technology category due to their similar nature of content which confuses the models to distinguish them.

Table 2 Precision, Recall and F-Measure of category classification of all the algorithms

Category	ML Models		Precision	Recall	F-Measure
Business and Economy	MNB	TF	0.9250	0.9024	0.9135
		TF-IDF	0.9000	0.9000	0.9000
	LR	TF	0.8500	0.8717	0.8607
		TF-IDF	0.9000	0.9000	0.9000
	SVM	TF	0.8000	0.8205	0.8101
		TF-IDF	0.8500	0.9189	0.8831
	RF	TF	0.9000	0.7659	0.8275
		TF-IDF	0.9250	0.7400	0.8222
	KNN (k=7)	TF-IDF	0.8750	0.8139	0.8433
Education and Career	MNB	TF	0.9000	0.7200	0.7999
		TF-IDF	0.9250	0.6607	0.7708
	LR	TF	0.7750	0.7948	0.7848
		TF-IDF	0.8500	0.8292	0.8395
	SVM	TF	0.7500	0.7500	0.7500
		TF-IDF	0.8500	0.8500	0.8500
	RF	TF	0.7750	0.8378	0.8051
		TF-IDF	0.8250	0.7857	0.8048
	KNN (k=7)	TF-IDF	0.9000	0.7826	0.8372
Entertainment	MNB	TF	0.9487	0.8043	0.8705
		TF-IDF	0.9487	0.8222	0.8809
	LR	TF	0.9230	0.8571	0.8889
		TF-IDF	0.9487	0.8409	0.8915
	SVM	TF	0.8974	0.9210	0.9090
		TF-IDF	0.9230	0.8372	0.8780
	RF	TF	0.9230	0.7500	0.8275
		TF-IDF	0.9230	0.7826	0.8048
	KNN (k=7)	TF-IDF	0.9230	0.9230	0.9230
Food and Health	MNB	TF	0.8000	0.9143	0.8533
		TF-IDF	0.7750	0.8857	0.8266
	LR	TF	0.7750	0.8378	0.8052
		TF-IDF	0.7500	0.8824	0.8108
	SVM	TF	0.6750	0.9000	0.7714
		TF-IDF	0.7500	0.9090	0.8219
	RF	TF	0.6750	0.7500	0.8275
		TF-IDF	0.7750	0.9393	0.8493
	KNN (k=7)	TF-IDF	0.8000	0.9142	0.8533
International	MNB	TF	0.8250	0.9705	0.89189
		TF-IDF	0.8000	0.9411	0.8648
	LR	TF	0.8250	0.7857	0.8049
		TF-IDF	0.8750	0.8333	0.8536
	SVM	TF	0.8750	0.7142	0.7865
		TF-IDF	0.8750	0.8974	0.8860
	RF	TF	0.7750	0.6739	0.7209
		TF-IDF	0.7750	0.7948	0.7848
	KNN (k=7)	TF-IDF	0.8250	0.8684	0.8461

Table 2 continued

Category	ML Models		Precision	Recall	F-Measure
PoliticsandGovernance	MNB	TF	0.9117	0.9688	0.9394
		TF-IDF	0.9411	0.9411	0.9411
	LR	TF	0.9117	0.9117	0.9117
		TF-IDF	0.8235	0.9655	0.8888
	SVM	TF	0.8529	0.9666	0.9062
		TF-IDF	0.7941	0.9642	0.8709
	RF	TF	0.7352	0.8620	0.7936
		TF-IDF	0.7941	0.8709	0.8307
	KNN (k=7)	TF-IDF	0.8529	0.8529	0.8529
Science and technology	MNB	TF	0.7000	0.7777	0.7368
		TF-IDF	0.5750	0.7931	0.6667
	LR	TF	0.7000	0.7368	0.7179
		TF-IDF	0.7500	0.7142	0.7317
	SVM	TF	0.7750	0.6595	0.7126
		TF-IDF	0.8250	0.6600	0.7333
	RF	TF	0.5750	0.71875	0.6388
		TF-IDF	0.4750	0.6551	0.5507
	KNN (k=7)	TF-IDF	0.7250	0.7631	0.7435
Sports	MNB	TF	0.9750	**1.0000**	0.9873
		TF-IDF	**1.0000**	**1.0000**	**1.0000**
	LR	TF	0.9750	0.9286	0.9512
		TF-IDF	**1.0000**	0.9756	0.9877
	SVM	TF	0.9750	0.9750	0.9750
		TF-IDF	**1.0000**	0.9756	0.9876
	RF	TF	**1.0000**	0.8888	0.9411
		TF-IDF	0.1000	0.9302	0.9638
	KNN (k=7)	TF-IDF	**1.0000**	**1.0000**	**1.0000**

Table 3 Time taken during the training period by the classifiers using different vectorizers for categorization.

ML Models	Count Vectorizer	TF-IDF Vectorizer
Multinomial Naive Bayes(MNB)	**1.82s**	**3.86s**
Random Forest(RF)	6.88s	6.74s
Support Vector Machine(SVM)	130s	147s
K Nearest Neighbor(KNN)	28.9s	29s
Logistic Regression(LR)	37.3s	16.2s

Table 3 gives a running time analysis of all machine learning models with both Counts vectoriser and TF-IDF vectoriser. The time taken is maximum in SVM and least in proposed model MNB. The MNB performs the best in both the cases of accuracy and time efficiency in the classification of news articles.

6 Conclusion and future works

Digital technologies based information sharing platforms such as news websites and social media platforms such as Facebook, Twitter, Instagram, What's app etc have flooded the information space. Information or text classification has an important role to play so that important and vital information can be classified based on the already

predefined categories. Nowadays the process of classification and segregation of textual information has become challenging due to the flow of diverse, vast information. Automated and intelligent tools which can classify the information and text accurately and efficiently are needed to reduce human efforts, time and increase productivity. The research work contributes in three ways. It proposed a model which takes news titles and news stories for classification of the news articles which is an improvement over previous work. It proposed a machine learning-based model which is very efficient and effective in news categorisation. Secondly, it performs classification and comparison on newly developed data sets of news stories. This paper presents an intelligent, efficient and robust intelligent machine learning model. The proposed IDF integrated MNB model achieves classification accuracy of 87.22 per cent and results have been compared with state of art machine learning algorithms such as LR, KNN, SVM and RF. The proposed model Count Vectorizer integrated MNB and IDF integrated MNB is also better in terms of efficiency with running time of 1.82s and 3.86s respectively. The work can also be improved by introducing more selection methods and deep neural network architects in future works.

Acknowledgements Acknowledgements, if any, should follow the conclusions, and be placed above any Appendices or the references.

Funding No funding received for this research work.

Declaration

Conflict of Interest There is no conflict of interest in this work.

References

Ahmed M, Mahmood AN, Islam MR (2016) A survey of anomaly detection techniques in financial domain. Future Gener Comput Syst 55:278–288

Al Omran FNA, Treude C (2017) Choosing an nlp library for analyzing software documentation: a systematic literature review and a series of experiments. In: 2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR), IEEE, pp 187–197

Alhothali A, Hoey J (2015) Good news or bad news: Using affect control theory to analyze readers'reaction towards news articles. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp 1548–1558

Ashuri T (2016) When online news was new: Online technology use and constitution of structures in journalism. Journal Stud 17(3):301–318

Bail CA (2017) Taming big data: Using app technology to study organizational behavior on social media. Sociol Methods Res 46(2):189–217

Boumans JW, Trilling D (2016) Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. Digit Journal 4(1):8–23

Canito J, Ramos P, Moro S, Rita P (2018) Unfolding the relations between companies and technologies under the big data umbrella. Comput Ind 99:1–8

Carstens L, Toni F (2017) Using argumentation to improve classification in natural language problems. ACM Trans Internet Technol (TOIT) 17(3):30

Castillo C, El-Haddad M, Pfeffer J, Stempeck M (2014) Characterizing the life cycle of online news stories using social media reactions. In: Proceedings of the 17th ACM conference on Computer supported cooperative work and social computing, ACM, pp 211–223

Chava RVSP, Dhar S, Gaur Y, Rambhakta P, Shetty S (2018) Big data text summarization-hurricane irma

Conroy NJ, Rubin VL, Chen Y (2015) Automatic deception detection: Methods for finding fake news. Proc Assoc Inf Sci Technol 52(1):1–4

Daud A, Khan W, Che D (2017) Urdu language processing: a survey. Artif Intell Rev 47(3):279–311

Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Stanley HE, Quattrociocchi W (2016) The spreading of misinformation online. Proc Nat Acad Sci 113(3):554–559

Fong S, Zhuang Y, Li J, Khoury R (2013) Sentiment analysis of online news using mallet. In: Proceedings of the 2013 International Symposium on Computational and Business Intelligence, IEEE Computer Society, Washington, DC, USA, ISCBI '13, pp 301–304, 10.1109/ISCBI.2013.67

Gui Y, Gao Z, Li R, Yang X (2012) Hierarchical text classification for news articles based-on named entities. In: International Conference on Advanced Data Mining and Applications, Springer, pp 318–329

Habash N, Rambow O, Roth R (2009) Mada+ tokan: A toolkit for arabic tokenization, diacritization, morphological disambiguation, pos tagging, stemming and lemmatization. In: Proceedings of the 2nd international conference on Arabic language resources and tools (MEDAR), Cairo, Egypt, vol 41, p 62

Hakim AA, Erwin A, Eng KI, Galinium M, Muliady W (2014) Automated document classification for news article in bahasa indonesia based on term frequency inverse document frequency (tf-idf) approach. In: 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE), IEEE, pp 1–4

Hogenboom F, Frasincar F, Kaymak U, De Jong F, Caron E (2016) A survey of event extraction methods from text for decision support systems. Decis Support Syst 85:12–22

Hsu WL, Lang SD (1999) Classification algorithms for netnews articles. In: Proceedings of the Eighth International Conference on Information and Knowledge Management, ACM, New York, NY, USA, CIKM '99, pp 114–121, 10.1145/319950.319965

Hu Y, Ye X, Shaw SL (2017) Extracting and analyzing semantic relatedness between cities using news articles. Int J Geogr Inf Sci 31(12):2427–2451

Joachims T (1998) Making large-scale svm learning practical. Technical report, Tech rep

Kanan T, Fox EA (2016) Automated Arabic text classification with p-s temmer, machine learning, and a tailored news article taxonomy. J Assoc Inf Sci Technol 67(11):2667–2683

Kibriya AM, Frank E, Pfahringer B, Holmes G (2004) Multinomial naive bayes for text categorization revisited. In: Australasian Joint Conference on Artificial Intelligence, Springer, pp 488–499

Kiranyaz S, Ince T, Gabbouj M (2014) Multidimensional particle swarm optimization for machine learning and pattern recognition. Springer, Berlin

Korb KB, Nicholson AE (2010) Bayesian artificial intelligence. CRC Press, Florida

Kumar S, Kalia A, Sharma A (2018a) Predictive analysis of alertness related features for driver drowsiness detection. Adv Intell Syst Comput 736:368–377

Kumar S, Pal SK, Singh R (2018b) A novel method based on extreme learning machine to predict heating and cooling load through design and structural attributes. Energ Build 176:275–286

Kumar S, Singh R, Pal SK (2018c) A conceptual architectural design for intelligent health information system: Case study on india. Quality, IT and Business Operations: Springer Proceedings in Business and Economics, vol 1. Springer, Singapore, pp 1–15

Kumar S, Saibal KP, Singh R (2019) A novel hybrid model based on particle swarm optimisation and extreme learning machine for short-term temperature prediction using ambient sensors. Sustain Cities Soc

Kumar S, Singh J, Singh O (2020) Ensemble-based extreme learning machine model for occupancy detection with ambient attributes. Int J Syst Assur Eng Manag

Kurt I, Ture M, Kurum AT (2008) Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. Expert Syst Appl 34(1):366–374

Li J, Fong S, Zhuang Y, Khoury R (2016) Hierarchical classification in text mining for sentiment analysis of online news. Soft Comput 20(9):3411–3420

Lin WY, Hu YH, Tsai CF (2011) Machine learning in financial crisis prediction: a survey. IEEE Trans Syst, Man, Cybern, Part C (Appl Rev) 42(4):421–436

Lykourentzou I, Giannoukos I, Nikolopoulos V, Mpardis G, Loumos V (2009) Dropout prediction in e-learning courses through the combination of machine learning techniques. Comput Educ 53(3):950–965

Marconi F (2020) Newsmakers: artificial intelligence and the future of journalism. Columbia University Press, Columbia

McCallum A, Nigam K, et al. (1998) A comparison of event models for naive bayes text classification. In: AAAI-98 workshop on learning for text categorization, Citeseer, vol 752, pp 41–48

Medagoda N (2016) Sentiment analysis on morphologically rich languages: An artificial neural network (ann) approach. In: Artificial Neural Network Modelling, Springer, pp 377–393

Miller K, Oswalt A (2017) Fake news headline classification using neural networks with attention. Tech. Rep., California State University

Mukwazvure A, Supreethi K (2015) A hybrid approach to sentiment analysis of news comments. 2015 4th International Conference on Reliability. Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), IEEE, pp 1–6

Noble WS (2006) What is a support vector machine? Nat Biotechnol 24(12):1565–1567

Onan A, Korukoğlu S, Bulut H (2016) Ensemble of keyword extraction methods and classifiers in text classification. Expert Syst Appl 57:232–247

Paragios N, Chen Y, Faugeras OD (2006) Handbook of mathematical models in computer vision. Springer, Berlin

Pröllochs N, Feuerriegel S, Neumann D (2016) Negation scope detection in sentiment analysis: Decision support for news-driven trading. Decis Support Syst 88:67–75

Ramík DM, Sabourin C, Moreno R, Madani K (2014) A machine learning based intelligent vision system for autonomous object detection and recognition. Appl Intell 40(2):358–375

Rasouli K, Hsieh WW, Cannon AJ (2012) Daily streamflow forecasting by machine learning methods with weather and climate inputs. J Hydrol 414:284–293

Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ (2016) Deep learning for health informatics. IEEE J Biomed Health Inf 21(1):4–21

Salminen J, Yoganathan V, Corporan J, Jansen BJ, Jung SG (2019) Machine learning approach to auto-tagging online content for content marketing efficiency: A comparative analysis between methods and content type. J Bus Res 101:203–217

Sen PC, Hajra M, Ghosh M (2020) Supervised classification algorithms in machine learning: A survey and review. In: Emerging technology in modelling and graphics, Springer, pp 99–111

Singh G, Kumar B, Gaur L, Tyagi A (2019) Comparison between multinomial and bernoulli naïve bayes for text classification. 2019 International Conference on Automation. Computational and Technology Management (ICACTM), IEEE, pp 593–596

Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Inf Process Manag 45(4):427–437

Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. Neural Process Lett 9(3):293–300

Thomson T, Angus D, Dootson P, Hurcombe E, Smith A (2020) Visual mis/disinformation in journalism and public communications: current verification practices, challenges, and future opportunities. Journalism Practice pp 1–25

Tong S, Koller D (2001) Support vector machine active learning with applications to text classification. J Mach Learn Res 2(Nov):45–66

Ur-Rahman N, Harding JA (2012) Textual data mining for industrial knowledge management and text classification: A business oriented approach. Expert Syst Appl 39(5):4729–4739

Valdivia A, Luzón MV, Herrera F (2017) Sentiment analysis in tripadvisor. IEEE Intell Syst 32(4):72–77

Van Veldhoven Z, Vanthienen J (2021) Digital transformation as an interaction-driven perspective between business, society, and technology. Electron Mark pp 1–16

Wang N, Wang P, Zhang B (2010) An improved tf-idf weights function based on information theory. In: 2010 International Conference on Computer and Communication Technologies in Agriculture Engineering, IEEE, vol 3, pp 439–441

Witten IH, Frank E, Hall MA, Pal C, DATA M (2005) Practical machine learning tools and techniques. In: DATA MINING, vol 2, p 4

Zahid N, Abouelala O, Limouri M, Essaid A (2001) Fuzzy clustering based on k-nearest-neighbours rule. Fuzzy Sets Syst 120(2):239–247

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Engg. Research ICARI Special Issue pp. 22-31 DOI : 10.36909/jer.ICARI.15287

Assessment and modeling of sewer network development utilizing Arc GIS and SewerGEMS in Kabul city of Afghanistan

Ali Reza Noori* and S.K.Singh**

** Department of Environmental Engineering, Delhi Technological University, Delhi, India, & Department of Water Supply and Environmental Engineering, Kabul Polytechnic University, Kabul, Afghanistan*
*** Department of Civil and Environmental Engineering, Delhi Technological University, Delhi, India*
*** Corresponding Author : sksinghdce@gmail.com*

ABSTRACT

The absence of a wastewater collection, management, and disposal scheme is one of Kabul's most serious environmental issues. This has resulted in both health and ecological problems. This research used Arc GIS and SewerGEMS tools to assess the viability of a decentralized sewerage collection model in the research area. The research area was chosen as the city's 5th district. Land-use and land-cover, Digital Elevation Model (DEM), and Satellite data were used to construct the network's geometry in the Arc map environment. SewerGEMS software was used to perform hydraulic simulation and modeling. The variables were regulated based on the results of the study using conventional wastewater topology guidelines. Based on the outputs of hydraulic analysis, it is concluded that the decentralized wastewater collection system would be the best option for the area. It can be deduced from hydraulic design findings that the hydraulic model was successfully developed and built. The methodology can be applied for the development of future wastewater master plans of the city.

Keywords: Hydraulic analysis; Sewer networks; Kabul city; Wastewater collection network; Arc GIS; sewerGEMS.

INTRODUCTION

Wastewater collection, treatment, and proper disposal are the main sanitary requirements of every civilization. Methods for treating, preserving, collecting, and disposing of wastewater can pose risks to the environment and public health (Chawla & Singh, 2014). For organizations and management of urban water and sanitation systems, the effective approach for various areas of the city is the right and timely selection of alternatives that should be specified in terms of physical characteristics, financial, economic, and social acceptance results. The absence of a wastewater collection, management, and disposal scheme is one of Kabul's most serious environmental issues. This has resulted in both environmental and health issues. Nowadays, blackwater is disposed of using sewage and septic tanks, semi-insulated wells, traditional (dry) toilets, and absorption pits (wells). The greywater is also disposing of by utilizing the same methods but is mainly discharging into the open environment. Kabul Municipality has implemented a small-scale biological wastewater treatment scheme and claimed that groundwater would be unusable in the coming years unless urgent and comprehensive steps are taken to deter groundwater contamination in Kabul (Shafaqna, 2019).

Groundwater becomes Kabul's main reliable and accessible source of drinking water (Zaryab et al., 2017). Fecal coliforms have poisoned drinking water and several groundwater sources (Afzali, 2017). Kabul's groundwater is in

bad quality, according to reports (Afzali, 2017; Brati et al., 2019; Frahmand, 2011; Gesim & Okazaki, 2018; Hayat & Baba, 2017; Houben & Tünnermeier, 2005; Noori & Nasimi, 2019; Rhimi & Murakmi, 2017). The emissions have a negative impact on the capital's environment. There are no statistics available on the disease or its causes. According to a study on the BBC website on August 10, 2017, (Ali, 2017), tainted water is the primary source of pathogens in Kabul. According to Antani hospital physicians, polluted water supply is the primary factor of pandemics in more than 70% of patient populations enrolled at the hospital.

According to the existing literature, generally, it can be argued that the city currently lacks a system for collecting and disposing of sanitary wastewater. Just 1% of the town's population is connected to the municipal sewage network, which is also partially active. Based on the current statistics, of the total greywater of residential area 83% and commercial areas, 59% are discharging to open environment. Also, 49% of the total blackwater in residential areas and 81% in commercial areas are disposed of by holding tanks (Consortium for DEWATS Dissemination Society, 2017; Paiman & Noori, 2019). In general, it can be said that the collection and disposal of wastewater by mentioned methods in the city with a population of nearly five million people pose a serious health risk to the lives of all citizens. Furthermore, this requires urgent and immediate action to measure a regular and comprehensive management system for wastewater collection, treatment, and disposal. On the other hand, the utilization of treated wastewater can cause a clean environment and help water resources, especially in arid and semi-arid areas (Aleisa & Alshayji, 2019).

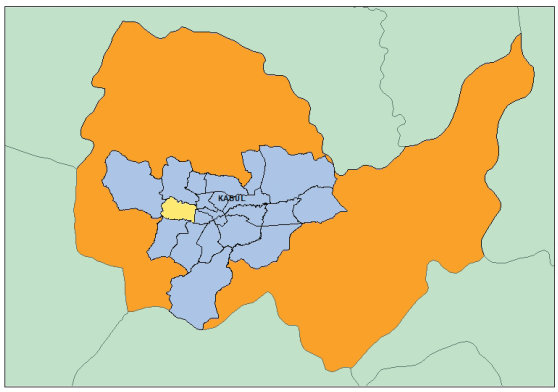
The integrated use of geostatistical approaches, remote sensing, and spatial analysis offers valuable knowledge for detecting, simulating, and discovering environmental components (Kanakiya et al., 2015). The aim of this article is to use Arc GIS and SewerGEMS tools to conduct a preliminary conceptual design for wastewater collection systems in the study area. A land-use and land-cover (LULC) mapping and a Digital Elevation Model (DEM) and aerial photographs are being used for geometric modeling in this analysis. Bentley's SewerGEMS Version 8i was used to perform hydraulic evaluation and network construction. Moreover, variables are monitored based on the study findings and generic sewerage network configuration guidelines.

STUDY AREA

More than half of the global population currently lives in cities, indicating rapid urbanization is a global phenomenon (Bhave & Rahate, 2018; Bu Qammaz et al., 2020). Kabul, Afghanistan's capital, is situated in the country's central east. The city located geographically in 34°39′20" N to 34°17′40" N latitude and 69°31′40″ E to 69°52′40″ E longitude. The annual average precipitation in the basin measured at around 330 mm per year (Source: Afghanistan meteorological department, data recorded 2008-2018). The temperature ranges with a monthly mean of a maximum of 32ºC in July and a minimum of -7ºC in January (Zaryab et al., 2017). It has a population of 4.3 million people (CIA, 2020; NSIA, 2020) and is the country's largest financial center.

.





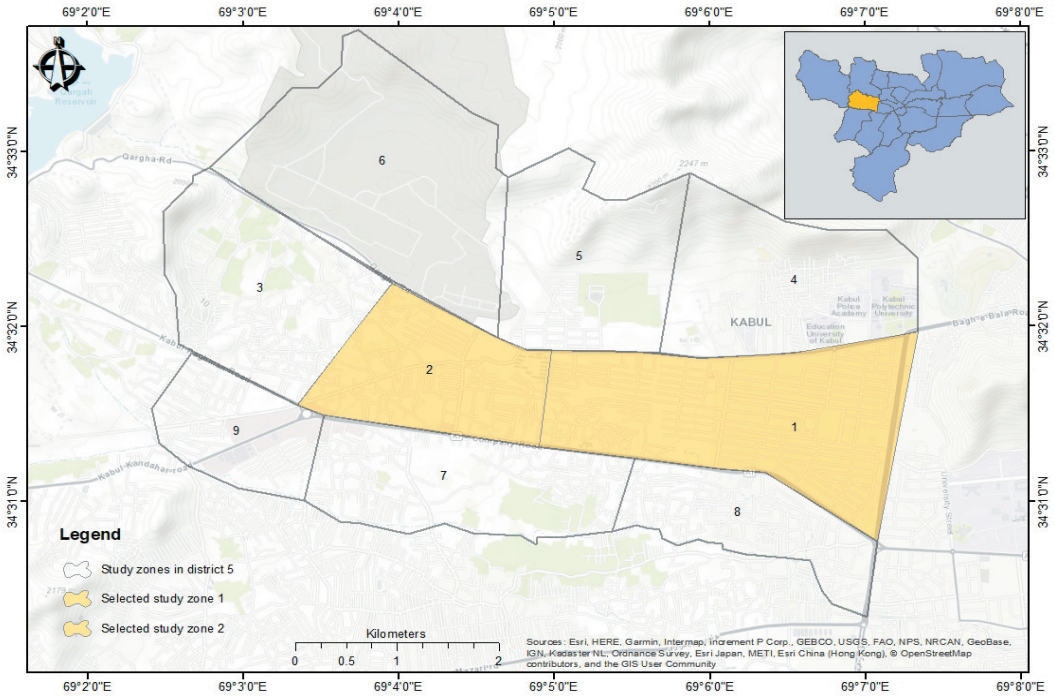


Figure 1. Location map of the study area

There are 22 separate city districts in the area. The planned research area is district five, situated in the western neighborhoods (Figure 1). This area is a built-up area of the capital, with an approximate total of 86000 people, as per Kabul municipal statistics. It has a total area of approximately 28 sq.km and a circumference of approximately 30 km. This field was chosen for the research because it encompasses both developed and unplanned zones. The planned area has covered only about 18% of it. The remaining 82% has developed in an unplanned manner. The study and modeling of the whole region are beyond the scope of this study. Consequently, the research area was already split into nine spatial variabilities, with the model established for just two of them (Figure 1). Both zones occupy a combined area of around nine sq. km, covering the developed region and the other an unplanned location. Topographic trends and projected planning were also taken into account when deciding on research zones. It would be easy to combine the concept to the current areas' primary sewer lines if it is developed in the future.

MATERIALS AND METHODS

The GIS technology greatly simplifies different geospatial scales and can reduce the expense and time of site observations, which is essential for environmental preservation (Kanakiya et al., 2015; Kumar et al., 2015). The sanitary sewer scheme for Kabul's 5th district was modeled in this study. The methodological flowchart is presented in (Figure 2). First, using Arc GIS software, the necessary details were calculated, such as the lengths of the sewage pipes, the surface, and ground level at the manhole positions. After which, in the target geographic region, similar details, including topography, statistical data such as population growth, water use, and sewage generation, were scheduled.

To create a geometric model, the availability of LULC is necessary for the study area. The LULC map created by Afghanistan's ministry of earth and urban development in 2016 was used. Built-up zones, farming, industrial, and administrative regions are included in the LULC. The relative position of motorways in the area has been depicted in LULC (Figure 3a). The number of homes in the area was estimated by the LULC using aerial imagery from the Kabul district. The households are used to estimate the populace and mass flow for hydraulic simulation of the network.

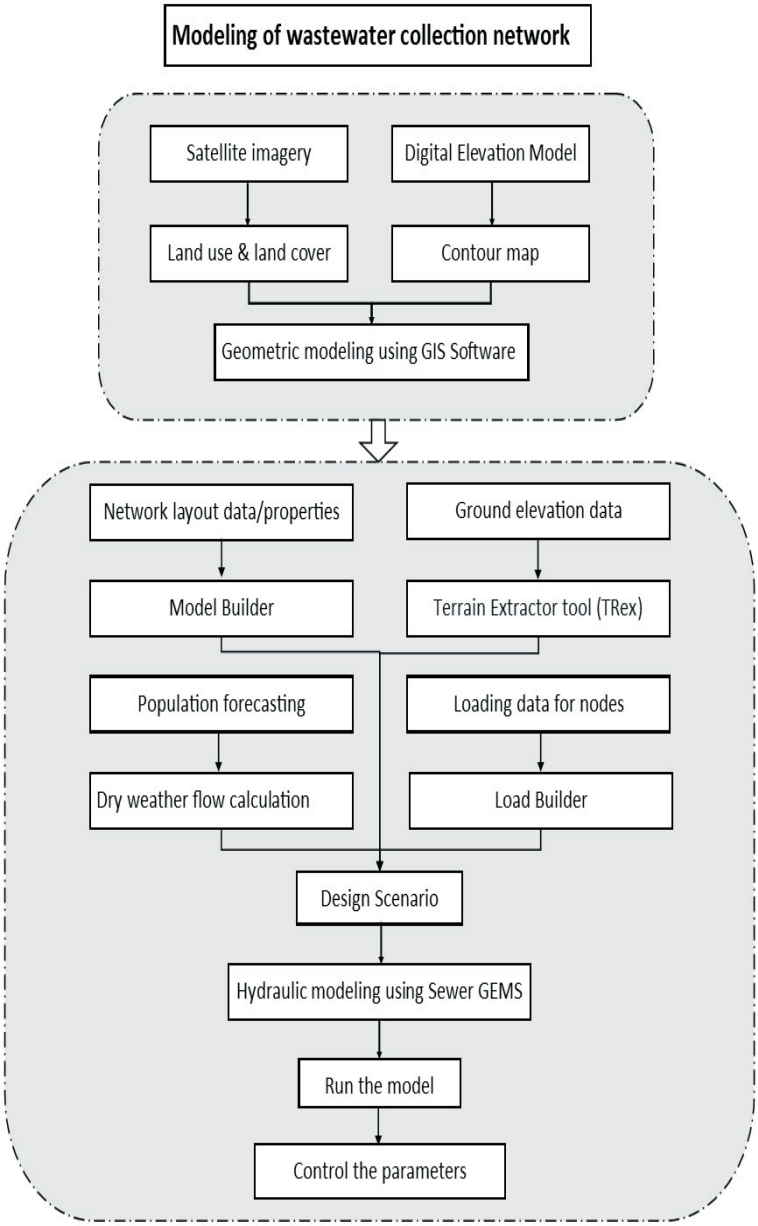
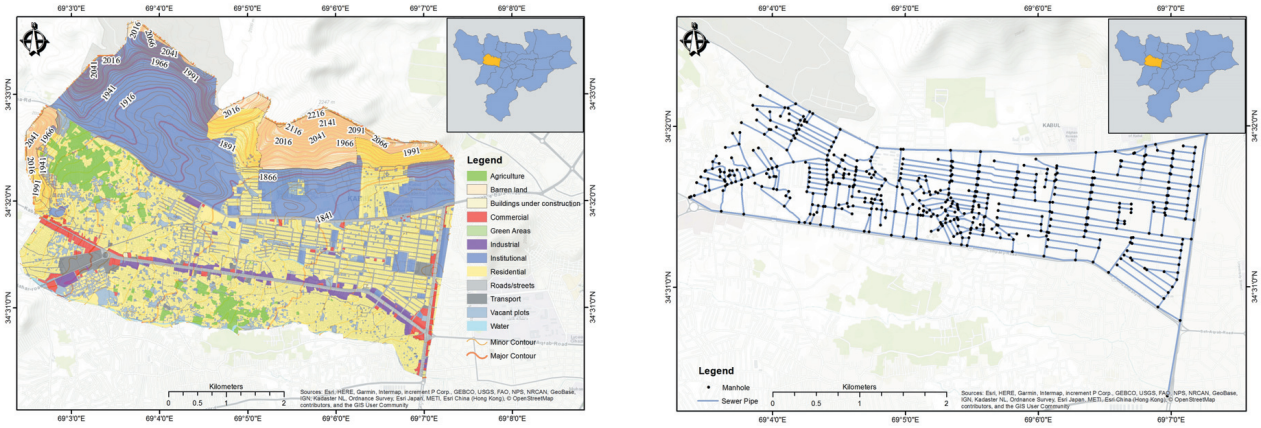


Figure 2. Methodology flowchart showing the modeling of a sewerage network.

For proper network navigation, the contour lines and topography of the region are also used. In practice, this is accomplished by a survey, which takes a significant amount of time and money. DEM, on the other hand, is utilized for analysis and the development of primary schema. DEM is a kind of open-source data that comes in a variety of resolutions. A 30-meter resolution was used in this study.

Bentley's SewerGEMS platform for designing hygiene and hybrid sewerage is one of the company's most popular products. The necessary details and schematic layout of the system were developed using Arc GIS 10.6 in this study. The volume of discharge was calculated based on per person wastewater generation in various locations and the necessary information inserted in SewerGEMS to hydraulically analysis the model. The network's architecture was developed using the research area's LULC and topographical map (Figure 3b). The network's architecture depicts the model's first figures. Based on the network's layout, the manholes' relative positions, and the critical network routes suggested.



a) LULC of the study area b) Geometric model of the sewer network

Figure 3. LULC and geometric of the model

ModelBuilder is a tool available in the Bentley SewerGEMS product that allows the creation and updates of a model from a source file (Akshaya & Jesse, 2020). The shapefiles created in the Arc GIS environment were utilized for modeling using the ModelBuilder technique in SewerGEMS. The Terrain Extractor (TRex) tool will automatically allocate elevation to any node based on the nearest elevation of the .dxf contour, shapefile, LandXML file, or location elevations (Dan & Jesse, 2020b). This method is applied in the model for ground elevation in manholes locations. In order to assign the loads to the hydraulic model, the LoadBuilder method is used in the SewerGEMS environment. In this model, proportional distribution by area (Dan & Jesse, 2020a) is applied to assign loads on the network.

RESULT AND DISCUSSION

Population Density and Discharge Estimation

Kabul has a population of 4.3 million people and a land area of 1023 sq. km. Its density of population is 4500 people per sq. km. The total space occupied by both study zones in the 5th district is 8.834 sq. km. As a result, the population density will be used to estimate the population of the coverage region, which is estimated to be around 40,000 inhabitants. From the other side, the research area is primarily made up of middle and high apartment blocks, which would directly affect population growth. As a result, the current population in the research area is given an extra multiplier (security factor = 1.8) to enhance continued stability. As a result, the research area's population is expected to be 72,000 residents. Kabul is among the fastest-growing cities in the world. Calculating the population growth is one of the principles and requirements for hydraulic analysis of the model. The geometric approach is a well-known and commonly used method for population estimation.

$$P_n = P_0(1 + r)^n \quad (1)$$

Where; Pn- Population in the desired year, P_0 – the current population, r – increasing rate, n – number of years of the design period.

The life of sanitation projects is typically estimated to be about 30 years based on design specifications. The population is expected to rise at a rate of (2.5%) each year. By considering these factors, the population for the next 30 years was estimated as follows.

$$P_{30} = 72000\left(1 + \frac{2.5}{100}\right)^{30} = 151025 \text{ capita}$$

The volume of sewage flowing into the system is equal to the volume of water used in residence. Just the volume of sewage flow in suburban areas without rainwater is determined since this model is not considered a hybrid. Based on the existing condition, water demand in cities of the country is taken equal to 100 liters per capita per day (Seddeqi, 2017)Afghanistan","title":"Water Supply Engineering","type":"book"},"uris":["http://www.mendeley.com/documents/?uuid=ed11a0e5-d019-414a-8729-a6a8e53337f1"]}],"mendeley": {"formattedCitation":"(Seddeqi, 2017. Therefore, the volume of sewage generation is also counted equal to 100 liters per capita per day. By using water demand and population density, the total volume of flow can be determined as follow:

$$Q_{avg} = 151025 \times 100 = \frac{15102500 \text{lit}}{\text{d}} = 15102 \text{m}^3/\text{d}$$

The determined volume of sewer flow from households (Dry Weather Flow = DWF) is equal to 15102 m^3/day. According to the design requirements for wastewater collection systems, a safety factor (s.f.=2) must be added for the efficiency operation of the model. Groundwater penetration, stormwater flows, frequent variation all contribute to this Safety aspect. The following formula is used to calculate the overall maximum outflow:

$$Q_{max} = 15102 \times 2 = 30204 \text{ m}^3/\text{day}$$

HYDRAULIC MODELING AND ANALYSIS

The main objective of a hydraulic analysis of the network is to examine the architectural model that was previously created in the Arc map environment. In many forms of sewer network analysis, hydraulic models are the most effective methodology (Zhang et al., 2018)Norway. The purpose of the control is to use the free space of the pipes to reduce overflow at the wastewater treatment plant (WWTP. Specifically, the goals of hydraulic design are to determine the required diameter of sewer pipes, calculate and control the slope and control the flow velocity in sewer pipes. Since velocity is a critical component in fluid dynamics, it has definite limits according to design criteria. Also, the sewer pipes are designed in incomplete filled, so the ratio of fullness and emptiness of the pipelines must be specified in hydraulic calculations.

The load on the model was applied using the *"area load"* technique in this analysis. The spread of discharge in the pipelines has also been achieved using *"Equal Flow Distribution."* The flow of all pipelines is distributed proportionally based on the area occupied per pipeline. Since some pipelines may have more loads and others may have fewer, such a circumstance will not manifest in practice. This approach is called the most suitable loading framework since this model was developed to investigate the viability of modeling.

The following design scenario was considered for the hydraulic design: a) The type of pipe material selected is, Polyvinyl chloride (PVC). b) Pipeline's diameter range is between 15 to 60 cm. c) The high and low-velocity limits are assumed 60-300 cm/sec. d) Depth of burial of pipelines according to frost depth considered within 90 to 400 cm. e) The slope for piping was set among 0.005 and 0.1 percent. f) All pipelines are considered half-full (50%) flow. The created model has been validated to evaluate the input data and design settings scenario. while the validation process was completed, the model was run to perform hydraulic analysis. The software measures and analysis numerical and analytical variables.

Presenting and Evaluating the Results

Different techniques are available to display the analysis results in SewerGEMS software. The *"design report"* illustrates a summarized overview of the analyzed network. The information such as specifications of the designer, title of the project, name of the organization, date of analysis, and design. It also gives brief information about the applied scenario in the creation of the model. It represents the exact amount of each element utilized in the network. In this study, a total number of 521 pipelines and manholes adding one outfall are being used. Furthermore, it gives the lengths and diameters of all diameter ranges of the pipelines used in the model (Figure 5). A total length of 92602.4 m pipelines were used in this modeling.

Hydraulic Model Inventory: 5.stsw

Title	مدل سازی شبکه جمع آوری فاضلاب ناحیه پنجم شهر کابل
Engineer	پوهنیار علی رضا نوری
Company	پوهنتون پولی تخنیک کابل
Date	7/26/2019
Notes	

Scenario Summary	
ID	1
Label	Base
Notes	
Active Topology	Base Active Topology
User Data Extensions	Base User Data Extensions
Physical	Physical Alternative - 3
Boundary Condition	Base Boundary Condition
Initial Settings	Base Initial Settings
Hydrology	Base Hydrology
Output	Base Output
Infiltration and Inflow	Base Infiltration and Inflow
Rainfall Runoff	Base Rainfall Runoff
Water Quality	Base Water Quality
Sanitary Loading	Base Sanitary Loading
Headloss	Base Headloss
Operational	Base Operational
Design	Base Design
System Flows	Base System Flows
SCADA	Base SCADA
Solver Calculation Options	Base Calculation Options

Network Inventory			
Conduits	521	Taps	0
-Circle	521	Transitions	0
-Box	0	Cross Sections	0
-Ellipse	0	Outfalls	1
-Virtual	0	Catchments	0
-Irregular Channel	0	Low Impact Development Controls	0
-Trapezoidal Channel	0	Ponds	0
-Triangular Channel	0	Pond Outlet Structures	0
-Rectangular Channel	0	Headwalls	0
-Pipe-Arch	0	Pumps	0
Laterals	0	Wet Wells	0
Channels	0	Pressure Junctions	0
Gutters	0	SCADA Elements	0
Pressure Pipes	0	Pump Stations	0
Catch Basins	0	Variable Speed Pump Batteries	0
Manholes	521	Air Valves	0

Circle Inventory			
Circle - 150.0 mm	80,508.8 m	Circle - 375.0 mm	1,405.3 m
Circle - 200.0 mm	3,732.6 m	Circle - 450.0 mm	2,145.7 m
Circle - 250.0 mm	2,159.0 m	Circle - 600.0 mm	881.0 m
Circle - 300.0 mm	1,770.1 m	Total Length	92,602.4 m

Figure 4. Summary of hydraulic design

It is possible to display separate tables in the software and observe the design details to see more information about manholes and pipelines. The materials utilized in the design of the pipelines, the length of each branch, and all the required figures can be displayed in the program report. Another method that can be used to see the results is diagram illustration. The graph representation in terms of time in various model components can be used to explain variation in the component in the context of a map. The adjustment in "hydraulic gradient" in a manhole over the course of 24 hrs is expressed in (Figure 5). If the extended duration approach is utilized in the design case, this technique yields a concrete outcome.

The sewer network profile is a crucial component that should be displayed. The profile of all pipelines is essential for the implementation of sanitation projects. The critical path, also known as the farthest path, can be seen in (Appendix A&B). The profile can include surface level, manhole depth, invert altitude, pipeline slope, width, and other required parameters.



Figure 5. Hydraulic Gradient for J-1

Control of Parameters

Some modules may be used as control parameters to consider the correctness and precision of the design and hydraulic analysis. For example, speed of flow (velocity) is a component that can be examined. Based on the design requirements, sewage network velocity limits vary from 30 to 500 cm/sec with an ideal range of 60 to 180 cm/sec. As a result, it can be said that in the wastewater collection modeling, the velocity should be within the applicable limits, if not the optimal height. The velocity of flow in all pipelines reported within the scope of 30 to 487 cm/sec, as seen in (Appendix C). Like pipeline diameters added as usual from the software library, other variables can be monitored, checked, and measured according to the flow volume. After running the program, all diameters were chosen based on typically available pipe diameter, as seen in the diameter column. SewerGEMS can also be used to verify variables linked to manholes (Appendix D). The ground level and the depth of the manholes, or invert altitude, are two important criteria to consider in manholes.

CONCLUSION

This study aimed to see whether decentralized wastewater collection schemes or regional networks were feasible. Consequently, the 5th district of Kabul city has been chosen as a research location, and two out of nine divided zones within it were investigated. The research area's LULC and topographic map were modified and generated in the Arc GIS setting using Satellite Photo and DEM. Different shapefiles for pipelines and manholes were created in the Arc map framework to develop the architectural model. The developed architectural model is imported to SewerGEMS software and adapted for a hydrological model.

SewerGEMS software was used to perform the model's hydraulic assessment. The geometric modeling and hydraulic analysis results show that decentralized and regional wastewater collection systems are viable options for Kabul city and look appropriate options. Nevertheless, after developing many simulations of various scenarios, the hydraulic model design at the 5th district level for two research zones by the SewerGEMS program was eventually completed based on maximum discharge. It can be inferred that the hydraulic model was successfully developed and operated after analyzing the hydraulic analysis findings and evaluating control parameters like flow rate, the diameters of pipelines, slopes, pipeline's profiles, and so on. Such models can be applied to city sizes at the district level. A large portion of Kabul city was built without a master plan. The lack of a consistent urban plan makes it difficult to create such models; this analysis aims to include additional tools in the gravity system for both planned and unplanned regions. The utilization of *"wet wells"* and *"lift stations"* will primarily overcome the system's dilemma.

REFERENCES

Afzali, A. O. (2017). Physicochemical and Bacteriological Quality of groundwater in Kabul basin, Afghanistan. IOSR Journal of Agriculture and Veterinary Science, 10(03), 58–61. https://doi.org/10.9790/2380-1003025861

Akshaya, N., & Jesse, D. (2020). Using ModelBuilder to Import External Data. Bentely Comunities.https://communities.bentley.com/products/hydraulics___hydrology/w/hydraulics_and_hydrology__wiki/2821/using-modelbuilder-to-import-external-data

Aleisa, E., & Alshayji, K. (2019). Analysis on reclamation and reuse of wastewater in Kuwait. Journal of Engineering Research, 7(1), 1–13.

Ali, H. (2017). BBC. https://www.bbc.com/persian/afghanistan-40787693

Bhave, P., & Rahate, S. (2018). Impact of Redevelopment Projects on Waste Water Infrastructure. Journal of The Institution of Engineers (India): Series A, 99(3), 503–509. https://doi.org/10.1007/s40030-018-0303-x

Brati, M. Q., Ishihara, M. I., & Higashi, O. (2019). Groundwater level reduction and pollution in relation to household water management in Kabul, Afghanistan. Sustainable Water Resources Management, 5(3), 1315–1325. https://doi.org/10.1007/s40899-019-00312-7

Bu Qammaz, A., AlHamed, S., AlNasser, F., & Alkhalidi, M. (2020). Integrated Coastal Zone Management in Kuwait: A Knowledge-Based System Model. Journal of Engineering Research, 8(4), 29–44. https://doi.org/10.36909/jer.v8i4.7931

Chawla, A., & Singh, S. K. (2014). Modelling of Contaminant Transport from Landfills. International Journal of Engineering Science and Innovative Technology, 3(5), 222–227.

CIA. (2020). The World Fact Book. https://www.cia.gov/library/publications/the-world-factbook/geos/af.html

Consortium for DEWATS Dissemination Society. (2017). Technical assessment of faecal sludge management in Kabul (Issue November).

Dan, I., & Jesse, D. (2020a). How do each of the LoadBuilder methods work? Bentely Comunities.https://communities.bentley.com/products/hydraulics___hydrology/w/hydraulics_and_hydrology__wiki/10862/how-do-each-of-the-loadbuilder-methods-work

Dan, I., & Jesse, D. (2020b). Using Terrain Extractor (Trex). Bentely Comunities. https://communities.bentley.com/products/hydraulics___hydrology/w/hydraulics_and_hydrology__wiki/8281/using-terrain-extractor-trex

Frahmand, A. S. (2011). GIS based Ground Water Quality Analysis in Kabul City Region. Geospatial World Forum, 1–13. https://www.researchgate.net/publication/284352960 GIS_based_Ground_Water_Quality_Analysis_in_Kabul_City_Region

Gesim, N. A., & Okazaki, T. (2018). Groundwater Quality Mapping Using Water Quality Index in the Kabul City , Afghanistan. August, 29–34.

Hayat, E., & Baba, A. (2017). Quality of groundwater resources in Afghanistan. Environmental Monitoring and Assessment, 189(7). https://doi.org/10.1007/s10661-017-6032-1

Houben, G., & Tünnermeier, T. (2005). Hydrogeology of the Kabul Basin Part II : Groundwater geochemistry and microbiology. BGR.

Kanakiya, R. S., Singh, S. K., & Shah, U. (2015). GIS Application for Spatial and Temporal Analysis of the Air Pollutants in Urban Area. International Journal of Advanced Remote Sensing and GIS, 4(1), 1120–1129. https://doi.org/10.23953/cloud.ijarsg.102

Kumar, A., Mishra, R. K., & Singh, S. K. (2015). GIS Application in Urban Traffic Air Pollution Exposure Study: A Research Review. Suan Sunandha Science and Technology Journal, 2(1January), 25–37.

Noori, K. M. A., & Nasimi, M. N. (2019). Kabul City Groundwater and Need for Artificial Recharge. Proceedings of the 4th International Conference on Civil, Structural and Transportation Engineering (ICCSTE'19), 1–7. https://doi.org/10.11159/iccste19.215

NSIA. (2020). Estimation-Population-Final-Year 2020.

Paiman, Z., & Noori, A. R. (2019). Evaluation of Wastewater Collection and Disposal in Kabul City and Its Environmental Impacts. Modern Environmental Science and Engineering, 5(5), 451–458. https://doi.org/10.15341/mese(2333-2581)/05.05.2019/012

Rhimi, O., & Murakmi, K. (2017). Rooftop rainwater harvesting and its efficiency in Kabul new city. 土木学会論文集G（環境）, 73(5), 25–30. https://www.jstage.jst.go.jp/article/jscejer/73/5/73_I_25/_pdf/-char/ja

Seddeqi, M. Q. (2017). Water Supply Engineering (2nd ed.). Nawesa.

Shafaqna. (2019). shafaqna. Shafaqna News Webpage. https://af.shafaqna.com/FA/341927

Zaryab, A., Noori, A. R., Wegerich, K., & Kløve, B. (2017). Assessment of water quality and quantity trends in Kabul aquifers with an outline for future drinking water supplies. Central Asian Journal of Water Research, 3(2), 3–11. http://www.water-ca.org/article/1925

Zhang, D., Martinez, N., Lindholm, G., & Ratnaweera, H. (2018). Manage Sewer In-Line Storage Control Using Hydraulic Model and Recurrent Neural Network. Water Resources Management, 32(6), 2079–2098. https://doi.org/10.1007/s11269-018-1919-3

Automatic Recognition of License Plates

1st Tanushri Bhagat
Electronics and Communications Department
Delhi Technological University
Delhi, India
tanushribhagat_2k20spd14@dtu.ac.in

2nd Rahul Thakur
Electronics and Communications Department
Delhi Technological University
Delhi, India
rahulthakur@dtu.ac.in

2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI) | 978-1-6654-1559-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICETCI51973.2021.9574072

***Abstract*—As our lives are getting automated with the Internet of Things, Machine Learning it is very important to get accurate results. This paper is a study of Automatic License Plate Recognition (ALPR) using various Python libraries and hence recognize the information on images and videos. We will discuss a novel image processing method which will be used for the recognition of Indian Number Plate Recognition in videos. Further, it will explain about the need of this system in India and also the importance of high intensity images so as to detect them. Also, some future prospects are proposed which will enhance the current working model keeping intact the security and privacy of the data of people. It can later be linked with the database of VAHAN and SARTHI for quick responses regarding the owner of the vehicle and other necessary information. The ALPR system is detected using Local Binary Pattern algorithm and is subsequently recognized using HTTP server using beanstalkd queue thus making a complete software implementation.**

***Index Terms*—license plate detection, image processing, automatic license plate recognition (ALPR), Local Binary Pattern.**

I. Introduction

Automatic License Plate Recognition is a software which is used to perform Optical Character Recognition on images of cars which contains the license plate. The information is supplied in the form of an image to the next stage of computer processing and then it can be translated, compared and stored to create an application deployed using ANPR. ANPR is a specialised tool which anticipates assistance to the imposition of the law and order, thus ensuring the continued safety of all people driving on the road. Safety and security of all the users has been braced by ANPR throughout the world and is strengthening the effective transport flow of people in real time.

The simulation used in this project takes into input an image and a video file, detetcts the characters present on it using the LBP algorithm [1] and then recognizes all the numerals present according to the beanstalkd queue and then returns the contents of the number plate/s as per the no. of license plates available along with their confidence value and dimensions of the number plate and various other parameters desired by the user. This system works fine for the recognition of license plates in varied environments and dimensions. It is also capable of detecting low contrast images to a certain extent.

The current stumbling block with the existing system is that this automated number plate recognition works fine for controlled environments only. To improvise this, the paper will propose a new method that will work for low illuminated and noisy images. The training of the openalpr library in python was done by training the system for a large set of manually clicked data. This system can be made smart by incorporating an IoT system in it, so that we can auto alert the nearby police station for notification of wanted, listed and stolen cars and take action at the earliest. With the help of the datasets of SARTHI and VAHAN the search for the cars can be accelerated this will make the system reliable and fast [2].

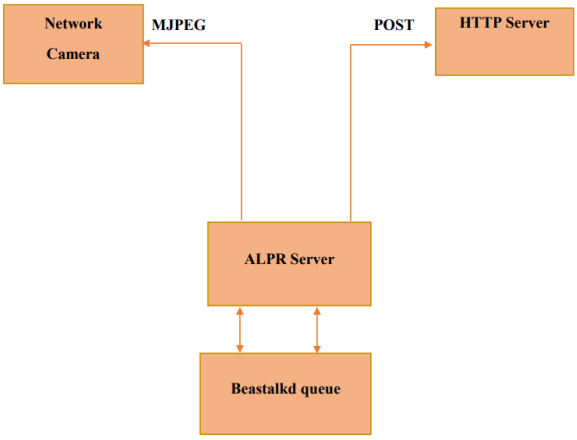


Fig. 1. Architecture of the Algorithm.

II. Literature Survey

To date, fuzzy logic and neural networks together are being used for the detection in Dutch number plates. Segmentation is performed using Fuzzy logic and for extracting and classifying the features present in the image or video file, neural networks are being used. Furthermore, in the case of distorted background, character extraction utilizing the wavelet transform method is used. Segmentation is done using an algorithm and then distinguishing between different characters in the number plate is done using Optical Character Recognition[3]. Furthermore, in advanced research, Morphological operations for plate extraction and character recognition can be used. This approach works fine for the recognition of license plates in varied environments and dimensions. It is also capable of detecting low-contrast images to a certain extent. It was

Authorized licensed use limited to: DELHI TECHNOLOGICAL UNIV. Downloaded on December 09,2021 at 10:55:00 UTC from IEEE Xplore. Restrictions apply.

also observed that in the latest developments, number plate detection uses the Support Vector Machines (SVM) algorithm [4].

After the segmentation of each character, Artificial Neural Networks (ANN) are used for recognizing each element. Additionally, for the segmentation of number plates, the border of the image is taken into consideration and then the contours are drawn on it. Now based on the spatial localization and dimensions of the characters filtration of contours is performed. Finally, the character recognition is done by using the algorithm based on K-nearest neighbor after the region of interest filtering and De-skewing.

In the paper written by Panahi [5], the aim was to detect and recognize the Persian Number Plate System. The localization of the license plate was performed using Random Sampling Consensus (RANSAC) algorithm, which is used to form the output data from the best mathematical model. For the recognition of characters, the number of active pixels is calculated which represents the highest intensity values. These pixels are used for calculating the distance (both shortest and longest) from the walls. This feature is used to recognize the vehicle plate and Principal Component Analysis is used as a reduction method.

In [6], the images were recorded with the help of an IR camera and an ALPR system. The colour images were converted into grayscale to calculate the vertical edges and then a Sobel filter was applied for horizontal derivative and noise reduction. Then the segmentation of the image was carried out using Threshold filter. Thereafter, Morphological operations were performed. The area and aspect ratio of the part left out has been done with the help of contouring. Finally, Support Vector Machine (SVM) Algorithm has been used to train the data and enhance the output results.

In [7], Noise was reduced in the image by the process of Smoothening. The gray scale image was converted into binary image using Binarization and Dilation was used for noise removal. The surrounding masked image was created using minimum containing rectangle. The license plate candidate areas have been identified with the aspect ratio range allowable. Lastly with this procedure data was trained to observe the results.

In recent years, License Plate has been an engaging field for research. Various models and methods have been proposed on different methods like morphological extraction, colour segmentation, colour modeling, wavelet transform, neural network, and texture feature analysis of gray-scale images[8]. For Chinese number plate detection, extraction of number plates has been done using hue-saturation intensity and colour edge mapping-based transformation methods. The characters have been identified using OCR techniques [9]. Theoretically, it is believed to be 90% accurate but following the simulation method used in this paper, the accuracy is 65% were the noisy, blurred, and skewed image is not being taken into consideration. In Argentina, the number plates are matched using intelligent template matching.

In India, the number plate used for 4-Wheeler's is of the format "MP 04 CL 7704" where the first 2 characters are alphabets depicting the state of registration of the vehicle. The second 2 numerals are segregating the RTO districts and the cities numbered according to the capital of the state and so on. The third pair of characters portray the title of the record register book and the last numeric represents the entry number in the directory.

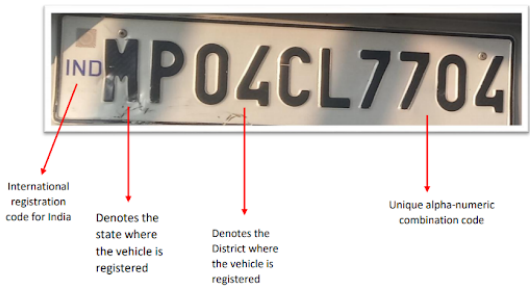


Fig. 2. Number Plate Description.

III. Methodology

The aim of this project is to take the image as an input, perform various image processing operations on it and then finally detect the number plate data and display it in the form of text to the user. The various processes occurring at the back end is explained herewith and the output data can be saved in the form of a JSON/CSV file for data exchange. The process continues in the form of a pipeline and the various steps are-

A. Detection

For starting the detection in the case of a video, first we need to extract frames from the video. After extracting the frames, for each frame, the Local Binary Pattern (LBP) algorithm is used which aims at finding the predefined plate of 'x' height and 'y' width. This is generally the most time taking and processing-intensive phase in the whole process. Local Binary Pattern- In this all the values of the 3*3 pixels are thresholded by the centre pixel. If the neighbouring value of the pixel is less than the centre pixel then it is given as 0 else the value is given as 1. All of them are combined to form a 8 bit value and this converted into the decimal is known as the LBP feature.

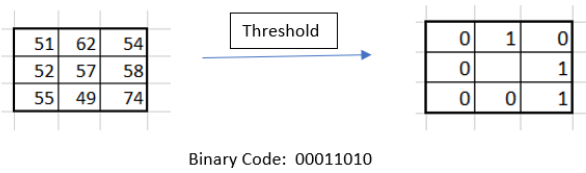


Fig. 3. Local Binary Pattern.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:55:00 UTC from IEEE Xplore. Restrictions apply.

B. Binarization

This will occur for all the plates which were achieved after the first step. Using the Wolf-Jolion method [10], this stage will form multiple binary plates for a single plate detected. It generally helps to detect all the characters properly in case of noise, improper sunlight, etc. Wolf-Jolion Method- This is done by taking into account the dilation and conditional erosion for neighbouring pixels. If the base pixel and the neighbouring pixel combined does not exceed the given threshold, then only dilation is performed. The conditional erosion is done on the basis of gray levels of the current and the neighbouring pixels [11].

C. Character Analysis

The third step in this process aims at localizing all small-size characters which may be present in the plate region. After this, it aims to find large size characters in the region. If suppose no character is found, it will discard this plate and move on to the next.

D. Edge Detection

The detection phase is supposed to detect a region that must be slightly bigger than the plate area so as to reduce the chances of missing any character. Now, this is the precision step in which we use the configurable weights method to determine the significant edges.

E. De-Skew

This step focuses on converting the image into an ideal plate of standard height and width and also removing the rotation and skewness if any.

F. Character Segmentation

This step tries to extract all the characters that may be present in the plate area. It also removes regions which aren't qualified to be under the category of characters and precisely mark the edges.

G. Optical Character Recognition

This step aims at the electronic conversion of data in the image in form of text. Also, it suggests the characters with their respective confidence percentage enabling better results [12].

H. Post Processing

This is a combination of all the characters which were earlier recognized by the OCR. Now it combines them into a complete license plate and then specifies the confidence percentage of each plate.

Thus, this way we can select the plate with the most confidence percentage at the end, and in the case of video, it detects all the plates which are present in the video.

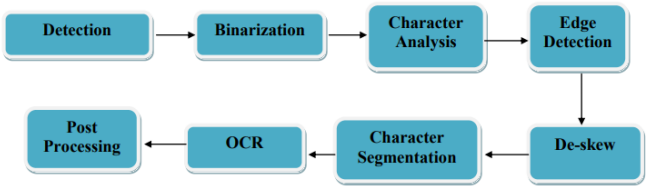


Fig. 4. Block Diagram.

IV. Simulation and Output

The coding was done on Python and several classes were imported from Python to do the coding. The various libraries imported were-

1. openalpr- It is an open source code written in C++ and it has bindings in Python, Java, C etc. These bindings provide a way to extract and use the file or code into different software (generally by mapping one thing to the other) [13]. This software is used to- recognize license plates from camera streams, Analyze still images from command prompt, Integrate license plate recognition into your application.
2. argparse- This library makes it easy to write user-friendly command-line arguments.
3. Open-cv- This particular library in Python is used for performing computer vision problems, machine learning, and image processing operations.
4. Numpy- It is used for working with arrays. It has functions for working in the domain of linear algebra, Fourier transform, and matrices. It was used because lists in python were really slow.
5. JSON- This is used for the storage and exchange of data through the web. It is text, written with Javascript object notation [14].
6. Itertools-They are a core set of fast, memory-efficient tools that allow us to construct and return iterations.

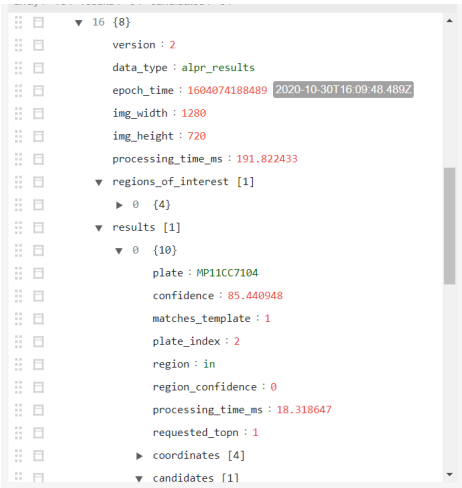


Fig. 5. JSON file.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:55:00 UTC from IEEE Xplore. Restrictions apply.

The code attached can be used for correct number plate detection in case of an image and video for all the number plates in use in India. The size of the number plate and its font and shape are already used for training this module and hence, we can see that the output can be read via text.

The JSON file (Fig 5.) that is used for the transfer of data on the web. The complete data can be transferred from the web. In our code, the Json file will tell us the options of all the number of plates detected with a confidence of more than 65%. Moreover, all the coordinates of the number plate are specified and all the possible options are stated.

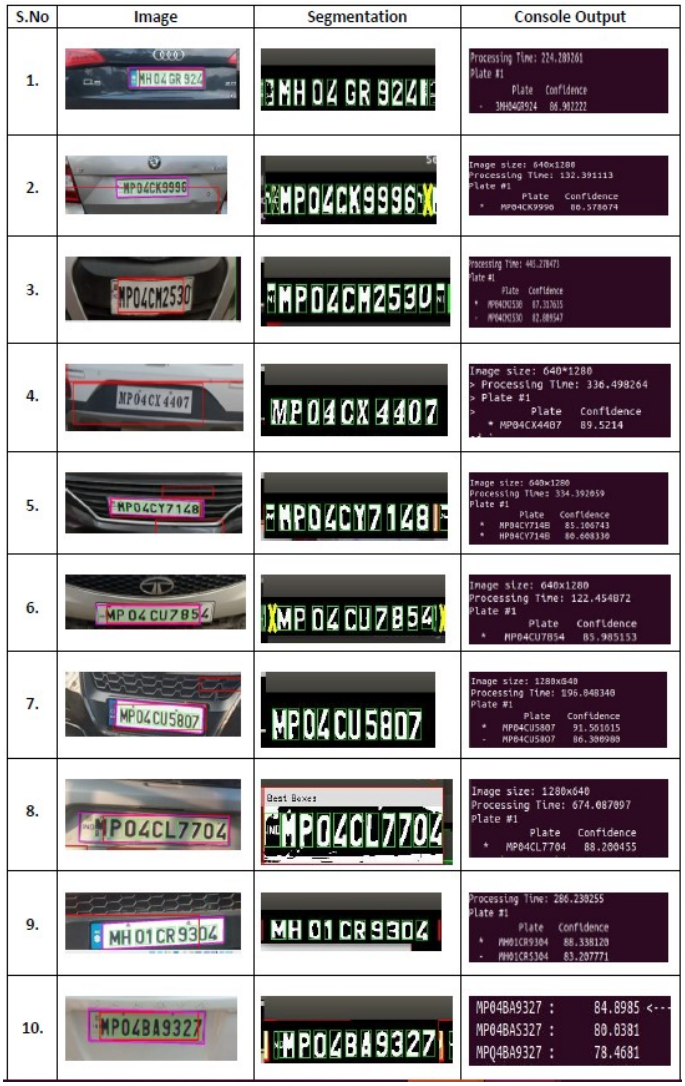


Fig. 6. Frame extraction and license reading from video.

The table attached (Fig 6.) is a compilation of the images taken from a video recorded from a moving car in a parking lot. The mobile used for recording this video was using 30 frames per second. The first row depicts the plate detection in the video followed by the recognition of the character using Segmentation. In the last row, the console results are attached which depicts the output parameters. Also, it has been noticed that in the figure above the number plate the blue rectangle depicts the number plate the system has detected. The recognition was done at every 10Th frame. This saved a lot of memory and also increased the speed and performance of the code. In comparison with the latest studies, it has been found that the efficiency of this code is approximately 88% which is at par with the existing methods.

The performance analysis (Fig 7.) for this project was calculated considering all the above images in the table. While the code simulation, the output consoles displays the Confidence Percentage of all the individual letters and the total Confidence Percentage is the mean of all the letter and numbers present in the number plate. Also, the time required for detection and recognition of the license plate is termed as processing Time and the graph is plotted considering both the parameters and hence the efficiency of the code is determined.

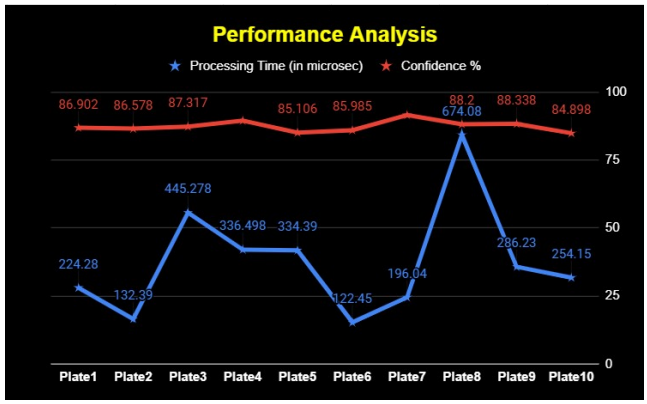


Fig. 7. Performance Analysis.

V. Need for ALPR

1. Highway Monitoring- In India there is one death in every four minutes with most of them occurring due to over speeding. ANPR is employed to watch the vehicles' average speed and may identify the vehicles that exceed the regulation [15]. In this case, a fine ticket is often automatically generated by calculating the space between two cameras. This helps to take care of law and order which, in turn, can minimize the amount of road casualties.

2. Law Enforcement- It can be used for taking the criminals into custody and exonerating the innocent. Due to the automatic nature of this system it allows continuous tracking of the information from different sources and then it is coordinated and enhanced by a group of analytical tools to store the data in an organized, safe, feasible and convenient database.

3. Recovery of Stolen Vehicles and number plates- On an average around 2 Lakh cars are stolen every year. ALPR system can be used for continuous plate after plate detection against the active lists of stolen vehicles and vanished number plates [16], sending automatic alerts and notification instantaneously to an officer on duty in an immediate area who can suspect and monitor the situation.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:55:00 UTC from IEEE Xplore. Restrictions apply.

4. Missing Person- In the US, this system is equipped with alerts, so that the officer on duty in the required area can be alerted as soon as possible and necessary actions could be taken.
5. Surveillance and Security- License plate reader data can help locate potential witnesses based on the last available location of the vehicle and can assist in validating an alibi for testimony.

VI. Challenges of ALPR

1.Variability- The first step in the detection of the license plates is its recognition in the image, but the variable plate size in India makes it difficult to generalize the algorithm.
2.Divergent Font- At present, there are 50 styles of the font used in the Indian number plate system. Adding to the problem some number plates are even handwritten. So, to make an algorithm that would be able to detect all these types is very difficult [14].
3.Quality of Camera- The camera installed for the purpose of detection on roads is generally of low quality. This adds to the accuracy problem in the detection of the plates. These cameras have limited visual coverage and lead to an improved number of blurred images.
All the above challenges are written taking into account only the car number plates in the country. Two-wheeler and Three-wheeler have a double line number plate. At this point in time, we need a generalized and standardized number plate so as to automate our requirements.

VII. Conclusion and Future Scope

While working on this project our software was tested with an IP camera which had the capability of video stream at 30fps FHD resolution 1080p and 10 frames were used per second keeping the time lapse at 4x, which is approximated to 0.12-second interval and the achieved number plate detection is 93% as per the software output and character recognition rate of 90% in worst condition including camera motion blurring, vehicle motion, and lighting condition. All the output achieved was performed on Python on Linux [18]. This system is currently able to detect the number of plates in images as well as videos with 30fps. In all the tests conducted during this experiment, it was observed that the current number plate detection fails while detecting non-standardised as two line number plates. Since standardization is difficult at this point of time for the vehicles already on road, we can train the existing model for maximum number of fonts and formats so that the confidence percentage of detection improves significantly. Moreover, we need permission to link the current system to the database of VAHAAN and SARTHI [19] so that we can link both of them. Also, we can connect all the cameras in the city as well as highways with this model to enable fast tracking of vehicles in case of crimes and traffic monitoring.

References

[1] Bai Hongliang. A hybrid license plate extraction method based on edge statistics and morphology In Pattern Recognition, ICPR 2004 proceedings, Vol. 2, 831 - 834, 23-26 August 2004.

[2] J. Barroso, A. Rafael, E. L. Dagless and J. Bulas- Cruz. Number plate reading using Computer vision. IEEE -International Symposium on Industrial Electronics ISIE'97,Universidade do Minho, Guimar˜aes, Portugal, Julho, 1997.

[3] X. Zhai, F. Bensaali and S. Ramalingam, "License Plate Localisation based on Morphological Operations", The 11th International Conference on Control, Automation, Robotics and Vision, pp. 1128-1132,Singapore, December 2010.

[4] Samabia Tehsin, Asif Masood, Sumaira Kausar,287520316_Survey_of_Region-Based_Text_Extraction _Techniques_for_Efficient_Indexing_of_ImageVideo_Retrieval.

[5] Panahi, Rahim, and Iman Gholampour. "Accurate Detection and Recognition of Dirty Vehicle Plate Numbers for High-Speed Applications." IEEE Transactions on Intelligent Transportation Systems (2016).

[6] Kumar Parasuraman, Subin P.S "SVM Based License Plate Recognition System "International Conference on Computational Intelligence and Computing Research2010 IEEE.

[7] Aiswarya Menon, Bini Omman "Detection And Recognition of Multiple License Plate From Still Images" presented at the IEEE International Conference on Systems, Man and Cybernetics, 2004.

[8] X. Pan, X. Ye and S. Zhang, "A hybrid method for robust car plate character recognition," presented at the IEEE International Conference,2018.

[9] Debkumaar Chowdury, Soumya Banerjee, "An Adaptive Technique for Computer Vision Based Vehicles License Plate Detection System," presented at the IEEE International Conference, 2019.

[10] Shigeharu Miyata, "Automated License Plate Detection using a Support Vector Machine," presented at 14th International Conference on Control, Automation, Robotics Vision Phuket, Thailand, 13-15th November 2016.

[11] Survey_of_Region-Based_Text_Extraction_ Techniques_for_Efficient_Indexing_of_ImageVideo_Retrieval.

[12] N Otsu.A threshold selection method from gray-level histogram[J].IEEE Transactions on SMC,1979,9:62-69.

[13] Charl Coetzee, C. Botha, C. Weber, D. Trier, "PC Based Number Plate Recognition System", Proceedings of IEEE International Symposium on Industrial Electronics, vol. 2, pp. 605 - 610, July 1998.

[14] Prathamesh Kulkarni, Ashish Khatri, Prateek Banga, Kushal Shah, Automatic Number Plate Recognition (ANPR) System for Indian conditions.

[15] http:doc.openalpr.com/accuracy_improvements.html#: :text=recognizing %20license%20plates.-,OpenALPR%20Design,plate%20numbers%20in %20the%20image.text=Transforms%20the%20perspective%20to%20a, the%20ideal%20license%20plate%20size.

[16] M. Sarfraz, M. J. Ahmed, and S. A. Ghazi, "Saudi Arabian license plate recognition system," in Proc. Int. Conf. Geom. Model. Graph., 2003, pp. 36–41.

[17] D. Zheng, Y. Zhao, and J. Wang, "An efficient method of license plate location," Pattern Recognit. Lett., vol. 26, no. 15, pp. 2431–2438, 2005.

[18] E. R. Lee, P. K. Kim, and H. J. Kim, "Automatic recognition of a car license plate using color image processing," in Proc. IEEE Int. Conf. Image Process., vol. 2. Nov. 1994, pp. 301–305.

[19] O. Martinsky, "Algorithmic and mathematical principles of automatic number plate recognition systems," B.Sc. thesis, department of intelligent systems, faculty of information technology, Brno University of Technology (2007).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:55:00 UTC from IEEE Xplore. Restrictions apply.

Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

BNDNN: Batch Normalization Based Deep Neural Network for Predicting Flood in Urban Areas

Vinay Dubey
Delhi Technological University
Rahul Katarya (rahuldtu@gmail.com)
Delhi Technological University https://orcid.org/0000-0001-7763-291X

Research Article

Keywords: Artificial Neural Network, Batch Normalization, Decision Tree, Deep Neural Network, Flood prediction, Support Vector Machine

Posted Date: November 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1027530/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

BNDNN: Batch Normalization based Deep Neural Network for predicting flood in urban areas

Vinay Dubey, Rahul Katarya
Big Data Analytics and Web Intelligence Laboratory, Delhi technological University, New Delhi, India
vinaydubeyamit@gmail.com, rahuldtu@gmail.com

Abstract

Disaster is a very serious dissipation that arises for a short time period, but the impact of that disaster on human society is very dangerous and very long-lasting. Disasters are categorized into two types like natural disasters and manmade disasters. Among all disasters, of all the natural disasters, flood is the commonplace natural disaster. Flood disaster that causes huge loss of human life, diversity as well as economic loss, which is very dangerous for the developing countries and developed countries also. Nowadays during the monsoon season flood is dangerous for all the geographical areas located nearby water bodies. Much research has been done for flood detection. Machine Learning and many other recent technologies are playing a vital role in predicting the occurrence of floods. For prediction purposes, a huge amount of data is requiring collected from sensors deployed in various locations. In this paper, we used the Batch normalization with Deep Neural Network (BNDNN) technique for the classification of data in three classes as Low, Moderate, and High. The result obtained from our proposed model is compared with some other models like Decision Tree (DT), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Deep Neural Network (DNN). In this our proposed BNDNN provides 89% accuracy which is higher among all existing models. Models are compared based on some parameters like Accuracy, Precision, Recall, F – Score. The compression among all the models used in this paper shows that our proposed model provides better results.

Key Words: Artificial Neural Network; Batch Normalization; Decision Tree; Deep Neural Network; Flood prediction; Support Vector Machine

1. Introduction

Disaster is a very serious outbreak that is of very short duration but has a huge impact on human life and nature. According to the united nation, a disaster is

A disaster can be defined as a “serious severance of a biodiversity or society's working due to widespread human or environmental damage”.

Nowadays, technologies are sufficient to predict the severity of the disaster such as flood and also detect disaster in inaccessible areas. Flood risk residential life and property, altering the natural environment, profound impact on biodiversity, polluting water bodies and ecosystem [Hirabayashi et al., 2013, Dottori et al., 2016] Nevertheless, a huge number of studies and explorations are done on sensor technology and artificial intelligence. The combination of these vital areas provides immense business value and a huge number of conveniences for burgeoning artificial intelligence technology and big data analytics. Monitoring discrete events is important, but our main goal is to process data collected from deployed sensory sources. Change in the climate of the earth is contributing to increasing weather-related events. That encounters the increment in natural disasters like floods, earthquakes, and tsunami, etc. [Mosavi et al., 2018]. Out of all the disasters, flood is the most frequent one. The reason behind floods is sometimes natural such as excessive rainfall or cloudburst or sometimes man-made such as encroachment of water bodies in the catchment areas.

Flesh flood disaster is referring to loss that is influenced by the flood in the form of causalities, infrastructure loss, and diversity loss as well as economical loss. For developing countries, any natural disaster is very destructive [Li et al., 2019]. Flood disaster very commonly occurs in natural disasters among all-natural disasters such as earthquakes, volcanoes, tsunami, and landslides. The impact of these disasters on human life is very extensive [Liu et al. 2019]. According to a survey, statistical economic loss due to the flood annually around the world for 50 million dollars approximately. The people concerned by the flood are approximate 100 million [Ahmadisharaf et al. 2015]. However, detection and prediction of the flooded area are required time and it is a very complex task due to the dynamically change in the climate stage. So thus nowadays mostly data-oriented flood prediction models are developed and concern in very uncomplicated presumptions [Aazam et al. 2014]. The unceasing improvement of Machine Learning approaches over the past decades determined their appropriateness for flood detection with the plausible rate of outperforms conventional approaches [Saeed et al. 2018, Ranit and Durge et al. 2019]. We have used DNN to classify our input data collected from different sensors. We add batch normalization techniques between layers to improve the classification accuracy of our DNN model. The accuracy of our model is better than the accuracy previously used for flood prediction.

In the current section of the paper, we provide the basic introduction of the flood disaster as well as some statistical data regarding the flood. These sections also highlight the basic information of the model which we use in our paper. Paper flow is in section 2, we provide a literature survey. In this section, we discuss the previously used model and algorithm used in flood prediction. In section 3, We have discussed some of the standard algorithms and models used earlier in flood prediction. In section 4, we discuss our proposed model for flood prediction. In section4, we provide information regarding the

data set we used in our paper as well as compare our model with some other models. In section 5, we provide the conclusion of the paper as well we the future scope.

Our major contribution in this paper:

- In this paper, we introduce the new approach, where the batch normalization approach combines with DNN for meat flood event prediction.
- The accuracy of our model is also compared to some existing standard models and algorithms and our proposed model provides better prediction accuracy.
- In this paper, we consider datasets with different environmental parameters, such as cloud cover, precipitation, average. temp, min. temp, max. temp. along with this parameter year, the month is also included.

2. Related work

For reducing the risk of possible damage caused by the disaster government all over the world focus on developing the Early Warning System (EWS). The author develops an ML method for monitoring flood areas. The AI-based system is designed for detecting abnormal behavior of dike. The Bishop model was used to calculate the Dike behavior and used as a target for the neural network [Pyayt et al., 2011]. Prediction of drainage of water is crucial for flood control during heavy rainfall. The proposed work focuses on the predictive discharge BPNN model. Data is collected from Govindpur basins on the Brahmani River on daily basis. Models have been Trained and tested to predict model performance over different architectures [Ghose et al. 2018]. The author uses the Hadoop file system to maintain a huge amount of data collected from sensors. a Convolution DNN for the analysis of data. Model is tested on data collected from IoT-based devices. Results were compared with the existing ANN and DNN models. [Anbarasan et al. 2020]. IoT is an area of applied electronics that is deals with collecting data in a real-time environment and transfer via a wireless sensor network. Data collected from the various sensors is analyzed through an ANN. In this proposed model standard three-layer architecture is used with one hidden layer [Mitra et al. 2016]

The study area in this paper is disaster risk analysis in the Philippines due to cyclones. The researcher in this study is the focus to improve the output of the hybrid model. This model uses ARIMA and ANN with DWT for disaster risk prediction in the Philippines provinces in terms of casualties [Alquisola et al. 2018]. The existing tool for data classification is mostly used manual data input and prepared data. The author proposed a unique approach for flood forecasting by integrating geospatial, Hydrological, Metrological, and crowdsources that big data in an appropriate Machine Learning framework. Results suggest that MLP ANNs provide better accuracy among all ML techniques [Puttinaovarat et al. 2020]. Flood occurrence is depending on various

metrological and hydrological factors. An ANN-based model aims to intensify the accuracy of the flood detection system. Data from various sensors fed to the ANN for analysis. Experimental results show that the NARX network with Levenberg- Marquardt training algorithm provides better results and real-time prediction with alert [Bande et al. 2017].

In many cases, ANN is providing better prediction accuracy. Japan is the very venerable geographical area for floods, which facing an increase in water level after Typhoon. In this paper, the author used various datasets. This work aims to select the most relevant dataset for the ANN-based water flow forecast model [Kim et al. 2016]. The study area in this paper is the Pampanga river basin. The hardware setup contains a microcontroller, solar penal, ultrasonic sensor, and GSM module. A Feedforward network with backpropagation is used and for optimizing the network Levenberg- Marquardt training algorithm is used [Sahagun et al. 2017]. In some cases, a tree-based ML model is also used to predict the sensitivity of flooded areas based on the spatial parameters [Lee et al. 2017]. A novel framework is design by the author to detect the flooded area and depth of flood by using ML measures and a hybrid of hydraulic models. To achieve this a two-dimensional hydraulic model (iRIC), calibrated by continuous water level data, is used to estimate river depth by applying a two ML model to the domain for an arbitrary discharge. is done to train. [Hosseiny et al. 2020]. Nowadays IoT technology is very commonly used for real-time data. The author proposed IoT based approach, to cover wide-area and reliability LoRaWAN method is used. For forecasting the occurrence of flood GRU neural network, LASTM and ANN model is used [Mousavi et al. 2021].

For reducing the flood damage, An ANN is used to predict the flood in Sungai, Malaysia. Three different optimization algorithms named LM, back-propagation, GD, and BR with back-propagation for optimizing the ANN result. The result of the prediction of BR is satisfactory [Keong et al. 2016]. Eight different ML models are implemented and their results were compared. 201 flood events are included in this dataset and 10,000 were randomly selected in the Haraz watershed (Iran) Non-event number. Among all models, The proposed EMmedian provides a higher accuracy [Shafizadeh-Moghadam et al. 2018].

Neural network training is a very tedious task as well as computationally expensive. To reduce the complexity of training time normalizes the activation of neurons. In the latest research Batch normalization technique was used when training our network. The efficiency of batch normalization is depending on the mini-batch size. In this paper, the author uses the layer normalization concept to speed up NN training [Al Nuaimi et al. 2015]. In a neural network, the input of each layer depends on the output of the preceding layer, a small change can lead to a big difference. The batch normalization technique is used to solve this by normalizing for mini-batch. This paper is focused on BN for physical buckling. The author design the CNN model with or without batch

normalization and tested that modal on the MNIST dataset [Chen et al. 2017]. In this paper author efficiently us the concert of batch normalization for improving the speed of DNN. To calculate the effectiveness of the method results are validate on CPU, GPU, and Raspberry Pi. An experiment was done by using the Caffe framework by combining the batch normalization and previous liner layer [Duan et al. 2018].

Recurrent Neural Networks (RNNs) is a very dynamic and robust sequential model and can learn long-term dependencies. According to recent research, it shows that by using normalization of the intermediate result of the neural network convergence rate of the network is improved. In this paper, the author shows how batch normalization is used in RNN. This model is applied to both speech reorganization and language modeling tasks [Laurent et al. 2016]. During the training of neural network parameters of each layer are change as the parameter of the previous layer is changed. In this paper, the author uses the concept of internal covariant shift and solves the problem by normalizing the layer input. Made use of an ensemble Network, the author tries to improve on the best-published Results on ImageNet classification [Kang et al. 2017]. A DNN based spectral-spatial method is used to develop a hyperspectral Image classification. Every layer of the deep network uses spectral reduction as pre-processing and BN [Abbasi et al. 2019].

The idea of doing normalization before fusing classification scores of different levels that do not contain vectors. Experiment results are it has been shown that applied technique advances the learning process Better convergence and training for better neural networks the indoor RGBD provides a better understanding of visuals [Hayat et al. 2017]. Some standard technologies, as well as a model, are used for flood prediction are compared in this paper. The author provides a brief overview that the accuracy of every model depends on the different parameters [Dubey and Katarya 2020]. XGBoost and KNN are also providing better results for the data of flood classification. In this paper for flood, prediction the author uses various parameters such as elevation, the slope angle of that geographical area, and distance from the stream network. Among both of the techniques, XGBoost provides more accuracy [El-Magd et al. 2021].

PSO and Firefly have better results for optimizing weights for flash flood prediction [Khan et al. 2020]. Nowadays it is possible to forecast the occurrence of flash floods with the help of ML algorithms at the initial stage and is very helpful in minimizing the casualties. In various researches approaches for flood, detection the author uses a very small size dataset. So the accuracy of the model is very less. In many papers, authors were used ANN, but the layers are not so much dense. So the resultant accuracy is less.

187 **Table 1. Comparison table various for flood prediction approaches**

Paper	Dataset	Algorithm	Description	Limitations
Pyayt et al., 2011	Hyperspectral Images	Convolutional Neural Network and Principal Components Analysis(PCA)	For HSI classification this paper focuses on the PBCNN framework. PCA is used to reduce spectral dimensionality.	An increase in the number of images in the dataset can provide better accuracy
Kim et al. 2016	The dataset includes the presence of 201 floods from Iran's Mazandaran province.	Various Ensemble models were used with eight different ML models	Various Ensemble models (EM) were used with eight different ML models EMmedian resulted in the highest accuracy	Work performed on a limited data set, which leads to low model accuracy.
Saeed et al. 2018	In this paper, multiple sensor values are named as Humidity, Water Flow, Water Level.	CNN and Big data	This study IoT data IS manage using HDFS map-reduce is used. For classification Deep neural network is developed.	In the future we get more enhanced results with more sensors and less cost, it is possible to detect floods using the IoT.
Alquisola et al. 2018	Hydrodynamic parameters collected from local stations and 16 different datasets were collected from local meteorology	ANN-based prediction model	An ANN-based after-runner forecast model is developed to predict future volatility. That leads to Time (eg, 24 hours)	An increase in the number of neurons in the hidden layers and more thunderstorm events to be considered in the forecast model
Anbarasan et al. 2020	Input data meteorological and Gloves. Hydrological data obtained	Integrated Large and crowdsource data using ML techniques	The author uses various ML techniques, among them ANN, SVM and RF provides better classification accuracy	Improve the limitation is include Incorporating the Internet of Things (IoT) into the real measure Meteorological data
	A solar panel with	A nonlinear	The approach of this	The resulting

Puttinaovarat et al. 2020.	a microcontroller, GSM module, and ultrasonic sensor are used.	autoregressive network with feed-forward ANN with the Backpropagation technique is used.	paper is categorized into three parts, In this main data collection is done. then describe the hardware architecture and last is the architecture of the ANN model and predictions.	estimated flood water level matches well with the observed water levels for both the NARX and NAR models. The NARX model provides better prediction performance than the NAR model.
Mousavi et al. 2021	NASA Earth Data	Convolutional Neural Network	The experiment was done by using the Caffe framework and for CNN ResNet20, ResNet50, ResNet152, and Inception models were used	Efficient algorithms are needed to improve the training time of CNN's.
Khan et al. 2020	Multiple sensor data and Satellite images	AI and ML-based algorithms	All major AI-based algorithms, sensor-based and modeling based models for flash flood detection are discussed	The forecast is for better performance to optimize flood firefly and PSO is used in the future.
Liu et al. 2021	11 flood condition factors from the various standard dataset	SVM	The author introduced the FSCI to measure the vulnerability of floods. this is achieved by using SVM	Future studies should consider more comprehensive indicator systems, climate change factors, and higher quality flood points.
Charbuty and Abdulazeez et al. 2021.	Environmental sensor data	Decision tree classifiers	In this paper, the author provides a detailed description of decision trees for classification.	The algorithm provides better accuracy if more data instances are used.

188

189

The outcome of the Related Work

In the above survey, we provide an overview of the various AI, ML, and DL-based approaches for flood prediction as well as detection. The experiments are performed on various datasets that contain environmental parameters and satellite images for prediction.

- There is limited review is on the forecasts in flesh flood. Further, novel framework development is recommended.
- All research is done on the basic ML and AI-based algorithm which provides less accuracy. In the case of Deep neural networks, there is a need to add a more dense hidden layer for providing better accuracy.
- Dataset considered in various research is very small that causes the lower performance of the algorithm.

3. Preliminaries for Flood disaster detection system

Worldwide all the countries are concern about natural disasters. There are two types of disasters, natural disasters like tsunamis, forest-fire, lightning, flood, earthquake, and also landslide, and man-made disasters like leakage in gas production, leakage in an oil pipeline, industrial explosion, and man-made disaster can be controlled or saved by a man.

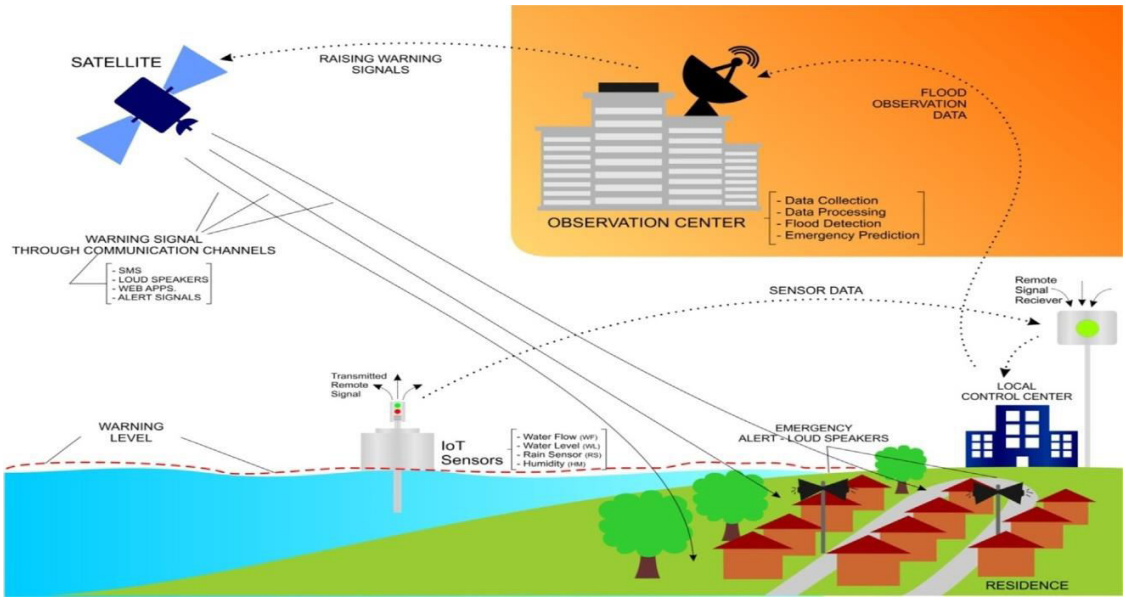


Fig 1: Working of Flood Disaster Early Warning System [Saeed et al. 2018]

In the above figure 1 give an overview that how with help of current technologies an effective Flood prediction system is designed. The proposed approach extracts various features that are used to predict flooding from the climate using different types of sensors and predict the probability of flooding with certain prediction algorithms. But in the case of natural calamities, it is possible we only create an early prediction system to predict the occurrence of the disaster so that we can take some precautions. Among all-natural disasters, flood very commonly occurs in natural disasters. In this paper, we have a focus on flood disaster detection systems. People will be prevented to face such a disaster from early detection of disaster. Here an overvirw of verious ML algorithms is provides.

3.1 Support Vector Machine

SVM is a supervised learning algorithm that may be used to tackle regression and classification tasks like SVC and SVR. Some hyperplanes may be utilized to separate different data instances in SVM. Here we focus on identifying the hyperplane with the largest margin, or distance between data points from both classes. The sets to differentiate are usually not linearly separable in that space, even though the original problem is described in a finite-dimensional space. Maximizing distance helps classify data points. Also, it allows data points to be assigned to different classes. [Yan et al. 2017].

In addition, the number of features determines the hyper plane's dimension. Because supervised learning is impossible when data is unlabeled, an unsupervised learning method is required, in which data is spontaneously grouped into groups and new data is mapped to these classes. In this approach, for generating a flood sensitivity map SVM model is used based on 11 factors of flood conditions [Liu et al. 2021]. Even though the initial problem is expressed in a finite-dimensional space, the sets to differentiate are often not linearly separable in that space.

3.2 Decision Tree

The DT is one of the regularly used machine learning algorithms. It is the most popular and powerful tool for data prediction and classification. DT is a kind of supervised learning algorithm. Where each data instance belongs to a specific class. This structure is a tree-like representation of the features of a dataset, which are represented by the branches and the leaf nodes. It shows the various decisions and solutions that can be derived from the dataset. This algorithm is used to build a model to identify the value of the target variable. For which a tree-like structure is used where the leaf node represents the class label for the corresponding problem. The internal node attributes are represented by that tree. [Somvanshi et al. 2016].

When we talk about flood prediction systems then DT is very useful and commonly used The classification algorithm is used for predicting the likelihood of floods based on the collected data. It also classifies the data according to its classification. [Charbuty and Abdulazeez et al. 2021].

3.3 Artificial Neural Network

An ANN is an ML model that draws on the principles of biological neurons. It learns and uses information flow to make predictions. ANN can be classified into a feed-forward (static) network and feed backward (Recurring) network. A feed-forward neural network implementation performs non-linear functions of provided input. It is a nonlinear function of neurons that is determined by nonlinear, bias functions and weight [Dutt et al. 2018].

ANN is the most generally used data processing technique. ANN has been applied in most studies with significant emphasis on prediction. The author was used ANN are used as a base model. Feedforward fully connected neural network is the very first neural network. This consists of several layers of neurons (nodes) connected. These layers are named the Output layer, Hidden layer, and Input layer. direct edges connect neurons in different layers, with each edge assigned some weight that is updated during the training of our network [Abiodun et al. 2019]. A fully connected feed-forward ANN is used here to give early warning for prediction and to communicate to the target users. In the neural network architecture, various activation functions are used in neural network architecture to activate neurons ReLU is used [Subeesh et al. 2018].

3.4 Deep neural network

Since the proposal of fast learning algorithms for deep belief networks, deep learning techniques have attracted ever-increasing research interest due to their inherent ability to overcome the shortcomings of traditional algorithms relying on hand-designed features. The basic unit of the neural network is called neurons/ perceptrons. So how neural network does is work? Multiple inputs are provided to the neurons named x1, x2, etc. it depends on the number of input data instances and produces output depends upon the output classes [Kwasigroch et al. 2019].

Not only can DNNs solve a task according to an algorithm, but they can also use their experience to predict its solution. [Liu et al. 2017]. DNN is a distributed system that works seamlessly in multiple layers. This is beneficial when you need to use it to replace human labor without compromising on its efficiency.

4. Proposed Model

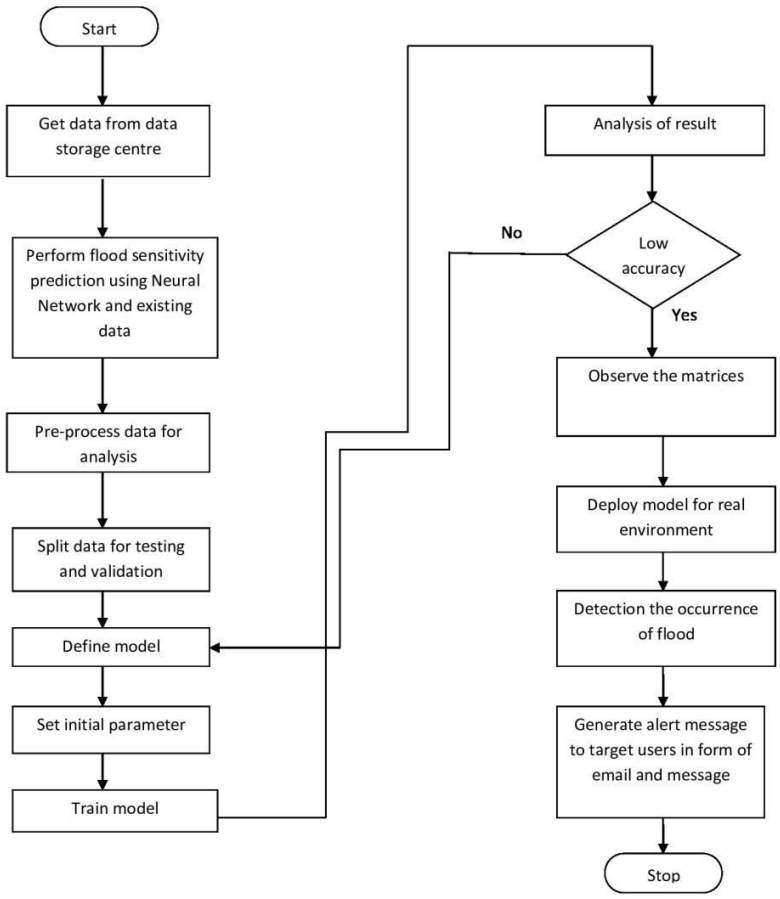


Fig 2: Represent Flow Chart of the proposed model

In the above figure 2, we present a brief overview of the working of the flood detection system. We demonstrate the working of our proposed methodology. The process is started from collect data from the data storage center then we pre-process the initial data. Some techniques like ANN, DNN, and Batch Normalization are used for the classification of data. Our data is classified into three classes named 'LOW',' MODERATE' and 'HIGH'. When real-time data is feed to the model then it will respond. Initially, the threshold value of each attribute is set then when the flood is cross its severity level then it sends the alert message regarding the flood to the high authorities

to take appropriate action. In this model, we also send the alert message to the public community live near the water bodies.

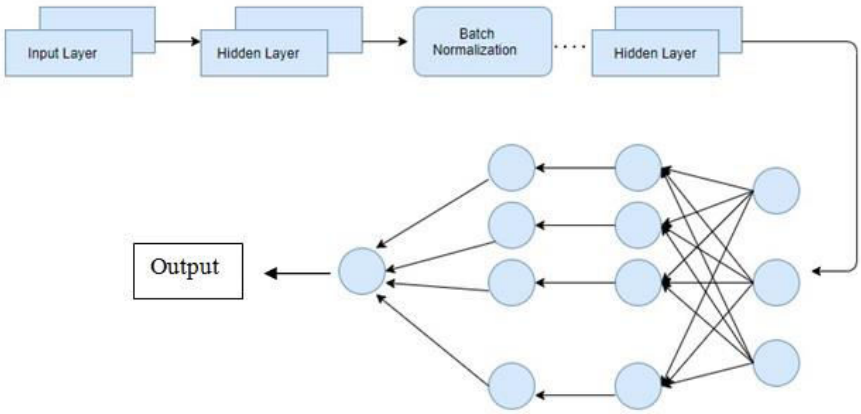


Fig. 3 fully connected Neural Network with Batch Normalization

In above figure 3, we explain the basic structure of our proposed BNDNN model for flood prediction based on the severity of the rain. The first layer of our model is the input layer. The number of neurons in the input layer is dependent on the data instances. The next layer of our model is the Hidden layer. The normalization technique is applied between the output of the first hidden layer and the input to the next hidden layer. Batch Normalization normalizes the input data to the activation function of each layer by using the small mini-batches. In the output layer, we use a single neuron and predict the possibility of flooding outside based on rainfall.

In our model, to introduce non-linearity into the model ReLU activation function is used. For classification, the softmax activation function is used to obtain the probability of being between 0 and 1. The equation of ReLU and softmax function is given below.

In neural networks, the Cross-entropy loss function is used while training our network. For prediction and classification, problem cross-entropy error is provide better results. It predicts a class highest likelihood evaluation based on input. It is an alternative to the cross-entropy loss function and the mean square error. The equation of the cross-entropy loss function is given below.

$$H_{Y'}(Y) = -\sum_i y_i' \log(y_i) \qquad (1)$$

In the above equation, 1 y_i is the predicted probability value for class i, and y_i' is the true probability for that class. $H_{Y'}(Y)$is the loss function. The error is calculated after the first iteration is complete and it spreads back to the input layer in the previous layers and computes the gradient. For the learning of neural networks, back-propagation techniques are used. The approach here for optimizing the neural network is called gradient descent for adjusts the weight. Weight is updated according to the formula below.

$$\omega = \omega - \mu \nabla Q(\omega) \quad (2)$$

In equation 2, Q is the phase of the shape or sometimes called the learning rate. ω is the weight assign to neurons. This reduces further errors. In this method stochastic gradient descent method is adopted, the method adopted is based on the calculation of cost gradient. For all the training data during the iteration, this process is repeated. Then our network is well learned to predict the outputs for the test inputs.

Batch Normalization normalizes the input data to the activation function of each layer by using the small mini-batches. The batch has the standard normal distribution (with standard deviation is 1 and mean is 0).

Let us consider mini-batch is B, size of mini-batch is M (consider as entire data-set) i.e

$$BN_{\gamma,\beta} : \quad x_{1.....m} \rightarrow y_{1.....m}$$

$$B = \{X_1, ..., X_M\}$$

The Batch Normalization is

$$BN_{\gamma,\beta} : \quad x_{1.....m} \rightarrow y_{1.....m} \quad (3)$$

After calculating variance and mean of small-batch data we normalize it

$$\widehat{x_i} \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 - \varepsilon}} \quad (4)$$

μ_B is sample mean

σ_B^2 is sample variance

Scale and shift is represented as follows

$$y_i \leftarrow \gamma \hat{x_i} + \beta \equiv BN_{\gamma,\beta}(x_i) \quad (5)$$

Where γ and β need to be calculated

When the mean and variance are calculated over the entire training dataset that normalization increases the speed of the network. While normalizing each input of layer will lose the information transferred from the previous layer.

In Batch Normalization, if we use to normalize the entire dataset in each training set, it is not useful when we use a stochastic optimizer. So another method of normalization is used. While using stochastic gradient training, in this approach each mini-batch is calculated mean and variance of each activation. In this, the data used for normalization can be fully used in gradient backpropagation. Consider the mini-batch B with size m, in this, we applied normalization to each activation function.

$$\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^{m} x_i \quad (6)$$

$$\sigma_B^2 \leftarrow \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2 \quad (7)$$

This approach is used to improve the performance, and speed of ANN. It normalizes the output of the input layer by re-scaling and re-centering. This allows us to be less careful about higher learning rates and initialization. In BN each layer has some impact according to random distribution. The input is changed during the training of the network due to the randomness of data. The effect of this randomness on the input to the internal layer is defined as the internal covariant shift. To resolve the issue of the internal covariant shift we have to standardize every input layer. It is very time-consuming and less effective. So to solve this problem batch normalization is introduced to reduce the internal covariant shift.

5. Result and Discussion

5.1 Data collection

The main thing to design a flood natural disaster prediction system is to need a huge dataset to process and generate the result; data is a very important part of any prediction model. Data can be used in any format such as images, videos, etc. For forecasting, data collected from any source sensor, satellite, the video camera is used. A large area of India is a flood-prone region as most of the human communities are located in nearby rivers. Therefore, they face floods for the maximum time during the rainy season. India is very suitable for flood detection. Following this, many of the states and union territories of India were facing major floods.

By doing this, it was found that Bihar has the highest risk of floods in the country. According to International Water Management Institute (IWMI), every year 73 percent of the total surface gets submerged. In Orissa to the frequency of flood events was very high. So we include two main states of India, Bihar, and Orissa for our system. Since Bihar and Orissa are both very large states, districts in these two states is 38 and 30 respectively, we select the ten most flood-prone cities from each state. For our prediction system, we use a standard dataset that includes parameters like minimum temperature, maximum temperature, precipitation; vapor pressure along with this parameter year, the month is also included in the dataset. We have considered over fifteen years between January to December.

5.2 Confusion matrix

The confusion matrix is a type of binary classifier that can be any size depends upon the number of different classes/ labels. Here the size of the matrix is (3X3) as we divide our data into three classes Low, Moderate and High. This is a table with four different combinations of estimated and actual values. These combinations are TP, FP, TN, and FN.

TP = true positive
FP = false positive
TN = true negative
FN = false negative
Where TP and TN denote the number of samples correctly classified the severity of the flood respectively. FP and FN denote the number of samples that have been incorrectly classified based on the severity of flood events.
In our paper, all four values of TP, FP, TN, and FN are extremely useful for calculating the various parameters named Precision, Recall, F1-score, and accuracy.

5.2.1 Precision

Precision is calculated as the fraction of correctly identifies instances among all instances which are identified as positive, which means how many instances are positive. For better results precision should be high. Precision is calculated using the following formula

$$\text{Precision} = \frac{TP}{TP+FP} \tag{10}$$

High precision requires a low FN rate and a low FP rate.

5.2.2 Recall

The Recall is a metric that measures the number of true positive predictions made from all positive predictions that could have been made. The percentage of the total positive is estimated to be positive. It is similar to TPR (true positive rate). For unbalanced learning, recall is typically used to measure minority class coverage. In our model or algorithm, we focus on the high recall value. The recall is calculated using the following formula

$$\text{Recall} = \frac{TP}{TP+FN} \tag{11}$$

High precision requires a low FN rate and a low FP rate.

5.2.3 F1-score

F1-score is a matrix; the accuracy and recall of both the values are needed to calculate the F1-score. It is a harmonic average of recall and precision wherein the best-case F1 value is 1 and in the worst case, the F1 value is 0. To calculate F1-score following formula is used

$$\text{F1-score} = 2 * \frac{Precision * Recall}{Precision + Reall} \tag{12}$$

5.2.4 Accuracy

Based on the confusion, matrix, for calculating the accuracy of the ratio of the correctly predicting instances to the total instances in the dataset. The formula to calculate accuracy is given below

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{13}$$

5.3 Analysis of the result

In this section, we have shone the accuracy, precision, F-score, and Recall of our proposed model and some other machine learning algorithms and compare their results. For this, we collect data from data storage then pre-processed the data to feed to our network. Here we have train five models and machine learning algorithms named SVM, DT, ANN, DNN, and our proposed model DNN with Batch Normalization (BNDNN). We divide the dataset into training, validation, and testing categories. From the dataset 70% of data is used for training the model 15% is used for validation and 15% is used for testing. According to the severity of the flood, data is divided into three classes named as Low, Moderate, and High chance of flood.

Table 1 Cooperative tabulation of Proposed Model (BNDNN), SVM, DT, ANN and DNN

Model/ Algorithm		**Precision**	**Recall**	**F1-score**
SVM	Low	90	70	85
	Moderate	76	87	83
	High	77	90	87
	Accuracy	**83**		
DT	Low	93	76	82
	Moderate	75	92	80
	High	73	87	83
	Accuracy	**80**		
ANN	Low	92	98	79
	Moderate	70	94	82
	High	75	89	85
	Accuracy	**82**		
DNN	Low	95	72	83
	Moderate	94	97	87
	High	78	82	91
	Accuracy	**84**		
Proposed Model (BNDNN)	Low	98	80	89
	Moderate	81	99	92
	High	84	95	94
	Accuracy	**89**		

In the above table 1, we have compared the four existing models and ML algorithms such as ANN, DNN, SVM, and DT with our proposed model. The comparison is done based on the parameter like F1-score, recall, precision, and accuracy of the models. Here we calculate the F1-score, recall, and precision for all the three classes of every technique and at the end, we calculate the accuracy of the model and algorithm. Based on the above

comparison, we say that the proposed model in this paper Batch Normalization with Neural Network (BNDNN) is performing better. We get higher accuracy with our model.

5.5 Performance Analysis

Here the BNDNN is performed better in comparison to SVM, DT, ANN, and DNN on the dataset collected from the data storage center. Based on the above comparison, we say that the proposed model in this paper Batch Normalization with Neural Network (BNDNN) is performing better. We get higher accuracy with our model. In this section, we discuss the graphical representation and identify that which machine learning algorithms such as ANN, SVN, DT, DNN, and our proposed model BNDNN are predicting the severity of the flood. Before the result, we provide a brief overview of performance metrics and parameters such as precision, recall, F1-score, and accuracy used for the analysis of the result.

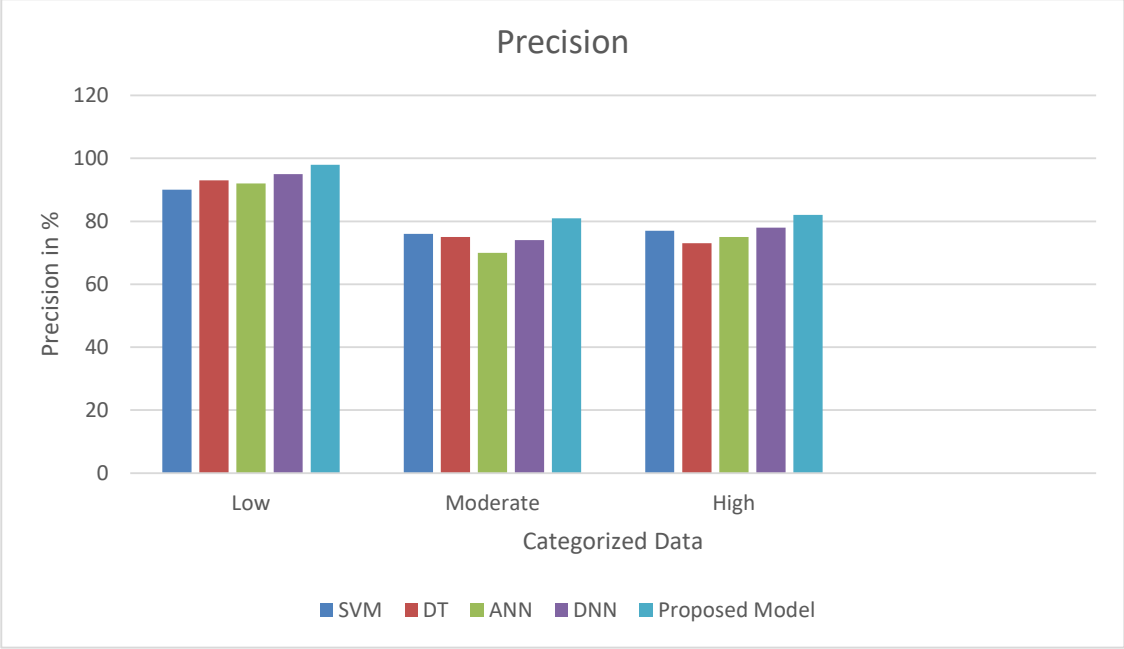


Fig. 4 Graph between Categorized data and Precision of SVM, DT, ANN, DNN, and Proposed Model

In the above figure 4, a graphical representation for the precision of various machine learning algorithms and our proposed model. The performance of each model varies in each data class but our proposed model BNDNN performs better in each case.

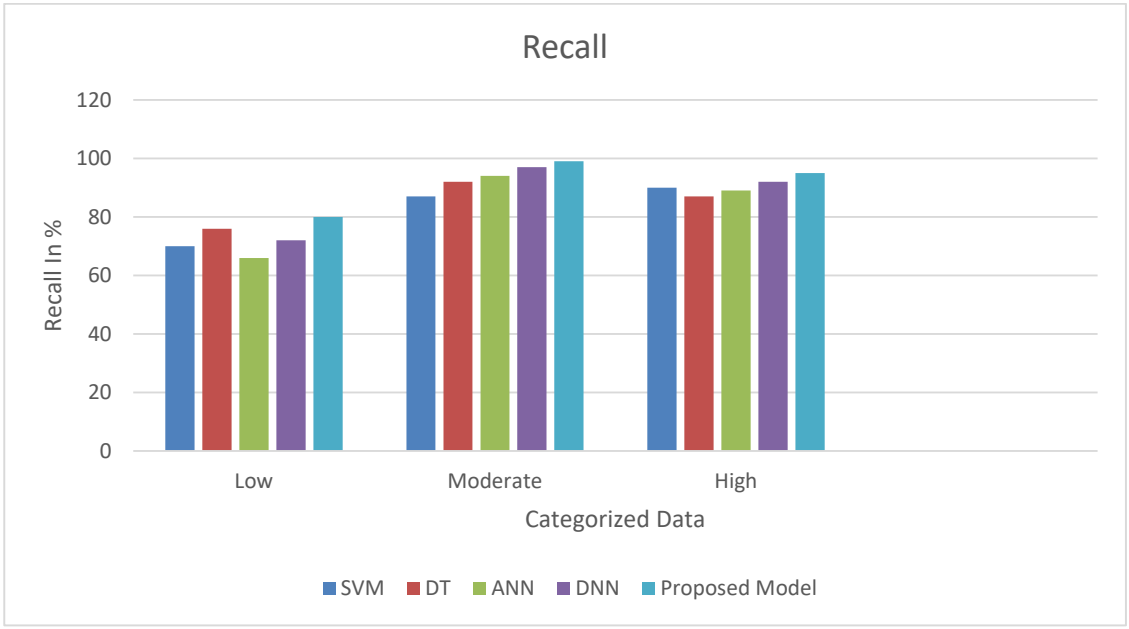


Fig. 5 Graph between Categorized data and Recall of SVM, DT, ANN, DNN, and Proposed Model

In the above figure 5, we represent the graphical comparison for the precision of various machine learning algorithms and our proposed model. Our proposed model BNDNN provides better prediction for all data classes.

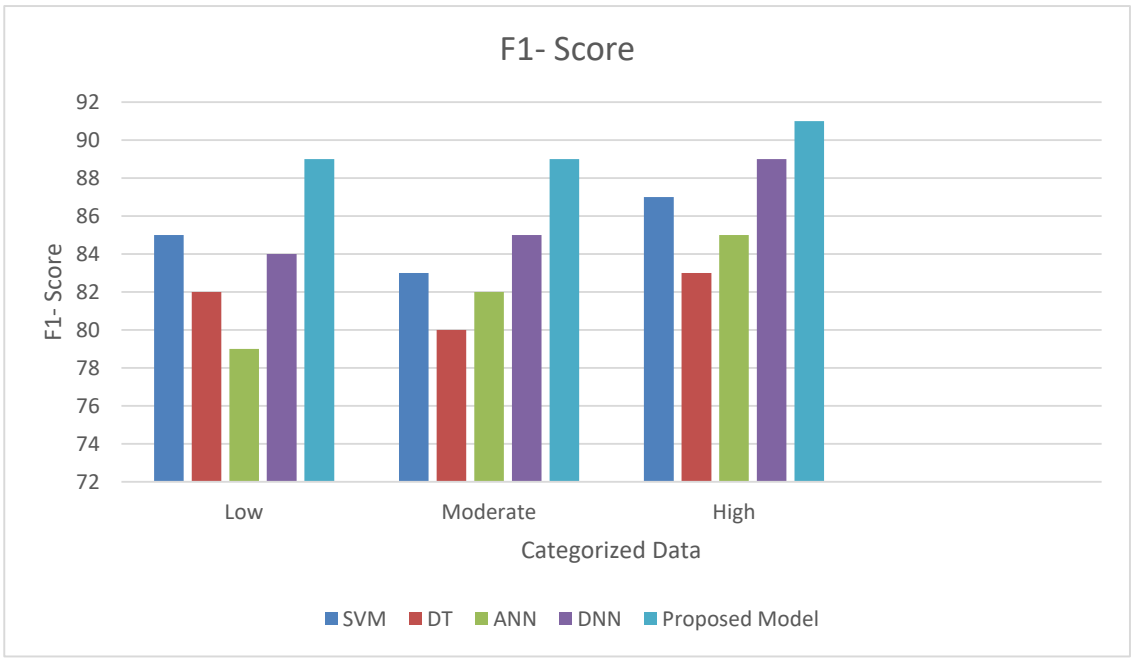


Fig. 6 Graph between Categorized data and F1-score of SVM, DT, ANN, DNN, and Proposed Model

In the above figure 6, the graphical representation for the F1-score of various machine learning algorithms and our proposed model is provided. Other machine learning models are performing poorly in each class. Our proposed model BNDNN provides far better results than the rest of the model.

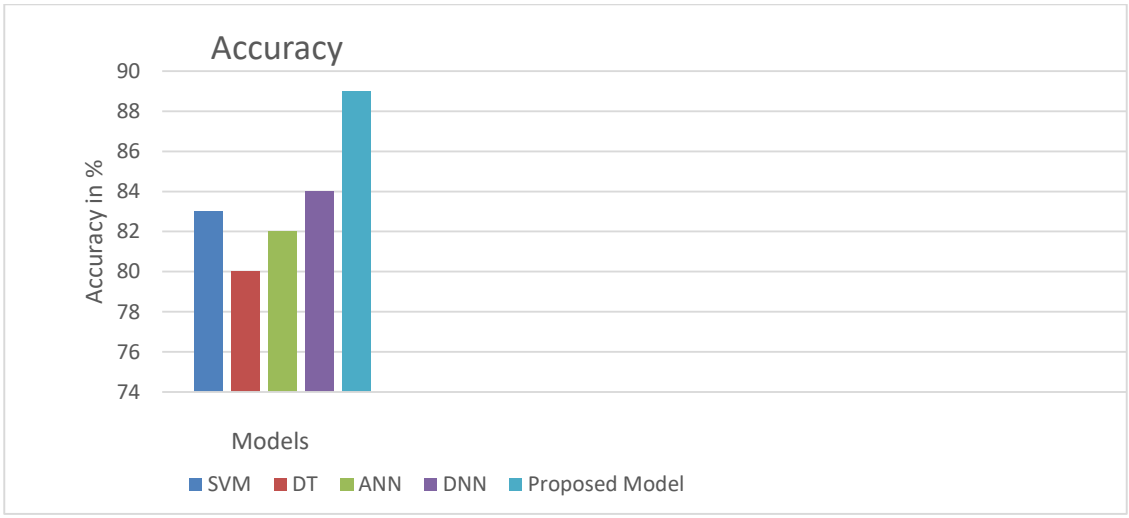


Fig. 7 Graph between Model Accuracy score of SVM, DT, ANN, DNN, and Proposed Model

In the above figure 7, The graphical representation of the accuracy of various machine learning models and our BNDNN proposed model is provided. As a result, our proposed model provides 89% accuracy which is the highest of all accuracy. DNN with 84% accuracy outperforms SVM with 83% accuracy. ANN provides 82% accuracy while DT provides very low 80% accuracy.

It is visible after analyzing the result based on various parameters that our proposed model (BNDNN) provides better accuracy than other algorithms. The rest of the parameters such as precision, recall, and F1-score are also providing better resultant value as compared to other algorithms. For an efficient model or algorithm, the value of precision and recall should be high. The value of the F1-score depends on the precision and recall value, so higher precision and recall give better results.

6. Conclusion and Future Scope

We cannot control human-made disasters because they are unpredictable. To control, we need human awareness and seriousness in human labor disasters. But in case of a natural disaster like flood, fire, cyclone, etc., we can warn people about their occurrence with previous experience and generate a vigilance system with the help of currently available technologies, so that we save diversity can. In this paper, we used a neural network for the prediction of the flood. In this paper, we present a new approach with the vast

capability of recent technologies like DNN combine with batch normalization for the implementation of an effective flood detection system that can combat flood natural disasters.

Our proposed model works on the concept of batch normalization technique. We compare our proposed model with SVM, DT, ANN, and DNN models. We train our model on the standard dataset. We compare our model based on the parameter like accuracy, precision, recall, F1-score. The accuracy of SVM is 83%, DT is 80%, ANN is 82% and DNN is 84% respectively. Whereas our proposed model (BNDNN) stores 89% accuracy higher than all other approaches. After comparison on the above parameter, we can conclude that our proposed model gives more accuracy than other models used for prediction. In the future, if we use more environmental parameters the there is a chance to achieve higher accuracy. Also if we combine other models and develop a hybrid model for flood prediction then also we get better accuracy.

Compliance with ethical standards

Conflict of interest There is "NO any" direct or indirectly related conflict of interest for this manuscript and with authors.

Reference

Hirabayashi Y, Mahendran R, Koirala S, Konoshima L, Yamazaki D, Watanabe S, Kim H, Kanae S (2013) Global flood risk under climate change. Nature Climate Change 3:816-821. doi: 10.1038/nclimate1911

Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F. A., & Feyen, L. (2016). Development and evaluation of a framework for global flood hazard mapping. *Advances in Water Resources*, *94*, 87–102. https://doi.org/10.1016/j.advwatres.2016.05.002

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. Water (Switzerland), 10(11), 1–40. https://doi.org/10.3390/w10111536

Li, X., Yan, D., Wang, K., Weng, B., Qin, T., & Liu, S. (2019). Flood risk assessment of global watersheds based on multiple machine learning models. Water (Switzerland), 11(8), 1–18. https://doi.org/10.3390/w11081654

Liu Q, Li Y, Yu M, Chiu L, Hao X, Duffy D, Yang C (2019) Daytime Rainy Cloud Detection and Convective Precipitation Delineation Based on a Deep Neural Network Method Using GOES-16 ABI Images. Remote Sensing 11:2555. doi: 10.3390/rs11212555

Ahmadisharaf E, Tajrishy M, Alamdari N (2015) Integrating flood hazard into site selection of detention basins using spatial multi-criteria decision-making. Journal of Environmental Planning and Management 59:1397-1417. doi: 10.1080/09640568.2015.1077104

Aazam M, Khan I, Alsaffar A, Huh E (2014) Cloud of Things: Integrating Internet of Things and cloud computing and the issues involved. Proceedings of 2014 11th International Bhurban Conference on Applied Sciences & Technology (IBCAST) Islamabad, Pakistan, 14th - 18th January, 2014. doi: 10.1109/ibcast.2014.6778179

Pyayt, A. L., Mokhov, I. I., Lang, B., Krzhizhanovskaya, V. V., & Meijer, R. J. (2011). Machine learning methods for environmental monitoring and flood protection. World Academy of Science, Engineering and Technology, 78(6), 118–123. https://doi.org/10.5281/zenodo.1075060

Saeed F, Paul A, Rehman A, Hong W, Seo H (2018) IoT-Based Intelligent Modeling of Smart Home Environment for Fire Prevention and Safety. Journal of Sensor and Actuator Networks 7:11. doi: 10.3390/jsan7010011

Ranit A, Durge P (2019) Flood Forecasting by Using Machine Learning. 2019 International Conference on Communication and Electronics Systems (ICCES). doi: 10.1109/icces45898.2019.9002579

Ghose D (2018) Measuring Discharge Using Back-Propagation Neural Network: A Case Study on Brahmani River Basin. Advances in Intelligent Systems and Computing 591-598. doi: 10.1007/978-981-10-7566-7_59

Anbarasan M, Muthu B, Sivaparthipan C, Sundarasekar R, Kadry S, Krishnamoorthy S, Samuel R. D, Dasel A (2020) Detection of flood disaster system based on IoT, big data and convolutional deep neural network. Computer Communications 150:150-157. doi: 10.1016/j.comcom.2019.11.022

Mitra P, Ray R, Chatterjee R, Basu R, Saha P, Raha S, Barman R, Patra S, Biswas S, Saha S (2016) Flood forecasting using Internet of things and artificial neural networks. 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). doi: 10.1109/iemcon.2016.7746363

Alquisola G, Coronel D, Reolope B, Roque J, Acula D (2018) Prediction and Visualization of the Disaster Risks in the Philippines Using Discrete Wavelet Transform (DWT), Autoregressive Integrated Moving Average (ARIMA), and Artificial Neural Network (ANN). 2018 3rd International Conference on Computer and Communication Systems (ICCCS). doi: 10.1109/ccoms.2018.8463238

Puttinaovarat S, Horkaew P (2020) Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques. IEEE Access 8:5885-5905. doi: 10.1109/access.2019.2963819

576 Bande S, Shete V (2017) Smart flood disaster prediction system using IoT & neural networks.
577 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon). doi:
578 10.1109/smarttechcon.2017.8358367

579 Kim S, Matsumi Y, Pan S, Mase H (2016) A real-time forecast model using artificial neural
580 network for after-runner storm surges on the Tottori coast, Japan. Ocean Engineering 122:44-53.
581 doi: 10.1016/j.oceaneng.2016.06.017

582 Sahagun M, Dela Cruz J, Garcia R (2017) Wireless sensor nodes for flood forecasting using
583 artificial neural network. 2017IEEE 9th International Conference on Humanoid, Nanotechnology,
584 Information Technology, Communication and Control, Environment and Management
585 (HNICEM). doi: 10.1109/hnicem.2017.8269462

586 Lee S, Kim J, Jung H, Lee M, Lee S (2017) Spatial prediction of flood susceptibility using
587 random-forest and boosted-tree models in Seoul metropolitan city, Korea. Geomatics, Natural
588 Hazards and Risk 8:1185-1203. doi: 10.1080/19475705.2017.1308971

589 Hosseiny H, Nazari F, Smith V, Nataraj C (2020) A Framework for Modeling Flood Depth Using
590 a Hybrid of Hydraulics and Machine Learning. Scientific Reports. doi: 10.1038/s41598-020-
591 65232-5

592 Mousavi F, Yousefi S, Abghari H, Ghasemzadeh A (2021) Design of an IoT-based Flood Early
593 Detection System using Machine Learning. 2021 26th International Computer Conference,
594 Computer Society of Iran (CSICC). doi: 10.1109/csicc52343.2021.9420594

595 Keong K, Mustafa M, Mohammad A, Sulaiman M, Abdullah N (2016) Artificial neural network
596 flood prediction for sungai isap residence. 2016 IEEE International Conference on Automatic
597 Control and Intelligent Systems (I2CACIS). doi: 10.1109/i2cacis.2016.7885321

598 Shafizadeh-Moghadam H, Valavi R, Shahabi H, Chapi K, Shirzadi A (2018) Novel forecasting
599 approaches using combination of machine learning and statistical models for flood susceptibility
600 mapping. Journal of Environmental Management 217:1-11. doi: 10.1016/j.jenvman.2018.03.089

601 Al Nuaimi E, Al Neyadi H, Mohamed N, Al-Jaroodi J (2015) Applications of big data to smart
602 cities. Journal of Internet Services and Applications. doi: 10.1186/s13174-015-0041-5

603 Chen L, Fei H, Xiao Y, He J, Li H (2017) Why batch normalization works? a buckling
604 perspective. 2017 IEEE International Conference on Information and Automation (ICIA). doi:
605 10.1109/icinfa.2017.8079081

606 Duan J, Zhang R, Huang J, Zhu Q (2018) The Speed Improvement by Merging Batch
607 Normalization into Previously Linear Layer in CNN. 2018 International Conference on Audio,
608 Language and Image Processing (ICALIP). doi: 10.1109/icalip.2018.8455587

609 Laurent C, Pereyra G, Brakel P, Zhang Y, Bengio Y (2016) Batch normalized recurrent neural
610 networks. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing
611 (ICASSP). doi: 10.1109/icassp.2016.7472159

Kang D, Park M, Kim H, Kim D, Kim S, Son H, Lee S (2017) Room Temperature Control and Fire Alarm/Suppression IoT Service Using MQTT on AWS. 2017 International Conference on Platform Technology and Service (PlatCon). doi: 10.1109/platcon.2017.7883724

Abbasi A, He M (2019) Convolutional Neural Network with PCA and Batch Normalization for Hyperspectral Image Classification. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. doi: 10.1109/igarss.2019.8899329

Hayat H, Yazhou Liu, Shah M, Ahmad A (2017) Batch regularization to converge the deep neural network for indoor RGBD scene understanding. 2017 2nd International Conference on Image, Vision and Computing (ICIVC). doi: 10.1109/icivc.2017.7984665

Dubey V, Katarya R (2020) An Analysis of Machine Learning Techniques for Flood Mitigation. Advances in Intelligent Systems and Computing 299-307. doi: 10.1007/978-981-15-5148-2_27

El-Magd S, Pradhan B, Alamri A (2021) Machine learning algorithm for flash flood prediction mapping in Wadi El-Laqeita and surroundings, Central Eastern Desert, Egypt. Arabian Journal of Geosciences. doi: 10.1007/s12517-021-06466-z

Khan T, Alam M, Shahid Z, Su'Ud M (2020) Investigation of Flash Floods on Early Basis: A Factual Comprehensive Review. IEEE Access 8:19364-19380. doi: 10.1109/access.2020.2967496

Yan J, Jin J, Chen F, Yu G, Yin H, Wang W (2017) Urban flash flood forecast using support vector machine and numerical simulation. Journal of Hydroinformatics 20:221-231. doi: 10.2166/hydro.2017.175

Liu J, Xiong J, Cheng W, Li Y, Cao Y, He Y, Duan Y, He W, Yang G (2021) Assessment of Flood Susceptibility Using Support Vector Machine in the Belt and Road Region. doi: 10.5194/nhess-2021-80

Somvanshi M, Chavan P, Tambade S, Shinde S (2016) A review of machine learning techniques using decision tree and support vector machine. 2016 International Conference on Computing Communication Control and automation (ICCUBEA). doi: 10.1109/iccubea.2016.7860040

Charbuty B, Abdulazeez A (2021) Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technology Trends 2:20-28. doi: 10.38094/jastt20165

Dutt M, Nunavath V, Goodwin M (2018) A Multi-layer Feed Forward Neural Network Approach for Diagnosing Diabetes. 2018 11th International Conference on Developments in eSystems Engineering (DeSE). doi: 10.1109/dese.2018.00060

Abiodun O, Kiru M, Jantan A, Omolara A, Dada K, Umar A, Linus O, Arshad H, Kazaure A, Gana U (2019) Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition. IEEE Access 7:158820-158846. doi: 10.1109/access.2019.2945545

647 Subeesh A, Kumar P, Chauhan N (2018) Flood Early Detection System Using Internet of
648 Things and Artificial Neural Networks. International Conference on Innovative
649 Computing and Communications 297-305. doi: 10.1007/978-981-13-2324-9_30

650 Kwasigroch A, Grochowski M, Mikolajczyk M (2019) Deep neural network architecture
651 search using network morphism. 2019 24th International Conference on Methods and
652 Models in Automation and Robotics (MMAR). doi: 10.1109/mmar.2019.8864624

653 Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi F (2017) A survey of deep neural
654 network architectures and their applications. Neurocomputing 234:11-26. doi:
655 10.1016/j.neucom.2016.12.038

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Carbon Nanotubes in Emerging Photovoltaics: Progress and Limitations

Sakshi Sharma and Vinod Singh, *Member, IEEE*

***Abstract*—Amid a wide-ranging search for materials that can aid the optimization of solar photovoltaic performances, propelled by the ever increasing demand for clean and renewable energy in the 21st century society, Carbon nanotubes (CNTs) offer an excellent avenue for progress. While multiple papers have reviewed and reported on their unique properties and individual applications, studies highlighting the role of CNTs in various solar cell architectures in a compiled form are relatively sporadic and harder to find. This article presents overviews of the development surrounding the incorporation of CNTs in different types of photovoltaic devices, namely organic photovoltaics, perovskite solar cells, dye sensitized solar cells, silicon-based photovoltaics, and Thin film solar cells, along with the important drawbacks that need to be addressed in order to ensure high efficiency, stable, and commercially viable photovoltaic cells. Based on the performance-cost tradeoff, future outlook regarding the prominence of CNT in the commercial solar cell market is also discussed.**

***Index Terms*—Carbon nanotubes (CNT), dye sensitized solar cells (DSSC), organic photovoltaics (OPV), silicon based photovoltaics, thin film solar cells.**

I. Introduction

THE discovery of cylindrical molecules formed of rolled up graphene sheets-named carbon nanotubes (CNTs)—popularly attributed in literature to Sumio Iijima in 1991 for multiwalled CNT (MWCNT) and in 1993 for single walled CNT, catapulted the material into the center of global scientific research [1]. Undoubtedly a material of the 21st century, CNTs now enjoy ubiquitous application as composite materials for the textile and automotive industries, catalysts for chemical processes, thin film transistors, sensors, displays, photovoltaics, and nanomedicine, among others. While single layer CNT (SWCNT) consisting of—as the name suggests-one single layer of hexagonally bonded network of carbon atoms can have a diameter between 1–5 nm, MWCNT consisting of concentrically linked walls typically has diameter equal to 100 nm. Their unique applications are a result of their unique properties. For example, their remarkably high mechanical tensile strength (∼400 times that of steel) and low density (∼one-sixth that of steel) makes them the ideal choice for reinforced composite materials used in sports equipment [2]. Their exceptionally large surface area, surface to volume ratio, and chemical stability grants them the ability to be used as direct catalysts or catalytic supports in photo- and electrocatalysis [3].

The current limitations in Si based conventional metal oxide semiconductor field effect transistors have been dealt with by employing device miniaturization. However, the limits of physical scaling and fabrication challenges have made it increasingly harder to achieve low power, high performance electronics. carbon nanotube field effect transistors (CNFET) presents an excellent alternative [4], [5]. Based on semiconducting CNTs as the channel between source and drain electrodes, CNFETs offer improved carrier mobility, absence of short channel effects in the nanoscale regime, room temperature processing and remarkably, the ability to stack circuits on top of each other creating 3-D chips with extremely high computing power-something that is not possible with Si-based architecture [6]. Additionally, hydrogen sensors based on palladium nanoparticle-decorated SWCNT as the sensing element provide enhanced mechanical flexibility and performance [7].

Unsurprisingly, CNTs are an emerging material of choice for upcoming high efficiency photovoltaic cells. Whether it be CNT based transparent conductive films (TCF) as a replacement for conventional indium in oxide (ITO) based TCF in organic photovoltaic (OPV) devices [8], the fiber shaped dye sensitized solar cells (DSSC) utilizing CNT in counter electrodes [9], or heterojunction solar cells combining Si with SWCNT thin films [10], CNTs in various forms are turning out to be tailorable, multifunctional materials. CNT provides the benefit of easy fabrication, stretchability combined with good mechanical strength, high optical transparency in the visible and infrared region, hole selectivity as well as solution processability [11], making them easily integrable with new generation PV cells.

The present review explores the numerous capacities and performance parameters achieved by photovoltaic devices fabricated using CNTs, alongside their limitations. The examples of devices provided are based on their ability to achieve higher-than-typical device efficiency, or to overcome the common material and device challenges encountered in CNT incorporation.

Manuscript received July 12, 2021; revised September 1, 2021; accepted September 14, 2021. The work of Vinod Singh was supported in part by the Delhi Technological University under the scheme =Research project Grant to the faculty of DTU DTU/IRD/619/2019/2127, and in part by the AICTE under the scheme –Research Promotion Scheme under Grant 8-123/FDC/RPS(POLICY-1)/2019-2020. *(Corresponding author: Vinod Singh.)*

Sakshi Sharma is with the School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: sakshisharma2510@gmail.com).

Vinod Singh is with the Department of Applied Physics, Delhi Technological University, New Delhi 110042, India (e-mail: vinodsingh@dtu.ac.in).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JPHOTOV.2021.3113317.

Digital Object Identifier 10.1109/JPHOTOV.2021.3113317

2156-3381 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

II. Application in Photovoltaics

A. In Organic Solar Cells

OPVs have garnered much interest over the years owing to their competitive edge over the conventional Si solar cells in terms of a number of advantageous features presented by the use of organic molecules ranging from ease of solution processing, large coefficient of optical absorption, ability to change the band gap and thus, electronic behavior by modifying functional groups of organic polymers, to flexibility and low cost of production, even for large volumes of material [12], [13]. Organic solar cells that employ pi-conjugated polymers or fullerene derivatives are seen as a cost effective solution for photovoltaic applications [14]. These conjugated materials, however, pose inefficiency in trapping the incident light because of the influence of low charge mobility [15], [16]. In addition to this, they also show substantial photochemical degradation [17]. A notable efficiency of 17.3% was demonstrated using tandem structure in OPV cell [18]. Still, improvements in OPVs call for materials that can contribute to enhancing exciton generation rate, longer exciton diffusion lengths, chemical and thermal stabilities, and ultimately higher photo conversion efficiencies (PCE). Thus, hybrid OPVs embedded with the combined benefits of both organic and inorganic materials have been explored to achieve improved values of the abovementioned properties [19].

The typical fabrication processes for OPV deposition utilize ITO coated glass substrates, which are brittle and nonflexible. ITO itself is expensive and cannot be solution processed; thereby limiting its application in future OPV cells for ubiquitous use. Other transparent electrode materials reported in literature for use in OPVs include doped metal oxides (ITO, gallium-doped zinc oxide, aluminium doped zinc oxide) [20]–[23], thin metal layers [20], [24]–[27], polymers such as PEDOT:PSS [28], [29], graphene [30], [31], and Ag nanowires [32], [33]. A comparison between material parameters such as sheet resistance, transparency, and flexibility is presented in Table I. For OPVs fabricated using CNT-based TCFs [34]–[39], the sheet resistance (Rs) continues to limit the PCE to less than desirable values, as summarized in Table II, along with the values of transmittance (T) and PCE.

Several works demonstrate the use of a highly flexible, solution-processed SWCNT film network for transparent anodes [37], [40]–[42]. By means of the transfer printing technique [43], homogenous, large area thin films can be obtained by vacuum-filtering SWCNT dispersion stabilized with sodium dodecyl sulfate surfactant. The films are then raised off using polydimethylsiloxane stamps and transferred onto the polyethylene terephthalate (PET) substrate. For device application [37], the films prepared using this method are spin-coated with poly3, 4-ethylenedioxythiophene:polystyrene sulfonate (PEDOT:PSS), which serves to enhance CNT conductivity and wettability and reduce roughness. The SWNT-based device had a photo-conversion efficiency of 2.5%, paralleling the performance of ITO-based devices (3%), and Rs of 200 Ω, which is relatively quite high (Rs = 40Ω for ITO/glass control device) and consequently impacts the fill factor (FF). The mechanical flexibility and stability for SWCNT/PET, however, was much greater than that for ITO/PET—a notable quality to ensure widespread deployment.

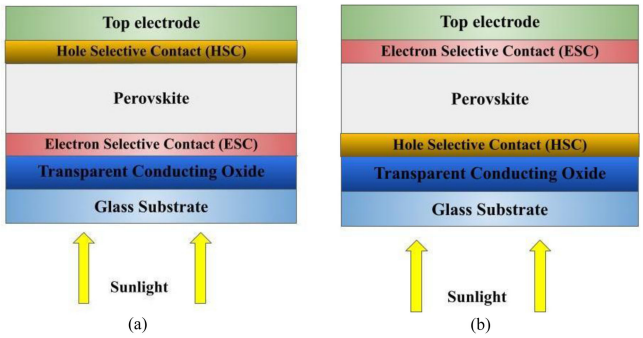


Fig. 1. Schematic diagram of PSC in (a) planar n–i–p architecture, where light passes through the ESC first and (b) planar p–i–n architecture, where light passes through the HSC first.

Jeon *et al.* [36] fabricated CNT-based transparent electrodes to be used in organic solar cells by dry deposition of aerosol chemical vapor deposition fabricated SWCNT films doped with MoO_3. Introduction of MoO_3 by thermal doping increased the transmittance of SWCNT films in comparison with pristine SWCNT, decreased its resistivity and increased the work function from 4.86 eV for undoped SWCNT to 5.4 eV for the doped version (resulting in better hole extraction). For the 90% and 80% SWCNT films with PTH3 donor, a higher transparency around 400–500 nm was observed, beyond which absorption by MoO_x flattens the curve, indicating that P3HT is compatible with 90% SWCNT electrode while PTB7 absorbing longer wavelengths is more suited to 65% SWCNT device. The device efficiency using 65%—SWCNT deposited over glass substrate and a photoactive layer of the low band gap polymer thieno [34-b]thiophene/benzodithiophene (PTB7) (investigated for CNT OPV for the first time in this report) was a record high of 6.04% to date. This value is higher than PI and PET based substrates as the former is limited by low intrinsic transparency and gives a PCE of 3.78% while the latter gave a PCE of 3.91%, ascribed to the fabrication induced damage on MoO3.

B. In Perovskite Solar Cells

CNTs have been incorporated into high efficiency perovskite solar cells (PSC). As Fig. 1 represents, the device architecture consists of perovskite absorber material sandwiched between charge selective contacts (CSC)-electron selective contact/layer and hole selective contact/layer—on either side. The charge carriers photo generated in the perovskite layer travel to ESC and HSC, respectively, and are subsequently transported to the electrodes. CNT based CSC have proved to be successful in steadily increasing device efficiency over the years. These contacts require the essential properties of low electrical resistance, high charge selectivity, inertness to the perovskite material itself as well as resistance to moisture and oxygen so as to establish prolonged device stability.

While some of the most commonly used contact materials for electron selective contact include metal oxides such as TiO2, SnO2, ZnO, or fullerene derivatives, where low temperature fabrication is a requirement (in p–i–n device geometry, for

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
MATERIAL PARAMETERS FOR VARIOUS TRANSPARENT CONDUCTIVE MATERIALS USED IN OPV DEVICES

MATERIAL	R_s (Ω/sq)	TRANSPARENCY (% in visible range)	FLEXIBLE	DRAWBACKS	REFERENCE
Indium Tin Oxide	10-50	80-90	No	Damage during sputter deposition, brittle, costly	20
Al doped zinc oxide	8-25	80-90	No	Damage during sputter deposition, requires high temperature fabrication	22 23
Au metal film	3-511	38-66%	Yes	Sporadic surface roughness, need for controlled deposition	24
Cu wire grid	~30	83	Yes	Complexity of process: requires nanoimprint lithography (NIL)	26
PEDOT:PSS	~200	78-87	Yes	Photo- and thermal degradation	28
Graphene	500-2300 (Growth temperature dependent)	80-90	Yes	Very high sheet resistance, Enhancement in film quality and structure needed	30
Ag nanowire	10-35	85-97	Yes	Corrodes upon oxidation, require buffer coating to prevent shorting	33

example), materials for hole selective contact include organic molecules such as PEDOT:PSS, poly(3- hexylthiophene-25-diyl) (P3HT), Polytriarylamine (PTAA) or inorganic HSCs consisting of Nickel oxide (NiOx), Cuprous thiocyanate (CuSCN), or Copper iodide (CuI) [44]. Despite a large range of available contact materials, factors such as high processing temperatures, heat induced morphology changes, and low carrier mobility that requires introducing dopants into the materials, are significant drawbacks.

CNTs in metal halide perovskite (MHP) solar cells have proved to be advantageous due to its intrinsic high carrier mobility and long range transport [45]. Li *et al.* first incorporated

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE II
MATERIAL PARAMETERS FOR CNT BASED TCF PREPARED WITH DIFFERENT FABRICATION TECHNIQUES FOR USE IN OPV DEVICES. PEDOT:PSS REFERS TO POLY3,4-ETHYLENEDIOXYTHIOPHENE:POLYSTYRENE SULFONATE; P3HT TO POLY(3 -HEXYLTHIOPHENE-25-DIYL); PTB7 TO THIENO [34-B] THIOPHENE/BENZODITHIOPHENE AND PCBM TO PHENYL-C_{61}-BUTYRIC ACID-METHYL ESTER

DEVICE STRUCTURE	CNT-TCF fabrication technique	R_s (Ω/sq)	TRANSMITTANCE (%)	PCE (%)	REFERENCE
Glass/CNT/PEDOT:PSS/P3HT:PCBM/Ca/Al	Ultrasonic spray	56	65	4.1	34
Glass/CNT/PEDOT:PSS/P3HT:PCBM/LiF/Al	Dip coating	188	92	2	35
Glass/CNT-MoOx/PEDOT:PSS/PTB7:PCBM/LiF/Al	Filtration	84	65	6	36
PET/CNT-MoOx/PEDOT:PSS/PTB7:PCBM/LiF/Al PET: polyethylene terephthalate	Filtration	84	65	3.9	36
PET/CNT/PEDOT:PSS/P3HT:PCBM/Al	Filtration	200	85	2.5	37
Glass/CNT-CuI/PEDOT:PSS/DIP/DBP/C_{60}/BCP/Al DIP: diindenoperylene DBP: tetraphenyl benzoperiflanthene BCP: bathocuproine	Doctor Blade	70	80	2.9	38
Glass/CNT/PEDOT:PSS/P3HT:PCBM/LiF/Al	Spray-coating	128	90	2.3	39

CNT in MHP architecture in 2014 by replacing metal counter electrode (Au) with a network of CNT films, achieving a device efficiency of 6.87% [46]. Habisreutinger *et al.* addressed the issue of thermal degradation in PSC by substituting organic HTC in favor of SWCNT/ Polymer composite [47]. Use of CNT electrodes eliminates the need for vacuum based thermal deposition of metals, thereby bringing down fabrication cost as well as metal ion migration. However, the major drawbacks consist of inadequate conductivity of CNT as compared with its metal counterparts and non-reflectivity, which decreases the short circuit current. This has resulted in limited PCE of CNT electrode-based PSCs [48].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

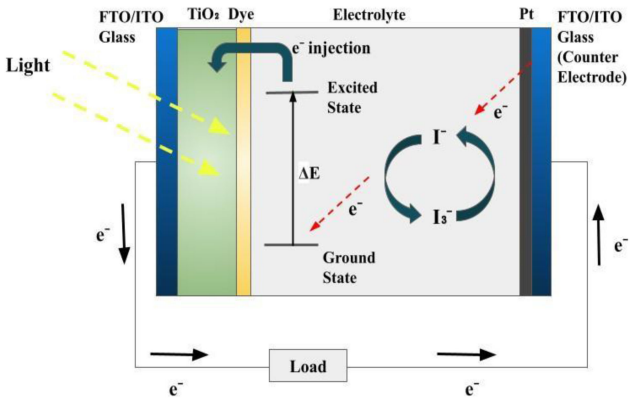


Fig. 2. Schematic structure and working of DSSC.

Zheng *et al.* in 2017 detailed a boron doped MWCNT (B-MWCNT) based counter electrode [49]. The solar cell PCE increased from 10.7% in case of undoped MWCNT electrodes to 14.6% in B-MWCNT electrodes. This is owed to the higher hole carrier extraction and conductivity resulting from the introduction of boron. Other examples of doped CNT electrodes contributing to higher device performances in literature include diluted HNO_3 doped CNT leading to increase in conductivity, transmittance of CNT films as well as PCE (6.3%) and MoO_3 doped CNT leading to MoO3-thickness dependent PCE with a thicker layer (6nm) resulting in energy level misalignment and thus, reduced PCE as compared with a thin layer (2 nm) [50]. Despite its proven role in upgrading device performance, doping is not a straightforward process with CNT electrodes as the dopant chemical damages the underlying perovskite layer. Research to find a solution has resulted in one recent work highlighting doping of Triflic acid (TFMS) over CNT sheets by vapor assisted doping method. This reported a device efficiency of 17.6% without damaging the underlying layer [51].

C. In Dye Sensitized Solar Cells

DSSC, in recent years, have found recognition as a substitute to traditionally employed Si solar cells [52], [53]. These hybrid, third generation cells work on the principle of conversion of photons that strike the organic molecular dyes (Photosensitizer) into electricity by transferring electrons out of the excited dye molecules to the conduction band of oxide semiconductor (TiO_2) that forms the working electrode. This process is depicted in Fig. 2 along with the components of the solar cell. Dye is sandwiched between two Fluorine-doped tin oxide (FTO)/ ITO coated conductive glass substrates that work as electrodes; working electrode with a coating of titanium dioxide nanoparticles and counter electrode that comprises of Platinum coated FTO glass, Pt working as the cathode material [54]. It is here that the role of carbon nanomaterials namely graphene, nanodiamond, and CNTs becomes noteworthy as Pt, although an efficient cathode material is costly and less abundant. Nanocarbon is seen as a replacement for Platinum in counter electrodes or photoanode additives because of its conductivity and efficient electro-catalytic activity [55], [56].

First DSSCs employing nanoporous TiO_2 photoanode, that improved the effective surface area of the working electrode, thereby leading to enhanced dye absorption, were an invention credited to the efforts of Brian O'Regan and Michael Gratzel at UC Berkeley [52], [57]. These cells showed light harvesting efficiencies of 7%, comparable with amorphous Si-solar cells. In the following years, efficiency of around 11% was also demonstrated [58]. But, DSSCs provide low cost and technologically simpler solutions to harvesting solar energy than Silicon cells. Their growing popularity has propelled research into ways to improve their efficiency.

Prasetio *et al.* [59] undertook the task of fabricating CNT based counter electrodes by depositing CNT films with varying masses (0.01 g, 0.02 g, 0.04 g) onto the FTO substrate through the doctor blade method. The performance parameters in terms of J_{sc} and efficiency were observed to be the highest for CNT-0.04 (6.413 mA/cm² and 0.91% respectively)—an effect ascribed to its relatively higher electrocatalytic activity compared to the other samples, although the values are still lesser than those for Pt electrodes. A higher efficiency along with the short circuit current density relative to the Pt/Si control sample ($J_{sc} = 20.6$mA/cm²; $\eta = 6.29\%$) was obtained in another experiment performed by Yilmaz *et al.* [60]. They deposited vertically aligned CNT (VACNT) over n-type Si substrate through hot filament chemical vapour deposition (HFCVD) process in planar and microarray patterned arrangements (P-CNT and MA-CNT, respectively). The improved efficiency of P-CNT (7.13%) over Pt is believed to be the result of its improved catalytic activity. The vertical CNT geometry is notable in that its 3-D network serves to make available a large contact area for the electrolyte to interact with the electrode, which consequently improves the charge transfer at CNT/electrolyte interface, and can be seen in the higher value of J_{sc} obtained (26.3 mA/cm²). Despite such improvements, a lowering of FF can be seen, possibly due to larger series resistance of the relatively thicker CNT/substrate system. MACNT electrodes, due to a lower amount of VACNT—leading to a lower contact area offered to the electrolyte- reported a drop in photovoltaic performance ($J_{sc} = 22.1$ mA/cm²; $\eta = 5.47\%$) as compared to P-CNT.

D. In Silicon Based Solar Cells

Crystalline Si based PV devices have secured a dominant portion of the global market share roughly for the last 30 years. This dominance of Si is based on multiple factors—abundance, near optimum band gap of 1.1 eV, nontoxicity, available infrastructure, high power conversion efficiencies (record efficiency of 26.6% has been reported for the heterojunction—interdigitated back contact technology [61]), and long lifetime. While efforts to overcome the theoretical PCE limit of 29% are underway and have yielded results in perovskite/silicon tandem solar cells [62], [63], conventional Si PV presents certain challenges [64]. First, obtaining pure, elemental silicon from the silicon dioxide found in nature requires pre-processing, which adds to the total production cost. In addition to this, silver used for silicon cell coating has been steadily becoming costlier and is potentially being replaced with copper [65]. Second, silicon solar panels

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

are rigid and heavy, negatively affecting their space efficiency as well as making large scale installation expensive. Third, while research challenging PCE limits in Si PVs exists, such reports are few and far between, at least where commercially attainable efficiencies are concerned [66].

Since Wei *et al.* first described CNT/Si heterojunction solar cells having efficiency of mere 1.31% [67], there has been an appreciable improvement in PCE values of these devices, with reports of PCE reaching 17% [68]. CNT/Si heterojunctions work in solar cells due to exceptional optical transparency of CNT films backed by good absorption of Si [69]. This effectively means that most of the incident sunlight is allowed to enter CNT and be subsequently absorbed by Si, where electron hole pairs are generated. Post charge separation at the interface, electrons travel to the n type Si and holes to the CNT film. One major benefit of this device architecture is that charge diffusion does not require a high temperature, as in case of traditional p–n junctions [70]. Nonetheless, whether the CNT/Si heterojunction functions as a traditional p–n junction or as a Schottky junction can be hard to understand due to the coexistence of both metallic and semiconducting states in CNT films [71]. Limitations of the window geometry used in Si/CNT heterojunction solar cells also arise due to the fact that presence of a window decreases the effective surface area, where photo-absorption should ideally take place. Scaling up this architecture also compromises the PCE [72].

Solar cell performance in CNT/Si cells also depends on the fabrication method, which influences the film property [73], as shown by the examples listed in Table III. For instance, dry process SWCNT film prepared by aerosol chemical vapour deposition (CVD) showed properties of high optical transparency, conductivity, and crystallinity, resulting in a record high PCE nearing 11% [74]. A report by Wang *et al.* utilized the floating catalyst CVD method to fabricate uniform SWCNT films having low resistivity and high transparency. Upon investigating the optimum device structure by varying the film thickness, transmittance, and the window size, the cell exhibited an efficiency of 12.5% without doping [75]. On the other hand, CNT films prepared by wet treatment typically result in lower cell PCE [76]–[78]. In [76], SWCNT network films prepared by sonication resulted in PCE of 2.4% due to high sheet resistance. A PCE of 5.93% was reported in [77] due to poor charge transfer caused by the development of a thick oxide layer on silicon; the efficiency had to be increased by the addition of antireflective polymer layers. One prominent issue responsible for the low cell PCE of heterojunction devices is the lack of surface passivation of silicon substrate, which results in high carrier recombination rate [79]. Recently, to overcome this issue, an approach utilizing solution—processable CNT:Nafion ink spin coated onto the back junction of industrially viable Si cells was presented [80]. Nafion works as a surface passivation agent, whereas CNT works as the HSC. A record PCE of 21.4% was reported. The CNT:Nafion solution has a stability period of six months, can be processed in the ambient atmosphere and is directly transferable onto the wafer, providing the material reproducibility required at the industrial scale. A current drawback, however, is that of long term performance degradation due to the hydrophobic nature of the sulfonic functional group present in Nafion.

E. In Thin Film Solar Cells

Thin film PV architecture has been heralded as the technology of the future. Thin film silicon solar cells, apart from the inherent advantages of silicon as discussed in preceding paragraphs, present the benefit of uniform appearance, low weight, flexibility (hence diversifying their application range), and ease of patterning. However, their PCE lies around 7%–10% for commercial modules [81]. While many studies exist for hybrid devices based on crystalline Si, amorphous Si, although less explored, presents the advantages of a high photo to dark conductivity ratio (making it ideal for indoor applications), low cost and processability using a relatively simple CVD process at room temperature [82]. One significant drawback arises from low carrier mobility in a-Si:H, necessitating a transparent conducting layer of ITO as the top contact. In addition, amorphous Si (a-Si:H) contains more defects than its crystalline counterpart (c-Si) and suffers from the Staebler–Wronski effect, i.e., degradation in photoconductivity upon long exposure to light [83]. Thus, efforts are focused on developing thin film technology that is not only cost effective, but also contributes to the search for ITO free PV devices while increasing efficiency.

To this end, CNT marks as a promising choice of material for transparent electrodes [84], [85], p-type window layer [86], light absorber [87], and light trapping structures [88]. SWCNT has been used as semitransparent electrodes that can replace conventional (ITO)/ZnO based electrodes in a-Si:H based solar cells [84]. Schottky barrier solar cells, formed of a-Si:H substrate deposited by Plasma enhanced CVD and SWCNTs spray deposited atop, reported external quantum efficiency value of 35% at wavelength around 470 nm. While SWCNT acted as a good substitute for ITO, they exhibited poor hole collection efficiency. A thin film hybrid solar cell based on PEDOT:PSS-SWCNT composite film employed as the window layer and transparent top electrode reported a PCE of 2.7% [86]. Coating of PEDOT:PSS polymer layer over the SWCNT synthesized by aerosol CVD is primarily for the purpose of conductivity enhancement, but also leads to decreased sheet resistance and increased work function of the composite film relative to pristine SWCNT films. While adding PMMA as an antireflective coating improved the PCE to 3.4%, this value is still a ways off for practical purposes.

Considering that Cadmium Telluride (CdTe) is the most prevalent variety of solar cells after c-Si available in the global market today [89], it seems imperative to discuss the impact of CNT incorporation on their performance. Structure of CdTe solar cells, as depicted in Fig. 3, consists of CdTe thin film as the p type semiconductor (where photo-absorption takes place), transparent conducting oxide (TCO) layer commonly made up of SnO_2, Cd_2SnO_4, ITO, or ZnO:Al and the intermediate n type CdS layer. This is deposited typically over a soda-lime glass substrate and enveloped on the other side by a back electrode (BC) formed of metals, MoO_x, ZnO/Al, ZnTe or NiO [90]. The high resistance transparent oxide layer, typically tin oxide (TO), and zinc oxide (ZnO), inserted between TCO and CdS, helps in obtaining optimized energy band alignment [91].

US based company First Solar set up a new benchmark of 22.1% efficiency obtained in laboratory for CdTe thin film PV

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III
PERFORMANCE PARAMETERS FOR CNT/SI HETEROJUNCTION SOLAR CELLS INCORPORATING CNT FILMS PREPARED THROUGH WET AND DRY PROCESSES

CNT FILM FABRICATION METHOD	J_{sc} (mA/cm^2)	V_{oc} (V)	PCE (%)	REMARK	REFERENCE
Dry Process	33.8	0.589	14.2	The values are reported for an optimised CNT/Si solar cell having an active area of 2.3mm^2. For larger active areas, PCE decreases due to a higher carrier recombination as well as series resistance.	71
	29.7	0.535	10.8	Highest PCE (10.8%), obtained for the sample TCF-80 (film with 80% transmittance) is ascribed to its superior crystallinity and long SWCNT tube length.	72
	30.2	0.586	12.5	Highest device PCE without carrier doping is achieved through optimization of SWCNT film thickness, transmittance (91%) and the window size (1 mm diameter).	73
Wet Process	14.6	0.39	2.4	For maximum PCE, an optimum CNT film thickness that balances between increased effective area of p-n junction and increased photo-absorption should be achieved.	74
	18.4	0.50	5.93	Low PCE is due to poor charge transfer caused by formation of a thick oxide layer on silicon, and is enhanced by the addition of anti-reflective polymer layers.	75
	26.1	0.51	8.52	Increase in PCE and Jsc is expected to be due to light concentration effect on hydrophobic SWCNT films.	76

cells in 2016, improving on its 21.5% value obtained in the preceding year [92] while also reporting average commercial module efficiency of ∼18% in 2020 [89]. Nevertheless, it suffers from several drawbacks, namely carrier recombination at the CdS-CdTe interface, insufficient transparency of TCO and window layer and lack of ohmic contact at the CdTe-metal BC interface [93].

SWCNT as transparent conductive thin films have found application as BC in CdTe solar cells [94]–[96]. Li *et al.* [96] proposed a CuTe nanoparticle (CuTe NPs)/CNTs paste as BC material, and obtained a maximum Voc of 824 mV for a CNT mass of 0.282 g in the paste and annealing temperature of 200 °C.

The performance parameters including FF, V_{oc}, and J_{sc} varied depending on the back contact annealing temperature and CNT mass as both these factors affect the diffusion of Cu ions in the CdTe layer. The effects of incorporating CNTs to Carbon BC in polycrystalline CdTe solar cells have also been explored and manifest as an increased value of FF resulting from lower

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
PERFORMANCE PARAMETERS FOR CDTE THIN FILM SOLAR CELLS INCORPORATING CNT BASED BACK CONTACT

BACK CONTACT	Jsc (mA/cm^2)	Voc (V)	PCE (%)	REMARK	REFERENCE
CuTe/SWCNT-network	22.25	0.818	12.4	The best device efficiency is obtained with 40-50% transmission in the 800-1500nm range. In addition to granting high near-infrared (NIR) transparency, low reflectivity and good hole selectivity, SWCNT networks are compatible with low cost deposition.	93
Au/SWCNT	21.15	0.773	11	Addition of Au to SWCNT electrode enhances the carrier collection efficiency at longer wavelength (700-800nm), which also reflects in the Jsc improvement of Au/SWCNT contacts over pure SWCNT contacts.	94
CuTe NPs/ CNT paste	-	0.824	~13	Maximum Voc and PCE are obtained for CNT mass of 0.282g and annealing temperature of 200°C. Fill Factor, Voc and Jsc vary with Back Contact annealing temperature and CNT mass as both these factors affect the diffusion of Cu ions in the CdTe layer.	95
CNT doped carbon	25	0.829	14.1	Incorporating CNTs in Carbon BC in polycrystalline CdTe solar cells results in higher Fill Factor (decrease in series resistance), higher Voc (decrease in back barrier) and overall improvement in PCE with CNT concentration from 1 to 5% by wt.	96
Au /GO doped SWCNT	14	0.65	-	GO doping rectifies the work function mismatch between pure SWCNT and CdTe by increasing SWCNT work function and barrier degradation (leading to better hole transport) at the CdTe/SWCNT interface.	97

series resistance, a higher V_{oc} due to decrease in back barrier and overall improvement in PCE with CNT concentration from 1% to 5% by weight [97]. Both, reduction in series resistance for the solar cell and improvement in FF were dependent on resistivity of carbon BC, which decreased with increasing CNT concentration upto 5wt%.

Composite thin films of SWCNT doped with graphene oxide (GO-doped SWCNT) have also found use as back contacts [98]. Compared with standard Cu/Au contacts, the V_{oc} of SWCNT/Au contact turned out to be significantly lower due to the work function mismatch between pure SWCNT (~4.9 eV) and CdTe (~5.15 eV). This was rectified by GO doping; the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

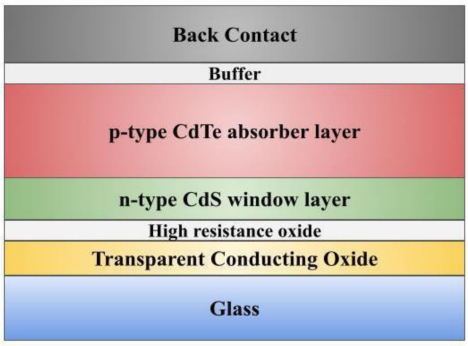


Fig. 3. Schematic diagram of a standard CdTe thin film solar cell.

device V_{oc} reported for the GO- SWCNT/Au combination ($\sim$0.6 V) exceeded that observed for devices with SWCNT/Au based BC (0.4–0.5 V)—an improvement attributable to the increase in SWCNT work function and barrier degradation at the CdTe/SWCNT interface (leading to better hole transport) caused by GO addition. Table IV recapitulates these devices consisting of CNT based back contacts.

III. Challenges and Outlook

One looming shortfall for CNT based solar cells is their inability to match the efficiency achievable through traditional Silicon based devices or the newer thin film solar cells. Despite this, scientists remain convinced of the untapped opportunities that the material provides, thanks to its low temperature processing, flexibility, cost effectiveness, and general versatility. Out of the numerous chiralities obtainable from thin, rolled up sheets of carbon into a nanotube, scientists commonly chose one specific chirality with the ability to absorb only a narrow range of optical wavelengths. However, a combination of poly-chiral nanotubes makes it possible to optimize photocurrent by harvesting light from a broader range of the electromagnetic spectrum. This includes previously unexplored parts in thin film photovoltaics such as the near infrared regime. But while this grants researchers the capability to develop solar cells with a wide range light absorption by selecting an appropriate chirality, it is a challenge to fabricate chirality selective CNT films on significantly large scales [99].

Another standard challenge facing integration of CNT into photovoltaic devices is the stability. While several experiments have reported long term stabilities for CNT/Si heterojunction solar cells, as highlighted in Table III, reports of high efficiency devices with PCE exceeding 20% that also analyze their stabilities are lacking. As the photoactive layer in OPV devices, CNT based cells have not been able to achieve efficiencies equaling those of the best tandem polymer solar cells [100]. One issue is to separate metallic SWCNT (m-SWCNT/) from semiconducting SWCNT (s-SWCNT), which coexist upon synthesis, so that the formation of Schottky barrier at the m- SWCNT/s-SWCNT junction-that reduces hole transmission probability-can be prevented [101]. Elimination of m-SWCNT maximizes the semiconducting performance and is achieved through density gradient ultracentrifugation (DGU), which sorts CNTs by diameter, band gap, and electronic behavior type [102]. In the case of functionalized CNT, for example, in conjugation with polymer systems [103], the morphology of functionalized nanocomposites influences the optical properties of the material and therefore, the overall photovoltaic performance of the device [104]. While the efficiency values for CNT/polymer hybrid cells hitherto achieved lie nowhere close to those of other PV technologies—such as thin film or Si-polymer solar cells still prove to be a commercially cheaper alternative.

Another important avenue for research moving forward is the thermal, environmental, and light stability of the material itself, considering that commercial solar cells would continually be exposed to the solar radiation and other natural elements. Research on thermal treatment of SWCNT through a combination of TEM, XRD, and Raman Spectroscopy shows that the material structure shows appreciable change within the 1600–2800 °C range such as decrease in crystalline order. This is believed to be due to coalescence of SWCNT, also increasing the tube diameter (2–4 nm) alongside the formation of bundles. At 2200 °C, SWCNT modify to MWCNT, eventually reaching outer diameters of 7 nm with multiple carbon layers [105]. Incorporation of semiconducting SWCNT in P3HT:PTBM device results in limiting severe crystallization of polymer molecules and thereby providing higher thermal stability ($<$5% decrease in PCE after 9 h of annealing) as compared with devices without SWCNT ($>$25% degradation in PCE for the same period). This unwanted decay is due to phase separation in bulk heterojunction active material, which leads to poor charge generation and exciton recombination. [106]

Optical characterization of P3HT/SWCNT devices revealed interesting facts [107]. While the absorption range of pristine SWCNT is expected to be around 1600 nm for S11 and 1000 nm for S22 transitions, the peaks obtained for SWCNT after deducting the P3HT spectrum from P3HT-SWCNT active layer spectrum is blue shifted and is present at 1200 and 900 nm for S11 and S22 transitions, respectively. This is expected to be due to hybridization between SWCNT valence subband and the P3HT LUMO. Based on efficiency analysis, CNT contributes to a higher carrier generation in the NIR regime, with a maximum internal quantum efficiency of $\sim$26%.

Celik *et al.* studied the environmental impact of CNT based solar cells using the life cycle assessment methodology, with mono-Si as the benchmark due to its widespread commercial use [108]. Environmental impact in this work was used to refer to potential for acidification, ecotoxicity, human toxicity, ozone depletion layer, and global warming to name a few. In comparison with other PV materials such as a-Si, mono-Si, poply-Si, CdTe, and CIGS on a kWh^{-1} basis, 1% SWCNT, with a lifetime of 3 years and average normalized impact of $\sim$18 turned out to be lease environmentally preferable, while a-Si was the most favorable option with a normalized impact of 0.48. On the other hand, 28% SWCNT shows some promising trends. In almost all categories, this device's impact values were close to other technologies, with its overall normalized impact equal to 0.76. Factors like moisture, oxygen, and light responsible for degradation in organic PV materials will not affect SWCMTs. Therefore, the lifetime assumption of three years and the impact estimates presented in this work for 28% SWCNTs could very well be an underestimation.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

With the global CNT market share expected to rise to USD 1714 million in 2026 from the current USD 876 million, and the energy and storage sector being the largest end use industrial avenue [109], CNT integration in semiconductor as a fiscally plausible objective has garnered much attention. Part of the motivation to employ more efficient photovoltaic systems stems from an unprecedented need to mitigate global carbon emissions and environmental effects including climate change by moving toward efficient, accessible, and market-friendly renewable energy sources capable of providing energy security [110], [111]. Therefore, it seems probable that despite the current limitations with the material, CNT will continue to engage the research interest as a material for upcoming flexible photovoltaics on part of both inventors and investors.

REFERENCES

[1] S. Iijima, "Helical microtubules of graphitic carbon," *Nature*, vol. 354, no. 6348, pp. 56–58, Nov. 1991, doi: 10.1038/354056a0.

[2] M. Berger, "Carbon nanotubes–what they are, how they are made, what they are used for," Apr. 2014. [Online]. Available: https://www.nanowerk.com/nanotechnology/introduction/introduction_to_nanotechnology_22.php

[3] M. Melchionna, S. Marchesan, M. Prato, and P. Fornasiero, "Carbon nanotubes and catalysis: The many facets of a successful marriage," *Catal. Sci. Technol.*, vol. 5, no. 8, pp. 3859–3875, Jul. 2015, doi: 10.1039/C5CY00651A.

[4] B. Ham, "Carbon nanotube transistors make the leap from lab to factory floor," Jun. 2020. Accessed: Oct. 15, 2021. [Online]. Available: https://news.mit.edu/2020/carbon-nanotube-transistors-factory-0601

[5] V. J. Kouzegaran, "Commercial value and benefits of carbon nanotubes transistors," Jun. 2021. Accessed: Oct. 15, 2021. [Online]. Available: https://nanografi.com/blog/commercial-value-and-benefits-of-carbon-nanotubes-transistors

[6] E. Gent, "Silicon is reaching its limits. Up next: Carbon nanotubes," Sep. 2019. Accessed: Oct. 15, 2021. [Online]. Available: https://singularityhub.com/2019/09/03/silicon-is-reaching-its-limits-are-carbon-nanotubes-next

[7] Y. Sun and H. H. Wang, "High- performance, flexible hydrogen sensors that use carbon nanotubes decorated with palladium nanoparticles," *Adv. Mater.*, vol. 19, no. 19, pp. 2818–2823, Oct. 2007, doi: 10.1002/adma.200602975.

[8] Y. Zhou and R. Azumi, "Carbon nanotube based transparent conductive films: Progress, challenges, and perspectives," *Sci. Technol. Adv. Mater.*, vol. 17, no. 1, pp. 493–516, Sep. 2016, doi: 10.1080/14686996.2016.1214526.

[9] X. Fu *et al.*, "A fiber-shaped solar cell showing a record power conversion efficiency of 10%," *J. Mater. Chem. A Mater. Energy Sustain.*, vol. 6, no. 1, pp. 45–51, 2018, doi: 10.1039/C7TA08637G.

[10] P. M. Rajanna, P. D. Lund, and A. G. Nasibulin, "Hybrid heterojunction solar cells based on single-walled carbon nanotubes and amorphous silicon thin films," *Wiley Interdiscip. Rev. Energy Environ.*, May 2021, doi: 10.1002/wene.402.

[11] S. Qiu, K. Wu, B. Gao, L. Li, H. Jin, and Q. Li, "Solution-processing of high- purity semiconducting single-walled carbon nanotubes for electronics devices," *Adv. Mater.*, vol. 31, no. 9, Mar. 2019, Art. no. e1800750, doi: 10.1002/adma.201800750.

[12] J. Nelson, "Polymer: Fullerene bulk heterojunction solar cells," *Mater. Today*, vol. 14, no. 10, pp. 462–470, Oct. 2011, doi: 10.1016/s1369-7021(11)70210-3.

[13] N. Yeh and P. Yeh, "Organic solar cells: Their developments and potentials," *Renewable Sustain. Energy Rev.*, vol. 21, pp. 421–431, May 2013, doi: 10.1016/j.rser.2012.12.046.

[14] Y. Liang *et al.*, "For the bright future-bulk heterojunction polymer solar cells with power conversion efficiency of 7.4%," *Adv. Mater.*, vol. 22, no. 20, pp. E135–E138, May 2010, doi: 10.1002/adma.200903528.

[15] M. M. Mandoc, L. J. A. Koster, and P. W. M. Blom, "Optimum charge carrier mobility in organic solar cells," *Appl. Phys. Lett.*, vol. 90, no. 13, Mar. 2007, Art. no. 133504, doi: 10.1063/1.2711534.

[16] S. Shoaee, M. Stolterfoht, and D. Neher, "The role of mobility on charge generation, recombination, and extraction in polymer-based solar cells," *Adv. Energy Mater.*, vol. 8, Apr. 2018, Art. no. 1703355, doi: 10.1002/aenm.201703355.

[17] M. Jørgensen, K. Norrman, and F. C. Krebs, "Stability/degradation of polymer solar cells," *Sol. Energy Mater. Sol. Cells*, vol. 92, no. 7, pp. 686–714, Jul. 2008, doi: 10.1016/j.solmat.2008.01.005.

[18] L. Meng *et al.*, "Organic and solution- processed tandem solar cells with 17.3% efficiency," *Science*, vol. 361, no. 6407, pp. 1094–1098, Sep. 2018, doi: 10.1126/science.aat2612.

[19] B. Sun, H. J. Snaith, A. S. Dhoot, S. Westenhoff, and N. C. Greenham, "Vertically segregated hybrid blends for photovoltaic devices with improved efficiency," *J. Appl. Phys.*, vol. 97, no. 1, Jan. 2005, Art. no. 014914, doi: 10.1063/1.1804613.

[20] J.-Y. Lee, S. T. Connor, Y. Cui, and P. Peumans, "Solution-processed metal nanowire mesh transparent electrodes," *Nano Lett.*, vol. 8, no. 2, pp. 689–692, Feb. 2008, doi: 10.1021/nl073296g.

[21] V. Bhosle, J. T. Prater, F. Yang, D. Burk, S. R. Forrest, and J. Narayan, "Gallium-doped zinc oxide films as transparent electrodes for organic solar cell applications," *J. Appl. Phys.*, vol. 102, no. 2, Jul. 2007, Art. no. 023501, doi: 10.1063/1.2750410.

[22] H. Saarenpää, T. Niemi, A. Tukiainen, H. Lemmetyinen, and N. Tkachenko, "Aluminum doped zinc oxide films grown by atomic layer deposition for organic photovoltaic devices," *Sol. Energy Mater. Sol. Cells*, vol. 94, no. 8, pp. 1379–1383, Aug. 2010, doi: 10.1016/j.solmat.2010.04.006.

[23] L. Sims, H.-J. Egelhaaf, J. A. Hauch, F. R.Kogler, and R. Steim, "1.20 - Plastic Solar cells," in *Comprehensive Renewable Energy*, Sayigh, Ed. Oxford: Elsevier, 2012, pp. 439–480.

[24] S. D. Yambem, A. Haldar, K.-S. Liao, E. P. Dillon, A. R. Barron, and S. A. Curran, "Optimization of organic solar cells with thin film Au as anode," *Sol. Energy Mater. Sol. Cells*, vol. 95, no. 8, pp. 2424–2430, Aug. 2011, doi: 10.1016/j.solmat.2011.04.019.

[25] N. Formica, D. Sundar Ghosh, T. L. Chen, C. Eickhoff, I. Bruder, and V. Pruneri, "Highly stable Ag–Ni based transparent electrodes on PET substrates for flexible organic solar cells," *Sol. Energy Mater. Sol. Cells*, vol. 107, pp. 63–68, Dec. 2012, doi: 10.1016/j.solmat.2012.08.002.

[26] M.-G. Kang, M.-S. Kim, J. Kim, and L. J. Guo, "Organic solar cells using nanoimprinted transparent metal electrodes," *Adv. Mater.*, vol. 20, no. 23, pp. 4408–4413, Dec. 2008, doi: 10.1002/adma.200800750.

[27] J.-S. Yu *et al.*, "Transparent conductive film with printable embedded patterns for organic solar cells," *Sol. Energy Mater. Sol. Cells*, vol. 109, pp. 142–147, Feb. 2013, doi: 10.1016/j.solmat.2012.10.013.

[28] D. Gupta, M. M. Wienk, and R. A. J. Janssen, "Efficient polymer solar cells on opaque substrates with a laminated PEDOT:PSS top electrode," *Adv. Energy Mater.*, vol. 3, no. 6, pp. 782–787, Jun. 2013, doi: 10.1002/aenm.201201061.

[29] Y. H. Kim, C. Sachse, M. L. Machala, C. May, L. Müller-Meskamp, and K. Leo, "Highly conductive PEDOT:PSS electrode with optimized solvent and thermal post- treatment for ITO-free organic solar cells," *Adv. Funct. Mater.*, vol. 21, no. 6, pp. 1076–1081, Mar. 2011, doi: 10.1002/adfm.201002290.

[30] M. Choe *et al.*, "Efficient bulk- heterojunction photovoltaic cells with transparent multi-layer graphene electrodes," *Org. Electron.*, vol. 11, no. 11, pp. 1864–1869, Nov. 2010, doi: 10.1016/j.orgel.2010.08.018.

[31] F. Brunetti *et al.*, "Doped and textured graphene as electrode for organic solar cells," in *Proc. IEEE 15th Int. Conf. Nanotechnol.*, Jul. 2015, pp. 560–563, doi: 10.1109/NANO.2015.7388666.

[32] J. Krantz, M. Richter, S. Spallek, E. Spiecker, and C. J. Brabec, "Solution-processed metallic nanowire electrodes as indium tin oxide replacement for thin-film solar cells," *Adv. Funct. Mater.*, vol. 21, no. 24, pp. 4784–4787, Dec. 2011, doi: 10.1002/adfm.201100457.

[33] D.-S. Leem, A. Edwards, M. Faist, J. Nelson, D. D. C. Bradley, and J. C. de Mello, "Efficient organic solar cells with solution-processed silver nanowire electrodes," *Adv. Mater.*, vol. 23, no. 38, pp. 4371–4375, Oct. 2011, doi: 10.1002/adma.201100871.

[34] T. M. Barnes *et al.*, "Carbon nanotube network electrodes enabling efficient organic solar cells without a hole transport layer," *Appl. Phys. Lett.*, vol. 96, no. 24, Jun. 2010, Art. no. 243309, doi: 10.1063/1.3453445.

[35] T. P. Tyler, R. E. Brock, H. J. Karmel, T. J. Marks, and M. C. Hersam, "Electronically monodisperse single-walled carbon nanotube thin films as transparent conducting anodes in organic photovoltaic devices," *Adv. Energy Mater.*, vol. 1, no. 5, pp. 785–791, Oct. 2011, doi: 10.1002/aenm.201100274.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[36] I. Jeon *et al.*, "Direct and dry deposited single-walled carbon nanotube films doped with MoO(x) as electron-blocking transparent electrodes for flexible organic solar cells," *J. Amer. Chem. Soc.*, vol. 137, no. 25, pp. 7982–7985, Jul. 2015, doi: 10.1021/jacs.5b03739.

[37] M. W. Rowell *et al.*, "Organic solar cells with carbon nanotube network electrodes," *Appl. Phys. Lett.*, vol. 88, no. 23, Jun. 2006, Art. no. 233506, doi: 10.1063/1.2209887.

[38] Y. Zhou *et al.*, "Fabrication of carbon nanotube hybrid films as transparent electrodes for small-molecule photovoltaic cells," *RSC Adv.*, vol. 6, no. 30, pp. 25062–25069, Mar. 2016, doi 10.1080/14686996.2016.1214526.

[39] S. Kim, J. Yim, X. Wang, D. D. C. Bradley, S. Lee, and J. C. de Mello, "Spin- and spray- deposited single-walled carbon-nanotube electrodes for organic solar cells," *Adv. Funct. Mater.*, vol. 20, no. 14, pp. 2310–2316, Jun. 2010, doi: 10.1002/adfm.200902369.

[40] R. C. Tenent *et al.*, "Ultrasmooth, large-area, high-uniformity conductive transparent single-walled-carbon-nanotube films for photovoltaics produced by ultrasonic spraying," *Adv. Mater.*, vol. 21, no. 31, pp. 3210–3216, Aug. 2009, doi: 10.1002/adma.200803551.

[41] J. Wan *et al.*, "Solution-processed transparent conducting electrodes for flexible organic solar cells with 16.61% efficiency," *Nano-Micro Lett.*, vol. 13, no. 1, Jan. 2021, Art. no. 44, doi: 10.1007/s40820-020-00566-3.

[42] J. Wang, J. Zhang, A. K. Sundramoorthy, P. Chen, and M. B. Chan-Park, "Solution processed flexible transparent conductors based on carbon nanotubes and silver grid hybrid films," *Nanoscale*, vol. 6, no. 9, pp. 4560–4565, May 2014 doi: 10.1039/c3nr06386k.

[43] Y. Zhou, L. Hu, and G. Grüner, "A method of printing carbon nanotube thin film," *Appl. Phys. Lett.*, vol. 88, no. 12, pp. 123109–123109–3, Mar. 2006, doi: 10.1063/1.2187945.

[44] B. Roose, Q. Wang, and A. Abate, "The role of charge selective contacts in perovskite solar cell stability," *Adv. Energy Mater.*, vol. 9, Dec. 2018, Art. no. 1803140, doi: 10.1002/aenm.201803140.

[45] S. N. Habisreutinger and J. L. Blackburn, "Carbon nanotubes in high-performance perovskite photovoltaics and other emerging optoelectronic applications," *J. Appl. Phys.*, vol. 129, no. 1, Jan. 2021, Art. no. 10903, doi: 10.1063/5.0035864.

[46] Z. Li *et al.*, "Laminated carbonnanotube networks for metal electrode-free efficient perovskite solar cells," *ACS Nano*, vol. 8, no. 7, pp. 6797–6804, Jul. 2014, doi: 10.1021/nn501096h.

[47] S. N. Habisreutinger, T. Leijtens, G. E. Eperon, S. D. Stranks, R. J. Nicholas, and H. J. Snaith, "Carbon nanotube/polymer composites as a highly stable hole collection layer in perovskite solar cells," *Nano Lett*, vol. 14, no. 10, pp. 5561–5568, Oct. 2014, doi: 10.1021/nl501982b.

[48] I. Jeon *et al.*, "Carbon nanotubes to outperform metal electrodes in perovskite solar cells via dopant engineering and hole- selectivity enhancement," *J. Mater. Chem. A Mater. Energy Sustain.*, vol. 8, no. 22, pp. 11141–11147, Jun. 2020, doi: 10.1039/D0TA03692G.

[49] X. Zheng *et al.*, "Boron doping of multiwalled carbon nanotubes significantly enhances hole extraction in carbon-based perovskite solar cells," *Nano Lett.*, vol. 17, no. 4, pp. 2496–2505, Apr. 2017, doi: 10.1021/acs.nanolett.7b00200.

[50] Q. Luo *et al.*, "Recent advances in carbon nanotube utilizations in perovskite solar cells," *Adv. Funct. Mater.*, vol. 31, Oct. 2021, Art. no. 2004765. doi: 10.1002/adfm.202004765.

[51] J.-W. Lee *et al.*, "Vapor-assisted ex-situ doping of carbon nanotube toward efficient and stable perovskite solar cells," *Nano Lett.*, vol. 19, no. 4, pp. 2223–2230, Apr. 2019, doi: 10.1021/acs.nanolett.8b04190, 10.1002/adma.200803551.

[52] B. O'Regan and M. Grätzel, "A low-cost, high efficiency solar cell based on dye sensitized colloidal TiO_2 films," *Nature*, vol. 353, no. 6346. pp. 737–740, 1991, doi: 10.1038/353737a0.

[53] M. Grätzel, "Recent advances in sensitized mesoscopic solar cells," *Acc. Chem. Res.*, vol. 42, no. 11, pp. 1788–1798, Nov. 2009, doi: 10.1021/ar900141y.

[54] K. Sharma, V. Sharma, and S. S. Sharma, "Dye-sensitized solar cells: Fundamentals and current status," *Nanoscale Res. Lett.*, vol. 13, no. 1, Nov. 2018, Art. no. 381, doi: 10.1186/s11671-018-2760-6.

[55] H. Zhu, J. Wei, K. Wang, and D. Wu, "Applications of carbon materials in photovoltaic solar cells," *Sol. Energy Mater. Sol. Cells*, vol. 93, no. 9, pp. 1461–1470, Sep. 2009, doi: 10.1016/j.solmat.2009.04.006.

[56] J. Chen, B. Li, J. Zheng, J. Zhao, and Z. Zhu, "Role of carbon nanotubes in dye-sensitized TiO_2-based solar cells," *J. Phys. Chem. C*, vol. 116, no. 28, pp. 14848–14856, Jul. 2012, doi: 10.1021/jp304845t.

[57] A. K. Chandiran, M. Abdi-Jalebi, M. K. Nazeeruddin, and M. Grätzel, "Analysis of electron transfer properties of ZnO and TiO_2 photoanodes for dye- sensitized solar cells," *ACS Nano*, vol. 8, no. 3, pp. 2261–2268, Mar. 2014, doi: 10.1021/nn405535j.

[58] Y. Chiba, A. Islam, Y. Watanabe, R. Komiya, N. Koide, and L. Han, "Dye-sensitized solar cells with conversion efficiency of 11.1%," *J. Appl. Phys.*, vol. 45, no. 7L, Jun. 2006, Art. no. L638, doi: 10.1143/JJAP.45.L638.

[59] A. Prasetio, A. Subagio, A. Purwanto, and H. Widiyandari, "Dye-sensitized solar cell based carbon nanotube as counter electrode," *AIP Conf. Proc.*, vol. 1710, no. 1, Feb. 2016, Art. no. 030054, doi: 10.1063/1.4941520.

[60] M. Yilmaz, S.-H. Hsu, S. Raina, M. Howell, W. P. Kang, and J.-H. Huang, "Dye- sensitized solar cells using carbon nanotube- based counter electrodes in planar and micro- array patterned configurations," *J. Renewable Sustain. Energy*, vol. 10, no. 6, Nov. 2018, Art. no. 063501, doi: 10.1063/1.5049482.

[61] K. Yoshikawa *et al.*, "Exceeding conversion efficiency of 26% by heterojunction interdigitated back contact solar cell with thin film Si technology," *Sol. Energy Mater. Sol. Cells*, vol. 173, pp. 37–42, Dec. 2017, doi: 10.1016/j.solmat.2017.06.024.

[62] Helmholtz-Zentrum Berlin fur Materialien und Energie, "Record: efficiency of perovskite tandem solar cell jumps to 29.15 percent," Jun. 2021. Accessed: Oct. 15, 2021. [Online]. Available: https://www.helmholtz-berlin.de/pubbin/news_seite?nid=21020;sprac he = en;seitenid =

[63] L. C. Andreani, A. Bozzola, P. Kowalczewski, M. Liscidini, and L. Redorici, "Silicon solar cells: Toward the efficiency limits," *Adv. Phys., X*, vol. 4, no. 1, Jan. 2019, Art. no. 1548305, doi: 10.1080/23746149.2018.1548305.

[64] J. Carbeck, "Solar power has big limitations. This wonder material could change that," Jun. 2021. Accessed: Oct. 15, 2021. [Online]. Available: https://www.weforum.org/agenda/2016/06/perovskite-solar-cells/

[65] C. Cui and D. Mattioli, "No silver lining left for users of the metal," *Wall Street J. (Eastern ed.)*, Apr. 2011. Accessed: Oct. 15, 2021. [Online]. Available: https://www.wsj.com/

[66] D. D. Smith, P. Cousins, S. Westerberg, R. D. Jesus-Tabajonda, G. Aniero, and Y.-C. Shen, "Toward the practical limits of silicon solar cells," *IEEE J. Photovolt.*, vol. 4, no. 6, pp. 1465–1469, Nov. 2014, doi: 10.1109/JPHOTOV.2014.2350695.

[67] J. Wei *et al.*, "Double-walled carbon nanotube solar cells," *Nano Lett.*, vol. 7, no. 8, pp. 2317–2321, Aug. 2007, doi: 10.1021/nl070961c.

[68] F. Wang *et al.*, "Considerably improved photovoltaic performance of carbon nanotube-based solar cells using metal oxide layers," *Nat. Commun.*, vol. 6, Feb. 2015, Art. no. 6305, doi: 10.1038/ncomms7305.

[69] M. Kuwahara, Y. Kim, and R. Azumi, "Measurement of the optical properties of a transparent, conductive carbon nanotube film using spectroscopic ellipsometry," *Jpn. J. Appl. Phys.*, vol. 54, no. 7, Jun. 2015, Art. no. 078001, doi: 10.7567/JJAP.54.078001.

[70] R. C. Neville, "Chapter V - basic theoretical performance," in *Solar Energy Conversion*, R. C. Neville, Ed., 2nd ed. Amsterdam, The Netherlands: Elsevier Science, 1995, pp. 155–195.

[71] J. Li *et al.*, "Direct identification of metallic and semiconducting single-walled carbon nanotubes in scanning electron microscopy," *Nano Lett.*, vol. 12, no. 8, pp. 4095–4101, Aug. 2012, doi: 10.1021/nl301561f.

[72] L. Wieland *et al.*, "Carbon nanotubes for photovoltaics: From lab to industry," *Adv. Energy Mater.*, vol. 11, 2021, Art. no. 2002880. [Online]. Available: https://doi.org/10.1002/aenm.202002880

[73] X.-G. Hu, P.-X. Hou, C. Liu, F. Zhang, G. Liu, and H.-M. Cheng, "Small-bundle single-wall carbon nanotubes for high-efficiency silicon heterojunction solar cells," *Nano Energy*, vol. 50, pp. 521–527, Aug. 2018, doi: 10.1016/j.nanoen.2018.06.004.

[74] K. Cui *et al.*, "Air-stable high-efficiency solar cells with dry-transferred single-walled carbon nanotube films," *J. Mater. Chem. A Mater. Energy Sustain.*, vol. 2, no. 29, pp. 11311–11318, Jul. 2014, doi: 10.1039/C4TA01353K.

[75] F. Wang *et al.*, "Fabrication of single- walled carbon nanotube/Si heterojunction solar cells with high photovoltaic performance," *ACS Photon.*, vol. 1, no. 4, pp. 360–364, Apr. 2014, doi: 10.1021/ph400133k.

[76] D. Kozawa, K. Hiraoka, Y. Miyauchi, S. Mouri, and K. Matsuda, "Analysis of the photovoltaic properties of single-walled carbon nanotube/silicon heterojunction solar cells," *Appl. Phys. Exp.*, vol. 5, no. 4, Apr. 2012, Art. no. 042304, doi: 10.1143/APEX.5.042304.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[77] L. Yu, D. D. Tune, C. J. Shearer, and J. G. Shapter, "Implementation of antireflection layers for improved efficiency of carbon nanotube–silicon heterojunction solar cells," *Sol. Energy*, vol. 118, pp. 592–599, Aug. 2015, doi: 10.1016/j.solener.2015.06.014.

[78] X. Li, Y. Jung, K. Sakimoto, T.-H. Goh, M. A. Reed, and A. D. Taylor, "Improved efficiency of smooth and aligned single walled carbon nanotube/silicon hybrid solar cells," *Energy Environ. Sci.*, vol. 6, no. 3, pp. 879–887, Feb. 2013, doi: 10.1039/C2EE23716D.

[79] R. Xie, N. Ishijima, H. Sugime, and S. Noda, "Enhancing the photovoltaic performance of hybrid heterojunction solar cells by passivation of silicon surface via a simple 1- min annealing process," *Sci. Rep.*, vol. 9, no. 1, Aug. 2019, Art. no. 12051, doi: 10.1038/s41598- 019-48504-7.

[80] J. Chen *et al.*, "A polymer/carbon-nanotube ink as a boron-dopant/inorganic-passivation free carrier selective contact for silicon solar cells with over 21% efficiency," *Adv. Funct. Mater.*, vol. 30, Jul. 2020, Art. no. 2004476, doi: 10.1002/adfm.202004476.

[81] M. Zendehdel, N. Y. Nia, and M. Yaghoubinia, "Emerging thin film solar panels," in *Reliability and Ecological Aspects of Photovoltaic Modules*, A. Gok, Ed. Rijeka, Croatia: IntechOpen, 2020, doi: 10.5772/intechopen.88733.

[82] M. Stuckelberger, R. Biron, N. Wyrsch, F.-J. Haug, and C. Ballif, "Review: Progress in solar cells from hydrogenated amorphous silicon," *Renewable Sustain. Energy Rev.*, vol. 76, pp. 1497–1523, Sep. 2017, doi: 10.1016/j.rser.2016.11.190.

[83] D. L. Staebler and C. R. Wronski, "Reversible conductivity changes in discharge-produced amorphous Si," *Appl. Phys. Lett.*, vol. 31, no. 4, pp. 292–294, Aug. 1977, doi: 10.1063/1.89674.

[84] S. Del Gobbo *et al.*, "Carbon nanotube semitransparent electrodes for amorphous silicon based photovoltaic devices," *Appl. Phys. Lett.*, vol. 98, no. 18, May 2011, Art. no. 183113, doi: 10.1063/1.3588352.

[85] P. M. Rajanna *et al.*, "Rational design of highly efficient flexible and transparent p- type composite electrode based on single-walled carbon nanotubes," *Nano Energy*, vol. 67, Jan. 2020, Art. no. 104183, doi: 10.1016/j.nanoen.2019.104183.

[86] P. M. Rajanna *et al.*, "Enhanced efficiency of hybrid amorphous silicon solar cells based on single-walled carbon nanotubes and polymer composite thin film," *Nanotechnology*, vol. 29, no. 10, Mar. 2018, Art. no. 105404, doi: 10.1088/1361-6528/aaa647.

[87] M. P. Ramuz *et al.*, "Evaluation of solution- processable carbon-based electrodes for all- carbon solar cells," *ACS Nano*, vol. 6, no. 11, pp. 10384–10395, Nov. 2012, doi: 10.1021/nn304410w.

[88] H. Zhou *et al.*, "Carbon nanotube arrays for optical design of amorphous silicon solar cells," *Int. J. Mater. Form.*, vol. 1, no. 2, Jun. 2008, Art. no. 113, doi: 10.1007/s12289-008- 0368-6.

[89] "Cadmium Telluride," Jun. 2021, [Online]. Available: https://www.energy.gov/eere/solar/cadmium-telluride.

[90] R. S. Hall, D. Lamb, and S. J. C. Irvine, "Back contacts materials used in thin film CdTe solar cells—A review," *Energy Sci. Eng.*, vol. ese3.843, pp. 606–632, Feb. 2021, doi: 10.1002/ese3.843.

[91] B. E. McCandless and K. D. Dobson, "Processing options for CdTe thin film solar cells," *Sol. Energy*, vol. 77, no. 6, pp. 839–856, Dec. 2004, [Online]. Available: https://optics.org/news/7/2/29, doi: 10.1016/j.solener.2004.04.012.22.1%efficiency.

[92] SPIE Europe Ltd., "First solar CdTe cell hits 22.1% efficiency," Feb. 2016. Accessed: Oct. 15, 2021. [Online]. Available: https://optics.org/news/7/2/29

[93] Z. Fang, X. C. Wang, H. C. Wu, and C. Z. Zhao, "Achievements and challenges of Cds/CdTe solar cells," *Int. J. Photoenergy*, vol. 2011, pp. 123–130, Oct. 2011, doi: 10.1155/2011/297350.

[94] T. M. Barnes *et al.*, "Single-wall carbon nanotube networks as a transparent back contact in CdTe solar cells," *Appl. Phys. Lett.*, vol. 90, no. 24, Jun. 2007, Art. no. 243503, doi: 10.1063/1.2748078.

[95] A. B. Phillips *et al.*, "Wiring-up carbon single wall nanotubes to polycrystalline inorganic semiconductor thin films: Low-barrier, copper-free back contact to CdTe solar cells," *Nano Lett.*, vol. 13, no. 11, pp. 5224–5232, Nov. 2013, doi: 10.1021/nl402659c.

[96] C. Li *et al.*, "CuTe nanoparticles/carbon nanotubes as back contact for CdTe solar cells," *J. Electron. Mater.*, vol. 47, no. 2, pp. 1250–1258, Feb. 2018, doi: 10.1007/s11664- 017-5897-7.

[97] T. Okamoto, R. Hayashi, Y. Ogawa, A. Hosono, and M. Doi, "Fabrication of polycrystalline CdTe thin-film solar cells using carbon electrodes with carbon nanotubes," *Jpn. J. Appl. Phys.*, vol. 54, no. 4S, Art. no. 04DR01, Jan. 2015, doi: 10.7567/JJAP.54.04DR01.

[98] L. Bao, B. Li, H. Zhou, H. Li, and X. Liu, "Graphene oxide doped single-walled carbon nanotube thin film as back electrode for CdTe solar cell," in *Proc. 10th IEEE Int. Conf. Nano/Micro Eng. Mol. Syst.*, Apr. 2015, pp. 391–396, doi: 10.1109/NEMS.2015.7147451.

[99] Northwestern University, "Breakthrough for carbon nanotube solar cells: Twice as efficient as current models," *Sci. Daily*, Sep. 2014. Accessed: Oct. 14, 2021. [Online]. Available: https://www.sciencedaily.com/releases/2014/09/140903105841.htm

[100] J. Y. Kim *et al.*, "Efficient tandem polymer solar cells fabricated by all-solution processing," *Science*, vol. 317, no. 5835, pp. 222–225, Jul. 2007, doi: 10.1126/science.1141711.

[101] M. S. Fuhrer *et al.*, "Crossed nanotube junctions," *Science*, vol. 288, no. 5465, pp. 494–497, Apr. 2000, doi: 10.1126/science.288.5465.494.

[102] M. S. Arnold, A. A. Green, J. F. Hulvat, S. I. Stupp, and M. C. Hersam, "Sorting carbon nanotubes by electronic structure using density differentiation," *Nat. Nanotechnol.*, vol. 1, no. 1, pp. 60–65, Oct. 2006, doi: 10.1038/nnano.2006.52.

[103] J. Arranz-Andrés and W. J. Blau, "Enhanced device performance using different carbon nanotube types in polymer photovoltaic devices," *Carbon N. Y.*, vol. 46, no. 15, pp. 2067–2075, Dec. 2008, doi: 10.1016/j.carbon.2008.08.027.

[104] N. Chehata, A. Ltaief, R. Bkakri, A. Bouazizi, and E. Beyou, "Conducting polymer functionalized multi-walled carbon nanotubes nanocomposites: Optical properties and morphological characteristics," *Mater. Lett.*, vol. 121, pp. 227–230, Apr. 2014, doi: 10.1016/j.matlet.2014.01.162.

[105] Z. Luo *et al.*, "Thermal stability of functionalized carbon nanotubes studied by in-situ transmission electron microscopy," *Chem. Phys. Lett.*, vol. 513, no. 1–3, pp. 88–93, Sep. 2011. doi: 10.1016/j.cplett.2011.07.072.

[106] T. Salim *et al.*, "Semiconducting carbon nanotubes for improved efficiency and thermal stability of polymer– Fullerene solar cells," *Adv. Funct. Mater.*, vol. 26, pp. 51–65, 2016. [Online]. Available: https://doi.org/10.1002/adfm.201503256

[107] S. Ren *et al.*, "Toward efficient carbon nanotube/P3HT solar cells: Active layer morphology, electrical, and optical properties," *Nano Lett.*, vol. 11, no. 12, pp. 5316–5321, Dec. 2011. doi: 10.1021/nl202796u.

[108] I. Celik *et al.*, "Environmental impacts from photovoltaic solar cells made with single walled carbon nanotubes," *Environ. Sci. Technol.*, vol. 51, no. 8, pp. 4722–4732, Apr. 2017. doi: 10.1021/acs.est.6b06272.

[109] "Carbon nanotubes (CNT) market, " Jun. 2021, [Online]. Available: https://www.marketsandmarkets.com/Market- Reports/carbon-nanotubes-139.html

[110] C. Ayoo, "Towards energy security for the twenty-first century," in *Energy Policy*, T. Taner, Ed. Rijeka, Croatia: IntechOpen, May 2020. [Online]. Available: https://www.intechopen.com/books/energy-policy/towards-energy-security-for-the-twenty-first-century, doi: 10.5772/intechopen.90872.

[111] P. A. Owusu and S. Asumadu-Sarkodie, "A review of renewable energy sources, sustainability issues and climate change mitigation," *Cogent. Eng.*, vol. 3, no. 1, Dec. 2016, Art. no. 1167990, doi: 10.1080/23311916.2016.1167990.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 11:00:19 UTC from IEEE Xplore. Restrictions apply.

arXiv:2111.11468v1 [hep-th] 22 Nov 2021

Causality Constraint on Circuit Complexity from COSMOEFT

Sayantan Choudhury [1,2],* Arghya Mukherjee [2,3],† Nilesh Pandey [4],‡ and Abhishek Roy [5]§

[1] *Institute of Physics, Sachivalaya Marg, Bhubaneswar, Odisha - 751005, India.*
[2] *Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai - 400085, India.*
[3] *National Institute of Science Education and Research, Jatni, Bhubaneswar, Odisha - 752050, India.*
[4] *Department of Applied Physics, Delhi Technological University, Delhi-110042, India. and*
[5] *Department of Physics, Indian Institute of Technology Jodhpur,Karwar, Jodhpur - 342037, India.*

In this article, we investigate the physical implications of the causality constraint via effective sound speed $c_s(\leq 1)$ on Quantum Circuit Complexity(QCC) in the framework of Cosmological Effective Field Theory (COSMOEFT) using the two-mode squeezed quantum states. This COSMOEFT setup is constructed using the Stückelberg trick with the help of the lowest dimensional operators, which are broken under time diffeomorphism. In this setup, we consider only the contribution from two derivative terms in the background quasi de Sitter metric. Next, we compute the relevant measures of circuit complexity and their cosmological evolution for different c_s by following two different approaches, Nielsen's and Covariance matrix method. Using this setup, we also compute the Von-Neumann and Rényi entropy, which finally establishes an underlying connecting relationship between the entanglement entropy and circuit complexity. Essentially, we study the behaviour of the circuit complexity measures and entanglement entropy with respect to the scale factor and c_s and find various interesting unexplored features within the window, $0.024 \leq c_s \leq 1$, which is supported by both causality and cosmological observation. Finally, we also comment on the connection between the circuit complexity, entanglement entropy and equilibrium temperature for different c_s lying within the mentioned window.

I. Introduction

The underlying physical concept of Quantum Circuit Complexity (QCC) is an instrumental tool not only in high energy physics but also in other branches of theoretical physics [1–49] This useful tool has now become very popular recently to probe many unexplored features in quantum field theory as well as in quantum information theory. Originally this concept was proposed to provide a way out to probe the features behind the horizon of black holes by making use of the well known "*Complexity = Volume* " and "*Complexity = Action*" conjectures [50–54]. After these conjectures' success, this concept has been frequently used in the framework of QFT and AdS/CFT [55]. Using both of these conjectures and the underlying concepts, it is possible to study various unexplored features of bulk gravitational physics with the help of quantum information theory.

Very recently, using both the concept of QCC and out-of-time ordered correlation (OTOC) functions,[56–60] it is possible now to probe the unknown features of quantum chaos . Using such studies, now it has been possible to extract the crucial information about the quantum Lyapunov exponent, scrambling time, and many more essential quantities to quantify the chaos in a quantum mechanical system. Applying these concepts, numerous significant contributions to a broad class of physics problems has been carried out. Various studies have been carried out in the framework of cosmological islands [61], inflationary cosmology [62, 63], bouncing cosmology [64], primordial gravitational waves (PGW) [65], black hole gas in arbitrary dimensions [66], supersymmetric quantum field theory [67], quantum neural networks [25] and in many other important places. The underlying connecting relationship between quantum entanglement in wormholes and emergent gravitational space-time in terms of spooky action at a distance has been an active area of interest in recent times. In this context the quantum entanglement entropy is interpreted as the minimal area of an *Einstein-Rosen bridge* (ER) in the context of the physics of wormhole. On the other hand, in the classical picture the minimal area of ER grow over a longer time span. But in the dual gravitational picture the corresponding system approaches to thermal equilibrium very fast, compared to the ER growth time scale. To express this phenomena in a very simpler language further in ref. [50] the author introduced the concept of quantum circuit complexity. In the context of black hole physics, it

* Corresponding Author
sayantan.choudhury@niser.ac.in,
sayanphysicsisi@gmail.com
† arghya@niser.ac.in
‡ nilesh911999@gmail.com
§ roy.1@iitj.ac.in

was further conjectured that the growth rate of quantum circuit complexity (C) with respect to time scale (t) is equal to the product of entropy (S) and the equilibrium temperature (T) of the black hole i.e.

$$\frac{dC}{dt} = TS, \tag{1}$$

which is completely consistent with the corresponding physical picture. Here it is important to note that, in terms quantum circuits this relationship corresponds to $K \sim S$ number of qubits in SU(2^K) gates (where the Hilbert space is 2^K dimensional.), which actually turns out to be the black hole entropy in this context. Very recently in refs. [68] the authors pointed an underlying connection between quantum circuit complexity and quantum entanglement in two different contexts. From the previous studies it is observed that if in this context the quantum entanglement entropy grows linearly with the corresponding time scale then the quantum circuit complexity also follow the same behaviour. However, it would be really good to explicitly study the underlying connection between complexity growth and entropy growth separately in various physical systems, because the previously mentioned relationship only holds for black hole systems. So the obvious expectation is that the relationship will be different for other systems as this relation is not universal at all. For this reason finding the new relation or the deviation from the previous relation is itself a very important study, which we are planning to pursue in this paper explicitly with the framework of Cosmological Effective Field Theory (COSMOEFT).

Now we will focus the spotlight on Effective Field Theory (EFT), which we will use in this paper to study quantum circuit complexity and quantum entanglement within the framework of cosmology in presence of the causal parameter, effective sound speed c_s. It is important note that, the concept of EFT is very useful in many areas of theoretical physics including cosmology, condensed matter physics, particle physics and hydrodynamics. See refs. [69–99] for more details on this topic in different contexts. Originally EFT was implemented as a model-independent version of the classes of underlying physical theories which is valid up to a specified cut-off scale at high/low energies, commonly known as UV/IR cut-off scale. For high energy EFT prescriptions the upper limit of this scale is fixed at the quantum gravity cut-off scale, which is the Planck scale $M_p(\sim 10^{19}$ GeV). For low energy EFT it is till now a free parameter. Depending on which low energy phenomena is described the EFT one can able to fix this parameter accordingly in low energy theories. Though in this work we will only restrict ourself to the high energy EFT framework which will be further useful to study COSMOEFT. In EFT we consider all the operators which are allowed by the underlying symmetry. There are two possible approaches towards the development of EFT, which are given by:

1. **Top-down method:**

   In this case, the key idea is to start with a UV complete framework which have many degrees of freedom in the theory. Using this setup one can derive an EFT below the high energy cut-off of the EFT by removing unwanted fields using the path integration method. For more details on this see refs. [80, 82, 95].

   To implement this idea let us consider a very simplest toy model of interacting scalar fields which is minimally coupled to the classical gravitational background via the space time metric and this setup is in principle UV complete. In this simplest version of the toy model let us consider a light scalar field ϕ which has a mass $m_\phi < M_p$ and many heavy scalar fields $\Phi_i \forall i = 1, 2, \cdots, N$ with mass $M_{\Phi_i} > M_p$. In principle one can formulate this setup in arbitrary $d+1$ dimensional gravitational space-time. But for simplicity and particularly since we are interested in 4 dimensional COSMOEFT setup we have restricted our discussion only in $3+1$ dimensional gravitational space-time. In this context of discussion, the light field can be easily probed by the EFT framework as its mass is below the Planckian UV cut-off and for this reason this field belongs to the visible sector of the theory in principle. On the other hand, the heavy scalar fields cannot be probed by using the EFT framework which belongs to the hidden sector of the theory. So to construct the EFT below the UV Planckian cut-off scale one needs to path integrate over all the hidden sector heavy fields. In principle this path integration can be performed in Lorentzian signature. But it turns out that technically it is very complicated to perform the path integral in that case. This is because finding out the proper saddle points is very very difficult in Lorentzian signature. So the next best possible way to do this path integration over hidden sector heavy scalar fields is in Euclidean signature which turn out be very very helpful in terms of finding proper saddles for the path integration. In this case, the starting point of the interacting toy model is described by the following representative action [80, 82, 95]:

$$\textbf{Toy model}: \quad S[\phi, \Phi_i, g_{\mu\nu}] = \int d^4x\sqrt{-g}\left[\frac{M_p^2}{2}R + \mathcal{L}_{\rm vis}[\phi] + \sum_{i=1}^{N}\mathcal{L}^{(i)}_{\rm hid}[\Phi_i] + \sum_{j=1}^{N}\mathcal{L}^{(j)}_{\rm int}[\phi, \Phi_j]\right], \tag{2}$$

where $g_{\mu\nu}$ represents the background gravitational metric in $3+1$ dimensional space-time, $\mathcal{L}_{\rm vis}[\phi]$ represents the Lagrangian density which describing the visible sector of the toy model having only light scalar field ϕ, $\mathcal{L}^{(i)}_{\rm hid}[\Phi_i]\ \forall i = 1, 2, \cdots, N$ represents the Lagrangian density which describing the hidden sector of the toy model having only heavy scalar fields $\Phi_i\ \forall i = 1, 2, \cdots, N$, and $\mathcal{L}^{(j)}_{\rm int}[\phi, \Phi_j]\ \forall j = 1, 2, \cdots, N$ is the Lagrangian density representing the interaction between light and heavy scalar fields. In principle any type of interaction can appear which are allowed by the underlying symmetry in the present context of discussion.

Now the next job is to perform the path integral in Euclidean signature using appropriate saddles on the heavy fields appearing in the UV complete toy model described in Eqn (2). Consequently, we left with only a EFT theory of the light fields whose corresponding partition function is given by the following expression:

$$\textbf{Euclidean partition function}: \quad Z_{\rm EFT}\left[\phi, g_{\mu\nu}\right] = \prod_{j=1}^{N} \int \left[\mathcal{D}\Phi_j\right] e^{-S_{\rm E}[\phi, \Phi_j, g_{\mu\nu}]} = \exp\left(-S^{E}_{\rm EFT}\left[\phi, g_{\mu\nu}\right]\right), \tag{3}$$

where in the above expression $S_{\rm E}[\phi, \Phi_j, g_{\mu\nu}]$ is the Euclidean version of the Lorentzian action $S[\phi, \Phi_j, g_{\mu\nu}]$ which can be obtained by replacing $S[\phi, \Phi_j, g_{\mu\nu}] \to iS_{\rm E}[\phi, \Phi_j, g_{\mu\nu}]$. After performing the path integration in Euclidean signature over the heavy hidden sector fields one can further compute the effective action or commonly known as the EFT action $S^{E}_{EFT}\left[\phi, g_{\mu\nu}\right]$ which is directly related with the partition function. To avoid any further confusion here it is important to note that this is semi-classical treatment where we treat the background gravitational space-time in classical footing and consider the scalar fields belong to both hidden and visible sectors in quantum footing. Also it is important to mention here that, after performing the path integration in the Euclidean signature and obtaining the expression for the Euclidean effective action one needs to go back in Lorentzian signature again, which will be useful for the further computation. For this we need to transform $S^{E}_{EFT}\left[\phi, g_{\mu\nu}\right] \to -iS_{\rm EFT}[\phi, g_{\mu\nu}]$. After performing this, at the end, the representative EFT action for the visible sector light field is described by following equation [80, 82, 95]:

$$\textbf{EFT action}: \quad S_{EFT}\left[\phi, g_{\mu\nu}\right] = \int d^4x\sqrt{-g}\left[\frac{M_p^2}{2}R + \mathcal{L}_{\rm vis}[\phi] + \sum_{p=1}^{N}\mathcal{C}^{(p)}(g)\frac{\mathcal{O}^{(p)}[\phi]}{M^{\Delta-4}}\right]. \tag{4}$$

Here the first two terms remains the same and only the effect can be observed in the last terms which is written in terms of the EFT operators appearing as an outcome of the previously mentioned path integration procedure. The coupling strength of these operators are identified as $\mathcal{C}^{(p)}(g)\ \forall p = 1, 2, \cdots, N$ which is in principle dimensionless and depend on the scale g of the UV complete toy theory. In technical language this coefficients are identified as Wilson coefficients and the associated operators are Wilson operators. Here the EFT operators are given by $\mathcal{O}^{(p)}[\phi]\ \forall p = 1, 2, \cdots, N$ where Δ represents the mass dimensional local EFT operators which are suppressed by $M^{\Delta-4}$ scale. It is important mention here that, the mathematical structure of the EFT Wilson operators will fully depend on the starting point where we have provided the toy model of UV complete theory. If anyone provides a different toy model for this purpose then the final mathematical form of these EFT operators will be different. However, the overall form of the representative EFT action in eqn 4 will not change. To check the validity of this statement one can try with different toy models of UV complete theories at the starting point. In each cases one needs to find the form of the EFT action which will support the mentioned comment.

2. **Bottom-up method:**

In this case, the key idea is to start with a

model-independent effective action which is valid below the UV Planckian cut-off scale and structure of the EFT action is restricted by symmetry. Using this setup the next goal is to find out a UV complete theory allowed by the symmetries [80, 82]. Consequently it is possible to determine the Wilson coefficients in terms of the UV complete theory. In this paper we have used this procedure. Though we have not find the exact form of the UV complete theory from the Wilson coefficients but this approach helps us to provide the causality constraint in terms of the effective sound speed parameter $c_s = \sqrt{\dot{p}/\dot{\rho}} = \sqrt{\partial p/\partial \rho}$ for COSMOEFT particularly in the parameter region where $c_s = 1$ or $c_s < 1$. Here the $\dot{}$ corresponds to the derivative with respect to the physical time which is explicitly appearing in the spatially flat FLRW space-time metric having the quasi de Sitter solution, which is given by:

$$ds^2 = -dt^2 + a^2(t)d\mathbf{x}^2, \tag{5}$$

where we have followed the $(-,+,+,+)$ metric signature along with velocity of light $c = 1$ in natural units. Here $a(t) = \exp(Ht)$ is the solution of the scale factor for quasi de Sitter background and here the Hubble parameter H is not exactly constant. Here in COSMOEFT setup p and ρ correspond to the effective pressure and density for an effective fluid, which can be very easily computed from the (ij) and (00) components of the energy momentum stress tensor, provided the explicit form of the effective action of the matter sector or EFT operators which mimics the same role are known. Now it is noted that in this context, causality breaks for the sound speed $c_s > 1$, which we will not consider in this paper. In literature it is usually identified as a-causal theories. Additionally it is important to note that though we will look into the region where the effective sound speed $c_s < 1$, but we will exclude the value $c_s = 0$ in this paper because this corresponds to ghost type of scalar field theories within the framework of EFT which is studied in the refs. [100, 101] in other contexts. Also this value is redundant from cosmological observation using Planck data [102], which restricts the lower bound on sound speed at $c_s = 0.024$. Now this means that in terms of EFT theories the effective sound speed will lie within the window, $0.024 \leq c_s \leq 1$, which will be supported by both cosmological observation and causality. But when the sound speed parameter $c_s \leq 1$ what it physically correspond to in terms of EFT theories? When $c_s = 1$ it describes a single scalar field slow roll model in COSMOEFT setup. On the other hand, for $c_s < 1$ it describes a wide class of non-canonical scalar field models within the framework of COSMOEFT. Fortunately both of these possibilities can be probed by using the constraint $c_s \leq 1$ within the present EFT setup.

In this paper, our prime objective is to compute the expressions for the quantum circuit complexity and entanglement entropy using the squeezed quantized modes appearing from the scalar part of the cosmological metric perturbation, which is commonly identified as the comoving curvature perturbation. For this computation we use unitary gauge which is useful to apply the Stückelberg trick [103, 104] which further generates the scalar component of the cosmological perturbation or the Goldstone modes by breaking the time diffeomorphism symmetry. Also it is important to note that, in this paper we have restricted the quantum initial state as the well known Bunch Davies or Euclidean state. Here the Stückelberg trick mimics the role of spontaneous symmetry breaking which is appearing in $SU(N)$ non-abelian gauge theory. In this formalism it appears that the Goldstone mode is completely destroyed by the quasi de Sitter background metric. To formulate this COSMOEFT setup the following prime components are necessarily required:

- First of all, to construct the EFT action polynomial powers of the temporal perturbed component of the metric (δg^{00}) is required which is identified as

$$\delta g^{00} = \left(g^{00} + 1\right). \tag{6}$$

  Here g^{00} represents the time component of the background metric.

- Further, to construct the EFT action polynomial powers of the temporal perturbed component of the extrinsic curvature at constant time slice $(\delta K_{\mu\nu})$ is required which is identified as

$$\delta K_{\mu\nu} = \left(K_{\mu\nu} - a^2 H h_{\mu\nu}\right), \tag{7}$$

  where $K_{\mu\nu}$ is the extrinsic curvature at constant time slice, a is the previously mentioned scale factor in quasi de Sitter space-time, H is the corresponding Hubble parameter and $h_{\mu\nu}$ is the spin-2 transverse, traceless graviton modes (gravitational waves).

The key idea of constructing an EFT action using Stückelberg trick can be also extended in the framework of having many scalar field contexts. Though for this paper we have restricted our discussion for only one canonical as well non-canonical scalar field in COSMOEFT.

In this paper, we are going to study the cosmological evolution of the quantum circuit complexity and entanglement entropy for the previously mentioned COSMOEFT setup for the different causal values of the sound speed parameter $c_s \leq 1$. This study actually going to help us to probe both the canonical and a class of non-canonical single field models with the help of a single EFT action. Also we have extended our analysis to study the evolution of the quantum circuit complexity and entanglement entropy with respect to the varying effective sound speed parameter within the previously mentioned window, $0.024 \leq c_s \leq 1$ for a given value of

the scale factor which is fixed at the early time scale of the evolution of our universe. These combined analysis will help us to extract the various unexplored important features of the quantum system under consideration for this COSMOEFT setup. Most importantly it will help us to constrain a wide class of cosmological single scalar field models in the EFT setup, which was not really performed before this work within the present context of discussion. In terms of technical details we will actually compute two very important measures of quantum circuit complexity and entanglement entropy, which are von Neumann and Rényi entropies in a very specific situation where the interacting part of the Hamiltonian of the EFT setup is parametrized by two mode squeezed states after performing the quantization. In terms of the technical computation we found that the two important relevant measures of quantum circuit complexity can be fully parametrized in terms of two parameters in the mentioned two mode squeezed state formalism, which are identified as the squeezing amplitude and squeezing angle. On the other hand, the entanglement entropy is only parametrized via squeezing amplitude. Both of these parameters evolve in the cosmological time scale and the explicit behaviour of these parameters can be computed in the corresponding Heisenberg picture for the given quasi de Sitter background metric which we use to develop the present COSMOEFT setup. By performing rigorous calculation it will turn out that in Heisenberg quantum picture the squeezing amplitude and the angle both satisfy two coupled linear differential equation where the cosmological evolution either can be explained through the time scale or in terms of the scale factor which is explicitly appearing in the expression for the quasi de Sitter metric. These coupled differential equations can be numerically solvable using the preferred set of given initial condition on the squeezing amplitude and the angle. For $c_s = 1$ case this mentioned procedure will be very simple and straightforward and it is not at all needed to provide any initial condition on the effective sound speed c_s to solve these equations numerically. In this case by solving these equation one can able to find implications for single scalar field models with the help of EFT. On the other hand, when $c_s < 1$ we have non-trivial contributions of the effective sound speed parameter c_s which will appear in the above mentioned coupled differential equations of squeezing amplitude and the angle. In this case one needs to provide an initial condition and a specific form of the sound speed parameter which is expressed in terms of the parameters appearing in the COSMOEFT action. This procedure is a bit complicated and involved compared to the previously mentioned case and give rise to significant implication to constraint a large class of non-canonical scalar models with the help of EFT approach. We have done a bit more in this paper to know in between quantum circuit complexity and entanglement entropy (1) which one capturing more information in quantum information theoretic sense? (2) which one gives best measure to describe the the underlying COSMOEFT setup? To justify to the proper answer to these questions we have further tried to see that how quantum circuit complexity and entanglement entropy are related to each other within the present context of discussion. We have done a parametric analysis of the rate of change of the quantum circuit complexity with respect to the scale factor and the entanglement entropy. In this analysis we have treated the effective sound speed c_s as a parameter to explicitly observe the outcome in the causal region $c_s \leq 1$ of the EFT. Further, this analysis is going to helps us to comment on the equilibrium thermodynamic features of the system under consideration and how much the results will be deviated from the previous results obtained from evaporating black holes. We have already mentioned before that for evaporating black hole context one can easily write down an analytical connecting relationship between the growth rate of the quantum circuit complexity and entanglement entropy. But it would be really interesting to know what would be the possible connection between these two quantities within the context of COSMOEFT. For this reason we have presented a detailed discussion to study this issue in the later half of this paper.

The key highlighting points of this paper are appended below point-wise:

- In this paper, the prime focus is to study the physical implication of the causality on quantum circuit complexity and entanglement entropy within the context of COSMOEFT.
- The COSMOEFT setup is constructed by following the model independent bottom up approach. For this computation we use Stückelberg trick which further generates the scalar component of the cosmological perturbation or the Goldstone modes by breaking the time diffeomorphism symmetry. Here the Stückelberg trick mimics the role of spontaneous symmetry breaking which is appearing in $SU(N)$ non-abelian gauge theory.
- For the different values of the sound speed parameter c_s, within the preferred range $0.024 \leq c_s \leq 1$, features of the quantum circuit complexity is studied using two well known methods, which are Covariance matrix method and Nielsen's wave function method. We have studied the behaviour (1) with respect to the scale factor in quasi de Sitter space and (2) with respect to the sound speed c_s.
- Features of the entanglement (von Neumann) entropy has been analysed (1) with respect to the scale factor in quasi de Sitter space and (2) with respect to the sound speed c_s.
- An underlying connection between the rate of change of circuit complexity and entanglement entropy has been established by treating the effective

sound speed c_s as a parameter within this COSMOEFT setup. Also we comment on the equilibrium thermodynamic behaviour of the system under consideration.

The organization of the paper is as follows:

- We start our discussion by providing an overview of the model independent bottom up approach of the generic COSMOEFT framework in section II.
- Further, in section III we have discussed about Stückelberg trick which helps further to construct an EFT in terms of Goldstone modes. This will further helps to develop a consistent COSMOEFT setup in this context.
- In section IV, we provide the analytical asymptotic solution for the mode functions generated from the scalar perturbation, which mimics the role of Goldstone modes in COSMOEFT. To fix the solution we have chosen Bunch Davies or Euclidean vacuum state as the initial choice of the quantum vacuum state.
- In section V, we further provide the details of the quantization of the Hamiltonian of the COSMOEFT in terms of two mode squeezed states. In this formalism we have shown that the interacting part of the quantized Hamiltonian is parametrized in terms of squeezing amplitude and squeezing angle, where each of them evolve either with respect to the evolutionary time scale or the scale factor.
- In section VI, we provide a short review on the subject of quantum circuit complexity, which we believe will be very helpful for the better understanding purpose of general readers. Here we provide sufficient details regarding the fundamental definition, appropriate measures, various properties along with the well known computational methods of the circuit complexity developed by Nielsen and collaborators.
- In section VII, we discuss about the computational details of the quantum circuit complexity using the previously mentioned two-mode squeezed state formalism within the framework of COSMOEFT. Here we have shown that the quantum circuit complexity can be quantified using the well known Covariance Matrix method and Nielsen's method of wave functions.
- In section VIII, we provide the computational details of the entanglement entropy of two-mode squeezed states. In this section we compute Rényi-entropy, and von-Neumann entropy in this connection in terms of squeezed amplitude parameter.
- In section IX, we discuss an underlying connection between quantum circuit complexity and entanglement entropy which will be helpful to know that which one is the best measure within the framework of COSMOEFT.
- In section X, we mention and analyse all the numerical outcomes from the features of quantum circuit complexity and entanglement entropy within the framework of COSMOEFT . In this section we have studied the behaviour of quantum circuit complexity and entanglement entropy (1) with respect to the scale factor in quasi de Sitter space for the fixed values of the sound speed c_s and (2) with respect to the sound speed c_s for the fixed value of the scale factor at the early time scale of the evolution of our universe. Also we have studied the parametric behaviour of the rate of change of the quantum circuit complexity with respect to the scale factor with the entanglement entropy in this context by considering the effective sound speed c_s as a parameter within the region $0.024 \leq c_s \leq 1$, which is both supported by cosmological observation and causality. This combined analysis helps us to extract various known physical implications of these quantum information theoretic measures for $c_s \leq 1$, using which one can further able to constraint both single scalar field models as well as a class of non-canonical scalar field models within the framework of COSMOEFT. Additionally, this analysis will going to help us to determine a connecting relationship between all of these quantum information theoretic measures. Most importantly, it will determine which is the best measure in this context.
- In section XI, we provide a comparison between the outcomes obtained from the computation of quantum circuit complexity using Covariance Matrix method and Nielsen's method of wave functions. This comparison will help us to decide from which method we are getting full information from the underlying COSMOEFT setup.
- Finally in section XII, we conclude our all findings of this paper along with some interesting future prospects of our present work.

II. Construction of COSMOEFT

In this section, our prime objective is to give a review of constructing the most generic COSMOEFT action written in the background of quasi de Sitter spacetime. The prime objective to use this specific framework in the present context is to generate scalar modes of cosmological perturbations which will be helpful for our further study. But in our EFT setup there is matter degrees of freedom whose fluctuation will generate such perturbations. The breakdown of time diffeomorphism symmetry will generate Goldstone modes from the various EFT operators which is nothing but the gravitational

fluctuations. Fortunately these Goldstone modes mimics the role of spontaneously symmetry breaking mechanism as appearing in the context of $SU(N)$ gauge theory. In our discussion such Goldstone mode serve the purpose of scalar modes as appearing in cosmological perturbation theory. In this setup, a scalar field transformed as scalar under the following space-time diffeomorphism symmetry within the framework of General Theory of Relativity (GTR):

$$\textbf{Space-time diffeomorphism symmetry:} \quad x^\nu \Longrightarrow \tilde{x}^\nu = x^\nu + \xi^\nu(t,\mathbf{x}) \quad \forall\ \nu = 0,1,2,3\ . \tag{8}$$

Here $\xi^\nu(t,\mathbf{x})$ represents the space-time diffeomorphism parameter in this context. Now if we think of a parallel situation where we are interested in scalar fields, then the perturbation on the field $\delta\phi$ transform (1) as a scalar under only the spatial part of the space-time diffeomorphism symmetry and (2) non-linearly under only the temporal part of the space-time diffeomorphism symmetry. Here separately these spatial and temporal part of the diffeomorphic transformations can be written as:

$$\textbf{Spatial diffeomorphism symmetry}: \quad t \Longrightarrow \tilde{t} = t,\ x^i \Longrightarrow \tilde{x}^i = x^i + \xi^i(t,\mathbf{x}) \quad \forall\ i = 1,2,3 \tag{9}$$

$$\longrightarrow \textbf{Outcome}: \quad \delta\phi \Longrightarrow \widetilde{\delta\phi} = \delta\phi, \tag{10}$$

$$\textbf{Time diffeomorphism symmetry}: \quad t \Longrightarrow \tilde{t} = t + \xi^0(t,\mathbf{x}),\ x^i \Longrightarrow \tilde{x}^i = x^i \quad \forall\ i = 1,2,3 \tag{11}$$

$$\longrightarrow \textbf{Outcome}: \quad \delta\phi \Longrightarrow \widetilde{\delta\phi} = \delta\phi + \dot{\phi}_0(t)\xi^0(t,\mathbf{x}). \tag{12}$$

In the above mentioned expressions $\xi^0(t,\mathbf{x})$ and $\xi^i(t,\mathbf{x})\forall i = 1,2,3$ represent the spatial and temporal diffeomorphism parameters. Here we choose the gravitational gauge such that, $\phi(t,\mathbf{x}) = \phi_0(t)$, where $\phi_0(t)$ is the background time dependent scalar field in homogeneous isotropic FLRW cosmological space-time. This further demands that in this gauge choice:

$$\textbf{Unitary gauge fixing condition}: \quad \delta\phi(t,\mathbf{x}) = 0\ , \tag{13}$$

In the framework of cosmology this is the unitary gravitational gauge fixing criteria which implies that all information of the field content are preserved in the background metric. The similar type of facts are also appearing in the context of $SU(N)$ gauge theory where the Goldstone modes transform non-linearly under spontaneous symmetry breaking. In this context similarly the temporal diffeomorphism symmetry transformation is capturing this non-linear effect. Now to construct the COSMOEFT we need to follow few steps:

1. First of all the EFT operators has to be function of the gravitational space-time metric $g_{\mu\nu}$.
2. Now using the derivatives of the space-time metric $g_{\mu\nu}$ one can compute further Riemann tensor $R_{\mu\nu\alpha\beta}$, Ricci tensor $R_{\mu\nu}$ and Ricci scalar R which are also be the components of the EFT action.
3. EFT operators has to be invariant under spatial diffeomorphism symmetry transformation. For this reason terms containing g^{00} and $\delta g^{00} = \left(g^{00}+1\right)$ are allowed in EFT action.
4. Additionally one can consider the extrinsic curvature $K_{\mu\nu}$ and its fluctuation $\delta K_{\mu\nu}$, which is explicitly defined before.

Consequently, the COSMOEFT action can be expressed as:

$$S = \int d^4x\sqrt{-g}\Bigg[\underbrace{\frac{M_p^2}{2}R}_{\textbf{Einstein Hilbert term}} + \underbrace{M_p^2\dot{H}g^{00} - M_p^2\left(3H^2+\dot{H}\right)}_{\textbf{Mimics the role of kinetic and potential term for a scalar field}} + \underbrace{\sum_{n=2}^{\infty}\frac{M_n^4(t)}{n!}(\delta g^{00})^n - \sum_{q=0}^{\infty}\frac{\bar{M}_1^{3-q}(t)}{(q+2)!}\delta g^{00}\left(\delta K_\mu^\mu\right)^{q+1} - \sum_{m=0}^{\infty}\frac{\bar{M}_2^{2-m}(t)}{(m+2)!}\left(\delta K_\mu^\mu\right)^{m+2} + \cdots}_{\textbf{Fluctuation part which will directly contribute in finding scalar perturbation}}\Bigg]\ . \tag{14}$$

where the dots corresponds to higher-order fluctuations on which we are not interested at present as we have restricted our analysis for operators which contain only the two derivatives in the space time metric. Here $\delta K_{\mu\nu}$ and δg^{00} is already defined before.

In this context extrinsic curvature $K_{\mu\nu}$, unit normal n_μ and induced metric $h_{\mu\nu}$ is defined as:

$$\textbf{Extrinsic curvature}: \quad K_{\mu\nu} = h^\sigma_\mu \nabla_\sigma n_\nu$$
$$= \left[\frac{\delta^0_\mu \partial_\nu g^{00} + \delta^0_\nu \partial_\mu g^{00}}{2(-g^{00})^{3/2}} + \frac{\delta^0_\mu \delta^0_\nu g^{0\sigma}\partial_\sigma g^{00}}{2(-g^{00})^{5/2}} - \frac{g^{0\rho}\left(\partial_\mu g_{\rho\nu} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu}\right)}{2(-g^{00})^{1/2}}\right], \tag{15}$$

$$\textbf{Induced metric}: \quad h_{\mu\nu} = g_{\mu\nu} + n_\mu n_\nu, \tag{16}$$

$$\textbf{Unit normal}: \quad n_\mu = \frac{\partial_\mu t}{\sqrt{-g^{\mu\nu}\partial_\mu t \partial_\nu t}} = \frac{\delta^0_\mu}{\sqrt{-g^{00}}}. \tag{17}$$

Now as we are interested to compute the two- and three-point correlation function, we have restricted to the following truncated EFT action :

$$\textbf{Required COSMOEFT action}: \quad S = \int d^4x\sqrt{-g}\left[\frac{M_p^2}{2}R + M_p^2\dot{H}g^{00} - M_p^2\left(3H^2+\dot{H}\right) + \frac{M_2^4(t)}{2!}\left(g^{00}+1\right)^2 + \cdots\right]. \tag{18}$$

where the dotted terms are not interesting contributions for our paper as it will not appear in the expression for effective sound speed c_s, which will going to be more clear in the next section. But for other purposes the other higher order fluctuating contributions are very useful. For example, to find three point and four point correlations from scalar perturbations these higher order terms are necessarily needed. For more details on these aspects see refs. [][**Add references**]. Last but not the least the corresponding gravitational metric components of describing spatially flat FLRW space-time having quasi de Sitter solution is explicitly written in the introduction, which is used here as a background gravitational space-time. In this case, the factor $\sqrt{-g} = a^3(t) = \exp(3Ht)$ as appearing in the required COSMOEFT action in the present context of discussion. Here it is important to note that, the Hubble parameter H is slowly varying with time and not a constant for quasi de Sitter space. For this reason in the later part of the discussion of the paper we have introduced a quantity, $\epsilon = -\dot{H}/H^2$, which basically measures the deviation from the constant behaviour of Hubble parameter which appears in the context of de Sitter space. This parameter is known as the slow roll parameter and our analysis is valid in the region $\epsilon \ll 1$ where all the approximations of the cosmological perturbation theory works well within the framework of COSMOEFT.

III. COSMOEFT and Goldstone Boson

In the section we elaborate the Stückelberg Trick for broken time diffeomorphism which will generate the Goldstone modes. We also discuss how the Goldstone action from the EFT can be constructed which will mimic the role of scalar modes for cosmological perturbation. For better understanding we divide the discussion in two different subsections which deal with both of these issues in detail.

A. Stückelberg Trick

Let us start with the time diffeomorphism symmetry as stated in eqn. (11) under which Goldstone mode ($\pi(t,\mathbf{x})$) transforms as:

$$\pi(t,\mathbf{x}) \to \tilde{\pi}(t,\mathbf{x}) = \pi(t,\mathbf{x}) - \xi^0(t,\mathbf{x}). \tag{19}$$

where $\xi^0(t,\mathbf{x})$ represents the local parameter which is appearing in eqn. (11). In this discussion this Goldstone mode mimics the role of the scalar modes in cosmological perturbation. Then the corresponding unitary gauge fixing condition is given by:

$$\textbf{Unitary gauge fixing}: \quad \pi(t,\mathbf{x}) = 0 \quad \Rightarrow \quad \tilde{\pi}(t,\mathbf{x}) = -\xi^0(t,\mathbf{x}) \ , \tag{20}$$

which will be useful further to construct the EFT action.

Now it is very useful to mention that how the compo-

nents of the space-time metric, Ricci tensor, Ricci scalar, perturbation on extrinsic curvature, time dependent coefficients and the slowly varying Hubble parameter transform under the application of broken time diffeomorphism symmetry:

1. **Transformation of space-time metric:** Under the application of broken time diffeomorphism symmetry contravariant and covariant metric transform as:

$$\textbf{Contravariant metric}: \quad g^{00} \Longrightarrow \tilde{g}^{00} = (1+\dot{\pi}(t,\mathbf{x}))^2 g^{00} + 2(1+\dot{\pi}(t,\mathbf{x}))g^{0i}\partial_i\pi(t,\mathbf{x}) + g^{ij}\partial_i\pi(t,\mathbf{x})\partial_j\pi(t,\mathbf{x}), \tag{21}$$

$$g^{0i} \Longrightarrow \tilde{g}^{0i} = (1+\dot{\pi}(t,\mathbf{x}))g^{0i} + g^{ij}\partial_j\pi(t,\mathbf{x}), \tag{22}$$

$$g^{ij} \Longrightarrow \tilde{g}^{ij} = g^{ij}. \tag{23}$$

$$\textbf{Covariant metric}: \quad g_{00} \Longrightarrow \tilde{g}_{00} = (1+\dot{\pi}(t,\mathbf{x}))^2 g_{00}, \tag{24}$$

$$g_{0i} \Longrightarrow \tilde{g}_{0i} = (1+\dot{\pi}(t,\mathbf{x}))g_{0i} + g_{00}\dot{\pi}(t,\mathbf{x})\partial_i\pi(t,\mathbf{x}), \tag{25}$$

$$g_{ij} \Longrightarrow \tilde{g}_{ij} = g_{ij} + g_{0j}\partial_i\pi(t,\mathbf{x}) + g_{i0}\partial_j\pi(t,\mathbf{x}). \tag{26}$$

2. **Transformation of Ricci scalar and Ricci tensor:** Under the application of broken time diffeomorphism symmetry Ricci scalar and the spatial component of the Ricci tensor on 3-hypersurface transform as:

$$\textbf{Ricci scalar}: \quad {}^{(3)}R \Longrightarrow {}^{(3)}\tilde{R} = {}^{(3)}R + \frac{4}{a^2}H(\partial^2\pi(t,\mathbf{x})), \tag{27}$$

$$\textbf{Spatial Ricci tensor}: \quad {}^{(3)}R_{ij} \Longrightarrow {}^{(3)}\tilde{R}_{ij} = {}^{(3)}R_{ij} + H(\partial_i\partial_j\pi(t,\mathbf{x}) + \delta_{ij}\partial^2\pi(t,\mathbf{x})). \tag{28}$$

3. **Transformation of extrinsic curvature:** Under the application of broken time diffeomorphism symmetry trace and the spatial, time and mixed component of the extrinsic curvature transform as:

$$\textbf{Trace of extrinsic curvature}: \quad \delta K \Longrightarrow \widetilde{\delta K} = \delta K - 3\pi\dot{H} - \frac{1}{a^2}(\partial^2\pi(t,\mathbf{x})), \tag{29}$$

$$\textbf{Spatial extrinsic curvature}: \quad \delta K_{ij} \Longrightarrow \widetilde{\delta K}_{ij} = \delta K_{ij} - \pi(t,\mathbf{x})\dot{H}h_{ij} - \partial_i\partial_j\pi(t,\mathbf{x}), \tag{30}$$

$$\textbf{Temporal extrinsic curvature}: \quad \delta K^0_0 \Longrightarrow \widetilde{\delta K}^0_0 = \delta K^0_0, \tag{31}$$

$$\textbf{Mixed extrinsic curvature}: \quad \delta K^0_i \Longrightarrow \widetilde{\delta K}^0_i = \delta K^0_i, \tag{32}$$

$$\textbf{Mixed extrinsic curvature}: \quad \delta K^i_0 \Longrightarrow \widetilde{\delta K}^i_0 = \delta K^i_0 + 2Hg^{ij}\partial_j\pi(t,\mathbf{x}). \tag{33}$$

4. **Transformation of time-dependent EFT coefficients:** Under the application of broken time diffeomorphism symmetry, time-dependent EFT coefficients transform after canonical normalization $\pi_c(t,\mathbf{x}) = Q^2(t)\pi(t,\mathbf{x})$ as:

$$\textbf{Time - dependent coefficients}: \quad Q(t) \Longrightarrow Q(t+\pi(t,\mathbf{x})) = \sum_{n=0}^{\infty}\frac{\pi^n(t,\mathbf{x})}{n!}\frac{d^nQ(t)}{dt^n} = \sum_{n=0}^{\infty}\underbrace{\frac{\pi_c^n(t,\mathbf{x})}{n!Q^{2n}(t)}}_{\textbf{Suppression}}\frac{d^nQ(t)}{dt^n} \approx Q(t)\ . \tag{34}$$

Here $Q(t)$ represents the time-dependent coefficients in the EFT action.

5. **Transformation of Hubble parameter:** Under the application of broken time diffeomorphism symmetry Hubble parameter transform as:

Hubble parameter :

$$H(t) \Longrightarrow H(t+\pi(t,\mathbf{x})) = \sum_{n=0}^{\infty} \frac{\pi^n}{n!} \frac{d^n H(t)}{dt^n} = \left[1 - \underbrace{\pi(t,\mathbf{x})H(t)\epsilon - \frac{\pi^2(t,\mathbf{x})H(t)}{2}\left(\dot{\epsilon} - 2\epsilon^2\right) + \cdots}_{\textbf{Correction terms}}\right] H(t) \ . \quad (35)$$

Here ϵ is the slow-roll parameter which we have defined before.

Now to construct the COSMOEFT action we need to understand the decoupling limit very clearly. In this limit the mixing contributions between the gravity and Goldstone modes can easily be neglected. To justify the validity of this statement let us start with the $-\dot{H}M_p^2 g^{00}$ operator contribution in the EFT action which will be very relevant for the further computation in this paper.

Under the application of the previously mentioned broken time diffeomorphism symmetry, this operator will transform by the following fashion:

$$-\dot{H}M_p^2 g^{00} \Longrightarrow -\tilde{\dot{H}}M_p^2 \tilde{g}^{00} \approx -\dot{H}M_p^2\left[(1+\dot{\pi}(t,\mathbf{x}))^2 g^{00} + \left(2(1+\dot{\pi}(t,\mathbf{x}))\partial_i\pi(t,\mathbf{x})g^{0i} + g^{ij}\partial_i\pi(t,\mathbf{x})\partial_j\pi(t,\mathbf{x})\right)\right] \ . \quad (36)$$

Now after perturbation the temporal component of the metric can be written as:

$$g^{00} = \bar{g}^{00} + \delta g^{00}, \quad (37)$$

where time component of the background quasi de Sitter metric is given by, $\bar{g}^{00} = -1$ and the perturbation is characterized by δg^{00}. Further using this in Equation (36) we get a kinetic contribution, $M_p^2\dot{H}\dot{\pi}^2 g^{\bar{0}0}$ and a mixing contribution, $M_p^2\dot{H}\dot{\pi}\delta g^{00}$ in this calculation. Furthermore, we use a canonical normalized metric perturbation, $\delta g_c^{00} = M_p \delta g^{00}$, one can write, the mixing contribution as, $M_p^2\dot{H}\dot{\pi}\delta g^{00} = \sqrt{\dot{H}}\dot{\pi}_c\delta g_c^{00}$. Here above the energy scale $E_{mix} = \sqrt{\dot{H}}$, one can easily neglect this mixing term in the decoupling limit. One can also consider mixing contributions $M_p^2\dot{H}\dot{\pi}^2\delta g^{00}$ and $\pi M_p^2\ddot{H}\dot{\pi}\bar{g}^{00}$, which can be rewritten after canonical normalization as, $M_p^2\dot{H}\dot{\pi}^2\delta g^{00} = \dot{\pi}_c^2\delta g_c^{00}/M_p$ and $\pi M_p^2\ddot{H}\dot{\pi}\bar{g}^{00} = \ddot{H}\pi_c\dot{\pi}_c\bar{g}^{00}/\dot{H}$. Here at the scale $E > E_{mix}$ one can neglect the contribution from $M_p^2\dot{H}\dot{\pi}\delta g^{00}$ term. At the end, in the decoupling limit Equation (36) can be further simplified as:

$$-\dot{H}M_p^2 g^{00} \Longrightarrow -\tilde{\dot{H}}M_p^2\tilde{g}^{00} \approx -\dot{H}M_p^2 g^{00}\left[\dot{\pi}^2 - \frac{1}{a^2}(\partial_i\pi)^2\right]. \quad (38)$$

The above mentioned simplified expression is very useful for the further calculation because this contribution will explicitly appear in the EFT action.

B. The Goldstone Action

In the decoupling limit the Goldstone action (S_π) is given by:

$$S_\pi = \int d^4x\ a^3\left[-M_p^2\dot{H}\left(\dot{\pi}^2 - \frac{1}{a^2}(\partial_i\pi)^2\right) + 2M_2^4\dot{\pi}^2\right]$$
$$= \int d^4x\ a^3\left(-\frac{M_p^2\dot{H}}{c_s^2}\right)\left[\dot{\pi}^2 - c_s^2\frac{1}{a^2}(\partial_i\pi)^2\right], \quad (39)$$

where we define effective sound speed $c_s$parameter as:

$$\textbf{Effective sound speed} : c_s \equiv \frac{1}{\sqrt{1 - \frac{2M_2^4}{\dot{H}M_p^2}}}. \quad (40)$$

Here if we set $M_2 = 0$ then in that case sound speed $c_s = 1$, which is true for single- scalar field canonical slow-roll models. But if $M_2 \neq 0$ then we get $c_s \neq 1$. But since we only interested in within the window, $0.024 \leq c_s \leq 1$, using this the following constraint on the parameter M_2 can be derived:

$$-867.556 \leq \frac{M_2^4}{\dot{H}M_p^2} \leq 0. \quad (41)$$

Here the linear spatial part of the metric perturbation can be written as:

$$g_{ij} = a^2(t)\left[(1+2\zeta(t,\mathbf{x}))\,\delta_{ij} + \gamma_{ij}\right] \quad \forall \quad i = 1,2,3, \quad (42)$$

where $a(t)$ is the scale factor which we have defined before in quasi de Sitter background. Here $\zeta(t,\mathbf{x})$ represent the curvature perturbation, γ_{ij} is the transverse and traceless tensor perturbation. Here under the application of broken time diffeomorphism symmetry scale factor $a(t)$ transforms as:

$$a(t) \Longrightarrow a(t-\pi(t,\mathbf{x})) \approx a(t)\left(1-H\pi(t,\mathbf{x})\right) \ . \quad (43)$$

This further implies that the curvature perturbation $\zeta(t,\mathbf{x})$ can be expressed in terms of Goldstone modes $\pi(t,\mathbf{x})$ as:

$$\zeta(t,\mathbf{x}) = -H\pi(t,\mathbf{x}). \quad (44)$$

Consequently, the Goldstone action can be expressed in terms of curvature perturbation $\zeta(t,\mathbf{x})$ variable as:

$$S_\zeta = \int d^4x\ a^3\ \left(\frac{M_p^2\epsilon}{c_s^2}\right)\left[\dot{\zeta}^2 - c_s^2\frac{1}{a^2}(\partial_i\zeta)^2\right]. \quad (45)$$

This action will be further used to study the squeezed state formalism in the present context.

IV. Mode Equation and Solution for Scalar Perturbation

Now our job is to simplify the form of the action stated in equation (45). To serve this purpose we introduce a field redefinition:

$$v(\tau,\mathbf{x}) = z\ \zeta(\tau,\mathbf{x})\ M_p \ \text{ where } \ z = \frac{a\sqrt{2\epsilon}}{c_s} \quad (46)$$

which helps us to recast the action for the curvature perturbation in terms of new variable as:

$$S = \int d^3x\ d\tau\ \left[v'^2 - c_s^2\frac{1}{a^2}(\partial_i v)^2 + \frac{1}{z}\frac{d^2z}{d\tau^2}v^2\right], \quad (47)$$

Here τ is the conformal time which can be expressed in terms of physical time t as, $\tau = \int \frac{dt}{a(t)}$. The conformal time described here is negative and lying within $-\infty < \eta < 0$. In terms of conformal time the background metric of quasi de Sitter space-time can be recast as:

$$ds^2 = a^2(\tau)\left(-d\tau^2 + d\mathbf{x}^2\right), \quad (48)$$

where the scale factor in terms of conformal time in quasi de Sitter space can be expressed as:

$$a(\tau) = -\frac{1}{H\tau}\left(1+\epsilon\right) \quad (49)$$

Now to compute the mode function for the scalar perturbations we map the whole problem in Fourier space by using the following ansatz for Fourier transformation:

$$v(\tau,\boldsymbol{x}) = \int \frac{d^3k}{(2\pi)^3}\ v_{\mathbf{k}}(\tau)\ \exp(i\mathbf{k}.\mathbf{x}) \quad (50)$$

After expressing the action in Fourier space and varying the action the equation of motion for scalar perturbation can be written as:

$$v_{\mathbf{k}}''(\tau) + \left(c_S^2k^2 - \frac{1}{z}\frac{d^2z}{d\tau^2}\right)v_{\mathbf{k}}(\tau) = 0. \quad (51)$$

In quasi de Sitter space one can write:

$$\frac{1}{z}\frac{d^2z}{d\tau^2} = \frac{\left(\nu^2-\frac{1}{4}\right)}{\tau^2} \quad (52)$$

where the parameter ν can be expressed as:

$$\nu = \left(\frac{3}{2} + 3\epsilon - \eta + \frac{s}{2}\right) \quad (53)$$

where ϵ, η and s are the slow-roll parameters which are defined as:

$$\epsilon = -\frac{\dot{H}}{H^2}, \quad \eta = 2\epsilon - \frac{\dot{\epsilon}}{2H\epsilon}, \quad s = \frac{\dot{c_s}}{Hc_s}. \quad (54)$$

The general analytical solution for $v_{\mathbf{k}}(\tau)$ can be expressed as:

$$v_{\mathbf{k}}(\tau) = \sqrt{-\tau}\left[C_1H_\nu^{(1)}\left(-kc_S\tau\right) + C_2H_\nu^{(2)}\left(-kc_S\tau\right)\right] \quad (55)$$

where $H_\nu^{(1)}\left(-kc_S\tau\right)$ and $H_\nu^{(1)}\left(-kc_S\tau\right)$ are the Hankel functions of first and second kind with order ν. Also C_1 and C_2 are the arbitrary integration constants which depends on the choice of the initial quantum vacuum. In this paper we use the well known Bunch Davies or the Euclidean vacuum state as the initial state which is fixed by $C_1 = 1, C_2 = 0$. Considering this the general solution for the scalar mode function can be simplified as:

$$v_{\mathbf{k}}(\tau) = \sqrt{-\tau}\ H_\nu^{(1)}\left(-kc_S\tau\right). \quad (56)$$

Now instead of using this full solution we need to use the asymptotic solution using which one can study the following cosmologically relevant regions:

1. **Superhorizon region:** $kc_s\tau \ll -1$,
2. **Horizon crossing:** $kc_s\tau = -1$,
3. **Subhorizon region:** $kc_s\tau \gg -1$.

At the end, capturing the information from these cosmologically relevant regions the simplified asymptotic solution of the mode function can be expressed as:

$$v_{\mathbf{k}}(\tau)=2^{\nu-\frac{3}{2}}\frac{1}{i\tau}\frac{1}{\sqrt{2}\left(kc_s\right)^{\frac{3}{2}}}(-kc_s\tau)^{\frac{3}{2}-\nu}\left|\frac{\Gamma(\nu)}{\Gamma\left(\frac{3}{2}\right)}\right|(1+ikc_s\tau)\exp\left(-i\left\{kc_s\tau+\frac{\pi}{2}\left(\nu+\frac{1}{2}\right)\right\}\right). \tag{57}$$

V. Quantized Hamiltonian and two-mode squeezed states

Using the previously derived scalar mode function from the EFT setup one can easily compute the classical Hamiltonian function which can be promoted as an operator in the Heisenberg picture. Consequently the quantized version of the Hamiltonian can be written as:

$$H(\tau)=\frac{1}{2}\int d^3\mathbf{k}\Bigg[\Omega_{\mathbf{k}}(\tau)\left(\hat{c}^{\dagger}_{\mathbf{k}}\hat{c}_{\mathbf{k}}+\hat{c}^{\dagger}_{-\mathbf{k}}\hat{c}_{-\mathbf{k}}+1\right)+i\lambda_{\mathbf{k}}(\tau)\left(e^{-2i\phi_{\mathbf{k}}(\tau)}\hat{c}_{\mathbf{k}}\hat{c}_{-\mathbf{k}}-e^{2i\phi_{\mathbf{k}}(\tau)}\hat{c}^{\dagger}_{\mathbf{k}}\hat{c}^{\dagger}_{-\mathbf{k}}\right)\Bigg], \tag{58}$$

where, $(\hat{c}^{\dagger}_{\mathbf{k}},\hat{c}^{\dagger}_{-\mathbf{k}})$ and $(\hat{c}_{\mathbf{k}},\hat{c}_{-\mathbf{k}})$ are the creation and annihilation operators for the two momentum modes having momenta $\mathbf{k}$ and $-\mathbf{k}$. In this quantized Hamiltonian first three terms are the outcome of the quantization from the free part of the EFT and rest of the contributions are appearing from the quantization of the interacting part of the EFT action. Most importantly, this interaction part in the present context parametrized in terms of two time evolving parameters, these are squeezing amplitude and squeezing angle. Particularly here the quantized part of the interaction Hamiltonian is written in terms of two mode squeezed states contributions after performing the quantization. Here we introduce two new symbols $\Omega_{\mathbf{k}}(\tau)$ and $\lambda_{\mathbf{k}}(\tau)$ which are defined as:

$$\Omega_{\mathbf{k}}(\tau):=2^{2(\nu-2)}\ k\ \left|\frac{\Gamma(\nu)}{\Gamma\left(\frac{3}{2}\right)}\right|^2\left(1+c_s^2\right), \tag{59}$$

$$\lambda_{\mathbf{k}}(\tau):=\sqrt{\left\{\frac{\Omega_{\mathbf{k}}(\tau)}{2}\left(\frac{1-c_s^2}{1+c_s^2}\right)\right\}^2+\left(\frac{z'(\tau)}{z(\tau)}\right)^2}. \tag{60}$$

and one can write:

$$\frac{z'(\tau)}{z(\tau)}=\frac{a'(\tau)}{a(\tau)}+\frac{1}{2}\frac{\epsilon'(\tau)}{\epsilon(\tau)}-\frac{c_s'}{c_s}. \tag{61}$$

As the quantized Hamiltonian is in our hand, we can now determine the expression for the unitary time evolution operator from these Goldstone modes in the COSMOEFT framework. Now as we have parametrized the interacting part of the quantized Hamiltonian in terms of two mode squeezed state it is very easy to write down the form of the unitary time evolution operator in this formalism, which is given by the following expression:

$$U(\tau,\tau_0)=\hat{\mathcal{S}}(r_{\mathbf{k}}(\tau,\tau_0),\phi_{\mathbf{k}}(\tau))\hat{\mathcal{R}}(\theta_{\mathbf{k}}(\tau)), \tag{62}$$

where the above factorization is only possible if we specifically use the singe field two mode squeezed state formalism in the present context. However, there are other possibilities using which one can able to determine the structure of the unitary time evolution operator. For more details on this issue see refs. [61, 64–66]. Here $\hat{\mathcal{R}}(\theta_{\mathbf{k}}(\tau))$ is the rotation operator which can be expressed as:

$$\hat{\mathcal{R}}(\theta_{\mathbf{k}}(\tau))=\exp\left(-i\theta_{\mathbf{k}}(\tau)\left(\hat{c}_{\mathbf{k}}\hat{c}^{\dagger}_{\mathbf{k}}+\hat{c}^{\dagger}_{-\mathbf{k}}\hat{c}_{-\mathbf{k}}\right)\right), \tag{63}$$

and $\hat{\mathcal{S}}(r_{\mathbf{k}}(\tau,\tau_0),\phi_{\mathbf{k}}(\tau))$ is the squeezing operator, which is defined as:

$$\hat{\mathcal{S}}(r_{\mathbf{k}}(\tau),\phi_{\mathbf{k}}(\tau))=\exp\left(r_{\mathbf{k}}(\tau)\left[e^{-i\phi_{\mathbf{k}}(\tau)}\hat{c}_{\mathbf{k}}\hat{c}_{-\mathbf{k}}-e^{i\phi_{\mathbf{k}}(\tau)}\hat{c}^{\dagger}_{-\mathbf{k}}\hat{c}^{\dagger}_{\mathbf{k}}\right]\right). \tag{64}$$

Here the squeezing amplitude and squeezing angle is characterized by $r_{\mathbf{k}}(\tau)$ and $\phi_{\mathbf{k}}(\tau)$, which evolve with respect to the evolutionary time scale. In the present context the rotation operator $\hat{\mathcal{R}}(\theta_{\mathbf{k}}(\tau))$ leads to an unwanted phase which is not at all important for the computation of the quantum circuit complexity and quantum entanglement entropy in the present context. For this reason we can easily ignore such unwanted contribution in the present computation. It would be good to provide few more information for the further computational purpose which will going to help us to determine the analytical expressions for the quantum circuit complexity and quantum entanglement entropy functions computed from various relevant measures. One of the prime input information is about the construction of initial quantum state which is necessarily required. In this description ground state is basically characterized by the free part of the quantized Hamiltonian and for the computation of quantum circuit complexity we fix this as the initial quantum mechanical state. Using this initial quantum state one can able to construct the two mode squeezed state in the present context which is considered to be the final target state for this computation. Now having this initial and target state one can now able to compute the expressions for the quantum circuit complexity and quantum entanglement entropy in terms of the squeezing amplitude $r_{\mathbf{k}}(\tau)$ and squeezing angle $\phi_{\mathbf{k}}(\tau)$.

Now our next job is to determine the time evolving behaviour of the squeezing amplitude $r_{\mathbf{k}}(\tau)$ and squeezing angle $\phi_{\mathbf{k}}(\tau)$ which will completely determine the quantum circuit complexity and quantum entanglement entropy in the context of COSMOEFT. This can be done by knowing the time evolutionary behaviour of the time dependent squeezed operator $\hat{\mathcal{S}}(r_{\mathbf{k}}(\tau),\phi_{\mathbf{k}}(\tau))$. Technically this is described by the Schrödinger equation within the present framework. This will finally give rise to the following two coupled differential equations of squeezing

amplitude $r_{\mathbf{k}}(\tau)$ and squeezing angle $\phi_{\mathbf{k}}(\tau)$:

$$\frac{dr_{\mathbf{k}}(\tau)}{d\tau} = \lambda_{\mathbf{k}}(\tau)\cos\left(2\left(\tilde{\phi}_{\mathbf{k}}(\tau) - \phi_{\mathbf{k}}(\tau)\right)\right), \tag{65}$$

$$\frac{d\phi_{\mathbf{k}}(\tau)}{d\tau} = \Omega_{\mathbf{k}}(\tau) + \lambda_{\mathbf{k}}(\tau)\coth\left(2r_{\mathbf{k}}(\tau)\right)\sin\left(2\left(\tilde{\phi}_{\mathbf{k}}(\tau) - \phi_{\mathbf{k}}(\tau)\right)\right), \tag{66}$$

where $\tilde{\phi}_{\mathbf{k}}(\tau)$ is given by the following expressions:

$$\tilde{\phi}_{\mathbf{k}}(\tau) = -\frac{\pi}{2} + \frac{1}{2}\tan^{-1}\left[\frac{\Omega_{\mathbf{k}}(\tau)}{2}\left(\frac{z'(\tau)}{z(\tau)}\right)\left(\frac{1-c_s^2}{1+c_s^2}\right)\right]. \tag{67}$$

Expressions for $\lambda_{\mathbf{k}}(\tau)$ and $\Omega_{\mathbf{k}}(\tau)$ are already mentioned before in this section. Here we fix the boundary condition at the late time scale $\tau = \tau_0$ which fix the squeezing amplitude and the squeezing angle at $r_{\mathbf{k}}(\tau_0) = 1$ and $\phi_{\mathbf{k}}(\tau_0) = 1$. This boundary condition will help us to know the time evolution of squeezing amplitude $r_{\mathbf{k}}(\tau)$ and squeezing angle $\phi_{\mathbf{k}}(\tau)$ at any scale.

However, solving these above mentioned equations are useful in terms scale factor rather with respect to the conformal time scale as it captures more information than the previous. Also solving these equations are much more simpler with respect to the scale factor. For this reason we use:

$$\frac{d}{d\tau} = a'\frac{d}{da}. \tag{68}$$

In terms of scale factor these coupled differential equations can be recast as:

$$\frac{dr_{\mathbf{k}}(a)}{da} = \frac{\lambda_{\mathbf{k}}(a)}{a'}\cos\left(2\left(\tilde{\phi}_{\mathbf{k}}(a) - \phi_{\mathbf{k}}(a)\right)\right), \tag{69}$$

$$\frac{d\phi_{\mathbf{k}}(a)}{da} = \frac{\Omega_{\mathbf{k}}(a)}{a'} + \frac{\lambda_{\mathbf{k}}(a)}{a'}\coth\left(2r_{\mathbf{k}}(a)\right)\sin\left(2\left(\tilde{\phi}_{\mathbf{k}}(a) - \phi_{\mathbf{k}}(a)\right)\right), \tag{70}$$

The boundary condition is fixed here at the late time scale $\tau = \tau_0$ where $a(\tau_0) = a_0 = 1$ which fix the squeezing amplitude and the squeezing angle at $r_{\mathbf{k}}(a_0 = 1) = 1$ and $\phi_{\mathbf{k}}(a_0 = 1) = 1$. This boundary condition will help us to know the time evolution of squeezing amplitude $r_{\mathbf{k}}(a)$ and squeezing angle $\phi_{\mathbf{k}}(a)$ at any value of the scale factor.

VI. Notes on Quantum Circuit Complexity

In this section, we give an overview on the topic of quantum circuit complexity which is becoming a very popular area research in theoretical physics. The prime objective of quantum information processing science is to build quantum circuits for search algorithms or Shor's factoring [105–107], which can able to perform various unitary operations. Similarly, in the framework of computer science the same job is performed by complexity [108, 109], which is technically defined as the minimum number of gates required to perform a computation. This is completely a classical definition till now. In the quantum mechanical context this is interpreted as the minimum number of quantum gates required perform unitary operation [110] to get the estimation of quantum circuit complexity [111, 112]. However, in terms of implementing these unitary operations in logic gates are not very simple and a deeper understanding is required to perform such operations in quantum circuits.

There are many useful approaches people have been following for the last few years to compute the quantum circuit complexity. One of the path breaking approach is to geometrically quantify quantum circuit complexity measure [113–115]. The key idea is to find out the optimal quantum circuit which is actually mimics the role of solving problems of determining geodesics in the context of Riemannian geometry. In this approach the Riemannian metric is defined in the space of n number of qubits. The corresponding distance measure $\mathcal{D}(I, U)$, which is measuring the distance between the identity I and the target unitary U quantum operation is exactly interpreted as the quantum circuit complexity measure in the present context of discussion. In this terminology, minimizing this distance measure $\mathcal{D}(I, U)$ is physically interpreted as the computing the geodesic length in the Riemannian geometry spanned by n number of qubits. From the systematic rigorous development of the problem it is expected that this distance measure $\mathcal{D}(I, U)$ gives a good measure of the quantum circuit complexity in this context. For this underlying reason one can reliably use various computational methods based on Riemannian geometry to study the unexplored features of quantum circuit complexity. These methods are, Levi-Civita connection, geodesic, curvature, which is frequently used in this computation.

To demonstrate this concept in a crystal clear fashion let us first consider an unitary operator U which is helpful transforms a reference state $|\Psi_R\rangle$ to the target state $|\Psi_T\rangle$ via the following operation:

$$|\Psi_T\rangle = U|\Psi_R\rangle \tag{71}$$

This unitary operation play a pivotal role in the context of quantum computation where it is interpreted as ordered operation of unitary gates $U_i \forall i = 1, 2, \cdots, d$ represented by:

$$U = U_1 U_2 \cdots U_d, \tag{72}$$

where d is the depth of the quantum circuit. To determine the ultimate success of this unitary operation a tolerance Δ is introduced here, which is given by:

$$|| \, |\Psi_T\rangle - U \, |\Psi_R\rangle \, ||^2 \leq \Delta \tag{73}$$

The inclusion of such tolerance is perfectly consistent with the realistic applications where it extremely difficult to perform an exact unitary operation which can be interpreted in terms of ordered discrete set of unitary $U_i \forall i = 1, 2, \cdots, d$ quantum operations. It is important to mention here that, in this literature there exists infinite number of possibilities using which one can able to get the required target quantum state $|\Psi_T\rangle$ from the reference state $|\Psi_R\rangle$. In all of these formalisms depth of the optimal circuit is actually measured by quantum circuit complexity.

In the context of the computation of quantum circuit complexity measure using geometric technique in refs. [113–115] the authors have used the principles of Hamiltonian control. This approach is commonly known as Nielsen's geometric approach towards circuit complexity. This idea is further used in various contexts of quantum mechanical systems to explore the unknown features of quantum circuit complexity measure. In this geometric technique people don't actually measure the number of discrete gates to perform the previously mentioned unitary operation. In Nielsen's geometric approach the unitary operator U is path ordered and can be expressed in terms of the time-dependent Hamiltonian $H(\tau)$ as:

$$U = \overleftarrow{\mathcal{P}} \, \exp\left[-i \int_0^1 d\tau H(\tau)\right], \tag{74}$$

where the Hamiltonian operator can be written in terms of hermitian operators $\mathcal{O}_I$ and control functions $Y^I(\tau)$ as:

$$H(\tau) = \sum_I Y^I(\tau) \mathcal{O}_I \tag{75}$$

Here it is important to note that, the path ordering $\mathcal{P}$ operation is actually the geometric version of time ordering operation. It physically signifies the quantum circuit is made of non-commuting operators which are acting from the right to left side. In this description $Y^I(\tau)$ interpreted as particular gates which is added at a time s.

These possible paths finally construct the form of the unitary operations which is expressed as:

$$U = \overleftarrow{\mathcal{P}} \, \exp\left[-i \int_0^\tau d\tau' H(\tau')\right]. \tag{76}$$

Here the boundary conditions are fixed at:

1. At time scale $\tau = 0$ the unitary operator is fixed at $U(\tau = 0) = 1$.

2. At time scale $\tau = 1$ the unitary operator is fixed at $U(\tau = 1) = U$.

Also, hermitian operators $\mathcal{O}_I$ and control functions $Y^I(\tau)$ satisfy the following equation:

$$Y^I(\tau) \mathcal{O}_I = \partial_\tau U(\tau) U^{-1}(\tau), \tag{77}$$

which is representing a time-dependent Schrödinger equation describing path ordered unitary time evolution.

Instead of finding out the infinite possible ways to construct the unitary operations here we will restrict ourself only for the procedures which gives optimal answer. For this reason, we consider a cost function $\mathcal{F}(U, \vec{Y}(\tau))$ which is a local functional. For each of possible optimal path the cost is given by the following expression:

$$\mathcal{D}(U(\tau)) = \int_0^\tau d\tau' \; \mathcal{F}(U(\tau'), \partial_\tau U(\tau')) \tag{78}$$

Nielsen explicitly elaborated in his computation that, the variational method of minimizing this cost functional is nothing but determining the optimized quantum circuit in the present context.

Here this cost function $\mathcal{F}(U, \vec{Y}(\tau))$ should satisfy the following four properties which are a appended below point-wise:

1. **Continuity**:
   The cost functional should be continuous which is described by the symbol $\mathcal{F} \in \mathcal{C}^0$. This is quite reasonable from the physics perspectives.

2. **Positivity**:
   The technical definition of cost function $\mathcal{F}$ demands positivity which can be mathematically stated as:

   $$\mathcal{F}(U, v) \geq 0, \tag{79}$$

   where equality holds iff $v = 0$ which further implies that the reference and target quantum states are basically identical.

3. **Positive homogeneity**:
   For any positive real number λ and any arbitrary vector v, the following condition has to be satisfied:

   $$\mathcal{F}(\lambda v) = \lambda \; \mathcal{F}(v). \tag{80}$$

   This condition in technical language is identified as the positive homogeneity criteria.

4. **Triangle Inequality**:
   Last but not the least, the cost functional $\mathcal{F}$ satisfies the triangle inequality in the present context:

   $$\mathcal{F}(U, v + u) \leq \mathcal{F}(U, v) + \mathcal{F}(U, u) \;\; \forall \; (u, v) \tag{81}$$

   The equality holds iff the tangent vectors v and u are along the same ray which is coming out from the origin.

If the continuity condition is extended to $\mathcal{F} \in C^{\infty}$ then the cost functional is identified to be in nature and the corresponding manifold is known as *Finsler Manifold*. In this connection it is importnat to note that the Nielsen's geometric approach is computing the expression for geodesic in Finsler geometry and in this calculation the geodesic length is interpreted as the quantum circuit complexity.

In literature, there are different choices of these cost functions $\mathcal{F}(U, v)$ exists using which one can estimate quantum circuit complexity. These choices depends on how one defines the elementary gates to compute complexity measure from the corresponding setup. Some of the possibilities are quoted below which will be useful for this purpose:

$$\mathcal{F}_1(U, Y) = \sum_I |Y^I|, \tag{82}$$

$$\mathcal{F}_2(U, Y) = \sqrt{\sum_I |Y^I|^2}, \tag{83}$$

$$\mathcal{F}_m(U, Y) = \sum_I m_I |Y^I|, \tag{84}$$

$$\mathcal{F}_n(U, Y) = \sqrt{\sum_I n_I |Y^I|^2}. \tag{85}$$

Comments regarding the various definitions of the cost functions are appended below point-wise:

- Here the cost function $\mathcal{F}_1$ is the linear one which measures the counting of individual gates in the quantum circuit.
- Here $\mathcal{F}_2$ is the quadratic cost functional which is physically interpreted as measuring the proper distance in the corresponding manifold under consideration.
- Here the cost functions $\mathcal{F}_m$ (for linear measure) and $\mathcal{F}_n$ (for quadratic measure) are defined in presence of an additional penalty factors m_I and n_I, which is used to physically represent the fact that the certain direction are preferred over all possibilities in this context. This possibilities are very useful to describe the physical situations where the non-local qubits are discarded for consideration and the elemantray gates only coupled to the local type of qubits.

One can further introduce a more general class of homogeneous and inhomogeneous family of cost functions, which are given by the following expressions:

$$\mathcal{F}_\beta(U, Y) = \sum_I |Y^I|^\beta \tag{86}$$

$$\mathcal{F}_{\frac{1}{\beta}}(U, Y) = \sum_I |Y^I|^{\frac{1}{\beta}} \tag{87}$$

where the parameter $\beta \geq 1$ physically signify the degree of homogeneity in the present context. Here it is important to point that $\mathcal{F}_\beta$ was first introduced in the context of AdS/CFT to match the results obtained from the well known "Complexity = Action" and "Complexity = Volume" conjectures, about which we have mentioned in the introduction before.

VII. Quantum Circuit Complexity from two mode squeezed states

The framework of the two-mode squeezed quantum state is the simplest realization of the generalized multi-mode entangled quantum states using which one can study various quantum mechanical and information theoretic features, such as entanglement entropy, quantum circuit complexity, quantum discord, logarithmic negativity and many more. In this article we have only focussed on two aspects out of all of these possibilities, which are entanglement entropy and quantum circuit complexity. In this section and in the next section we discuss that using such useful two-mode squeezed state formalism how one can quantify these two measures within the framework of COSMOEFT. See refs. [116] for more details on two-mode squeezed states.

We have already defined the two-mode squeezed state operator in previous half of the paper, to remind ourselves again let us write down this expression again in this section which will be useful for the further discussion. In this context the two mode squeezing operator is describe as:

$$\hat{\mathcal{S}}(r_{\mathbf{k}}, \phi_{\mathbf{k}}) = \exp\left(r_{\mathbf{k}}\left[e^{-i\phi_{\mathbf{k}}}\hat{c}_{\mathbf{k}}\hat{c}_{-\mathbf{k}} - e^{i\phi_{\mathbf{k}}}\hat{c}^\dagger_{-\mathbf{k}}\hat{c}^\dagger_{\mathbf{k}}\right]\right). \tag{88}$$

where $r_{\mathbf{k}}$ and $\phi_{\mathbf{k}}$ are the squeezing parameters which represent the corresponding amplitude and angle. Here it is important to note that, these parameters are lying within the window, $0 \leq r_{\mathbf{k}} < \infty$ and $0 \leq \phi_{\mathbf{k}} \leq 2\pi$. In our computation the two-mode squeezed state is considered to be the target quantum state and this can be constructed by operating squeezing operator $\hat{\mathcal{S}}(r_{\mathbf{k}}, \phi_{\mathbf{k}})$ on two-mode initial vacuum state:

$$|\Psi_{\rm sq}\rangle_{\mathbf{k},-\mathbf{k}} = \hat{\mathcal{S}}(r_{\mathbf{k}}, \phi_{\mathbf{k}})\,|0_{\mathbf{k}}, 0_{-\mathbf{k}}\rangle\,. \tag{89}$$

Here it is important to note that, the initial vacuum state structure is given by:

$$|0_{\mathbf{k}}, 0_{-\mathbf{k}}\rangle := |0\rangle_{\mathbf{k}}\,|0\rangle_{-\mathbf{k}}\,. \tag{90}$$

In terms of two-mode occupation number states, one can further express that the target two-mode squeezed state is given by the following expression:

$$|\Psi_{\rm sq}\rangle_{\mathbf{k},-\mathbf{k}} = \frac{1}{\cosh r_{\mathbf{k}}}\sum_{n=0}^{\infty}(-1)^n e^{in\phi_{\mathbf{k}}}(\tanh r_{\mathbf{k}})^n\,|n_{\mathbf{k}}, n_{-\mathbf{k}}\rangle\,. \tag{91}$$

Using this target two-mode squeezed quantum states we now can able to compute the expressions for quantum circuit complexity measure and entanglement entropy. Using these initial reference and target quantum states one

needs to construct wave function which is Gaussian in nature.

To serve this purpose let us first write down the position and momentum variables by the following expressions:

$$\hat{q}_{\mathbf{k}} = \frac{1}{\sqrt{2\Omega_{\mathbf{k}}}}\left(\hat{c}^{\dagger}_{\mathbf{k}} + \hat{c}_{\mathbf{k}}\right) \tag{92}$$

$$\hat{p}_{\mathbf{k}} = i\sqrt{\frac{\Omega_{\mathbf{k}}}{2}}\left(\hat{c}^{\dagger}_{\mathbf{k}} - \hat{c}_{\mathbf{k}}\right) \tag{93}$$

where they satisfy the following commutation condition:

$$[\hat{q}_{\mathbf{k}}, \hat{p}_{\mathbf{k}'}] = i\delta^3(\mathbf{k} - \mathbf{k}'). \tag{94}$$

Consequently, in the position space the reference and target state can be expressed as:

$$\begin{aligned}\Psi_R(q_{\mathbf{k}}, q_{-\mathbf{k}}) &= \langle q_{\mathbf{k}}, q_{-\mathbf{k}} | 0_{\mathbf{k}}, 0_{-\mathbf{k}}\rangle \\ &= \left(\frac{\Omega_{\mathbf{k}}}{\pi}\right)^{\frac{1}{4}} \exp\left(-\frac{\Omega_{\mathbf{k}}}{2}\left(q^2_{\mathbf{k}} + q^2_{-\mathbf{k}}\right)\right), \quad (95)\\ \Psi_T(q_{\mathbf{k}}, q_{-\mathbf{k}}) &= \langle q_{\mathbf{k}}, q_{-\mathbf{k}} | \Psi_{\rm sq}\rangle_{\mathbf{k},-\mathbf{k}} \quad (96)\\ &= \frac{e^{A\left(q^2_{\mathbf{k}} + q^2_{-\mathbf{k}}\right) - Bq_{\mathbf{k}}q_{-\mathbf{k}}}}{{\rm cosh}r_{\mathbf{k}}\sqrt{\pi}\sqrt{1 - e^{-4i\phi_{\mathbf{k}}}{\rm tanh}^2 r_{\mathbf{k}}}} \quad (97)\end{aligned}$$

where, the coefficients A and B can be expressed in terms of squeezing amplitude $r_{\mathbf{k}}$ and squeezing angle $\phi_{\mathbf{k}}$ by the following expressions:

$$A = \frac{\Omega_{\mathbf{k}}}{2}\left(\frac{e^{-4i\phi_{\mathbf{k}}}{\rm tanh}^2 r_{\mathbf{k}} + 1}{e^{-4i\phi_{\mathbf{k}}}{\rm tanh}^2 r_{\mathbf{k}} - 1}\right), \tag{98}$$

$$B = 2\Omega_{\mathbf{k}}\left(\frac{e^{-2i\phi_{\mathbf{k}}}{\rm tanh}r_{\mathbf{k}}}{e^{-4i\phi_{\mathbf{k}}}{\rm tanh}^2 r_{\mathbf{k}} - 1}\right). \tag{99}$$

We further introduce few more relevant quantities which are written in terms of A, B and $\Omega_{\mathbf{k}}$ as:

$$\Sigma_{\mathbf{k}} = (-2A + B), \tag{100}$$

$$\Sigma_{-\mathbf{k}} = -(2A + B), \tag{101}$$

$$\omega_{\mathbf{k}} = \omega_{-\mathbf{k}} = \frac{\Omega_{\mathbf{k}}}{2} \tag{102}$$

In the following subsections and next section we will discuss about covariance matrix method and Nielsen's method to compute the quantum circuit complexity and entanglement entropy using two-mode squeezed state formalism developed in this section.

A. Complexity via Covariance matrix method

In this subsection we will discuss using covariance matrix method how one can able to quantify the quantum circuit complexity measure. It will turns out that the final expression of circuit complexity computed in this method solely depend on the squeezing amplitude $r_{\mathbf{k}}$ and completely independent of the squeezing angle $\phi_{\mathbf{k}}$.

As we have already seen in that in the present context the initial reference and target states, represented by eqn (95) and eqn (96) are Gaussian wave functions, one can easily use the covariance matrix method for the computation of quantum circuit complexity measure. The corresponding covariance matrices for the initial reference state and the target state are given by the following expressions:

$$G^{s=0}_k = \begin{pmatrix} \frac{1}{\Omega_{\mathbf{k}}} & 0 & 0 & 0 \\ 0 & \Omega_{\mathbf{k}} & 0 & 0 \\ 0 & 0 & \frac{1}{\Omega_{\mathbf{k}}} & 0 \\ 0 & 0 & 0 & \Omega_{\mathbf{k}} \end{pmatrix} \tag{103}$$

$$G^{s=1}_k = \begin{pmatrix} \frac{1}{{\rm Re}(\Sigma_{\mathbf{k}})} & -\frac{{\rm Im}(\Sigma_{\mathbf{k}})}{{\rm Re}(\Sigma_{\mathbf{k}})} & 0 & 0 \\ -\frac{{\rm Im}(\Sigma_{\mathbf{k}})}{{\rm Re}(\Sigma_{\mathbf{k}})} & \frac{|\Sigma_{\vec{k}}|^2}{{\rm Re}(\Sigma_{\mathbf{k}})} & 0 & 0 \\ 0 & 0 & \frac{1}{{\rm Re}(\Sigma_{-\mathbf{k}})} & -\frac{{\rm Im}(\Sigma_{-\mathbf{k}})}{{\rm Re}(\Sigma_{-\mathbf{k}})} \\ 0 & 0 & -\frac{{\rm Im}(\Sigma_{-\mathbf{k}})}{{\rm Re}(-\Sigma_{\mathbf{k}})} & \frac{|\Sigma_{-\mathbf{k}}|^2}{{\rm Re}(\Sigma_{-\mathbf{k}})} \end{pmatrix} \tag{104}$$

where the symbols $\Sigma_{\mathbf{k}}$ and $\Sigma_{-\mathbf{k}}$ are explicitly defined in eqn (100) and eqn (101). The good part of the computation using covariance matrix method is the above mentioned covariance matrices for the initial reference and target state capturing the full information of the Gaussian wave functions written for these two states. Using this method the computed quantum circuit complexity measure the logical quantum gates to transform reference covariance matrix $G^{s=0}_k$ to target covariance matrix $G^{s=1}_k$. For the computational simplicity purpose we now factorize the 4×4 reference and target covariance matrix into two two non-trivial 2×2 symmetric blocks for each of the cases. For the initial reference state we have following two symmetric blocks:

$$G^{s=0}_{k=0} = \begin{pmatrix} \frac{1}{\Omega_{\mathbf{k}}} & 0 \\ 0 & \Omega_{\mathbf{k}} \end{pmatrix}, \tag{105}$$

$$G^{s=0}_{k=1} = \begin{pmatrix} \frac{1}{\Omega_{\mathbf{k}}} & 0 \\ 0 & \Omega_{\mathbf{k}} \end{pmatrix} \tag{106}$$

On the other hand , for the target state we have:

$$G^{s=1}_{k=0} = \begin{pmatrix} \frac{1}{\mathrm{Re}(\Sigma_{\mathbf{k}})} & -\frac{\mathrm{Im}(\Sigma_{\mathbf{k}})}{\mathrm{Re}(\Sigma_{\mathbf{k}})} \\ -\frac{\mathrm{Im}(\Sigma_{\mathbf{k}})}{\mathrm{Re}(\Sigma_{\mathbf{k}})} & \frac{|\Sigma_{\vec{k}}|^2}{\mathrm{Re}(\Sigma_{\mathbf{k}})} \end{pmatrix}, \tag{107}$$

$$G^{s=1}_{k=1} = \begin{pmatrix} \frac{1}{\mathrm{Re}(\Sigma_{-\mathbf{k}})} & -\frac{\mathrm{Im}(\Sigma_{-\mathbf{k}})}{\mathrm{Re}(\Sigma_{-\mathbf{k}})} \\ -\frac{\mathrm{Im}(\Sigma_{-\mathbf{k}})}{\mathrm{Re}(\Sigma_{-\mathbf{k}})} & \frac{|\Sigma_{-\mathbf{k}}|^2}{\mathrm{Re}(\Sigma_{-\mathbf{k}})} \end{pmatrix}, \tag{108}$$

Here the total complexity is given by the sum of the individual complexities of the two blocks, where we have to perform an additional sum over all $\Omega_{\mathbf{k}}$.

Now to make the calculation simpler we perform the following change of basis for the each blocks:

$$\tilde{G}^{s=1} = SG^{s=1}S^T, \tag{109}$$
$$\tilde{G}^{s=0} = SG^{s=0}S^T. \tag{110}$$

where, S is a similar matrix which is given by:

$$S = \begin{pmatrix} \sqrt{\Omega_{\mathbf{k}}} & 0 \\ 0 & \frac{1}{\sqrt{\Omega_{\mathbf{k}}}} \end{pmatrix} \tag{111}$$

This further implies:

$$\tilde{G}^{s=0} = \mathbb{1}_{2\times 2}, \tag{112}$$

$$\tilde{G}^{s=1} = \begin{pmatrix} \frac{\Omega_{\mathbf{k}}}{\mathrm{Re}(\Sigma_{\mathbf{k}})} & -\frac{\mathrm{Im}(\Sigma_{\mathbf{k}})}{\mathrm{Re}(\Sigma_{\mathbf{k}})} \\ -\frac{\mathrm{Im}(\Sigma_{\mathbf{k}})}{\mathrm{Re}(\Sigma_{\mathbf{k}})} & \frac{|\Sigma_{\mathbf{k}}|^2}{\Omega_k \mathrm{Re}(\Sigma_{\mathbf{k}})} \end{pmatrix}, \tag{113}$$

In this method the corresponding unitary transformation of wave functions can be represented as:

$$\tilde{G}^s = \tilde{U}(\tau)\tilde{G}^{s=0}\tilde{U}^T(\tau). \tag{114}$$

Further these unitary transformations parametrized with the help of logical gates which satisfy the $SL(2,R)$ algebra and the corresponding structure of the unitary matrix is given by:

$$\tilde{U}(\tau) = \begin{pmatrix} \cos\mu(\tau)\cosh\rho(\tau) - \sin\theta(\tau)\sinh\rho(\tau) & -\sin\mu(\tau)\cosh\rho(\tau) + \cos\theta(\tau)\sinh\rho(\tau) \\ \sin\mu(\tau)\cosh\rho(\tau) + \cos\theta(\tau)\sinh\rho(\tau) & \cos\mu(\tau)\cosh\rho(\tau) + \sin\theta(\tau)\sinh\rho(\tau) \end{pmatrix} \tag{115}$$

where, μ, ρ, θ are the coordinates which characterize $SL(2,R)$ group. Next, we set the following two boundary conditions:

$$\tilde{G}^{s=1} = \tilde{U}(\tau=1)\tilde{G}^{s=0}\tilde{U}(\tau=1)^T \tag{116}$$
$$\tilde{G}^{s=0} = \tilde{U}(\tau=0)\tilde{G}^{s=0}\tilde{U}(\tau=0)^T \tag{117}$$

These boundary conditions applied with the $SL(2,R)$ parametrized unitary transformations further gives:

$$\cosh 2\rho(1) = \frac{\Omega_{\mathbf{k}}^2 + |\Sigma_{\mathbf{k}}|^2}{2\Omega_{\mathbf{k}}\mathrm{Re}(\Sigma_{\mathbf{k}})}, \quad \tan(\theta(1)+\mu(1)) = \frac{\Omega_{\mathbf{k}}^2 - |\Sigma_{\mathbf{k}}|^2}{2\Omega_{\mathbf{k}}\mathrm{Im}(\Sigma_{\mathbf{k}})}, \quad \rho(0) = 0, \quad \theta(0)+\mu(0) = c. \tag{118}$$

For the simplification purpose we choose the following constraints:

1. $\mu(\tau=1) = \mu(\tau=0) = 0$.

2. $\theta(\tau=0) = \theta(\tau=1) = c = \tan^{-1}\left(\frac{\Omega_{\mathbf{k}}^2 - |\Sigma_{\mathbf{k}}|^2}{2\Omega_{\mathbf{k}}\mathrm{Im}(\Sigma_{\mathbf{k}})}\right)$.

Given this constraint conditions the final form of the $SL(2,R)$ metric for $\tilde{U}$ is given by:

$$ds^2 = \Big(d\rho^2 + \cosh 2\rho\, \cosh^2\rho d\mu^2 + \cosh 2\rho\, \sinh^2\rho\, d\theta^2 - \sinh^2 2\rho\, d\mu\, d\theta\Big). \tag{119}$$

Now from the boundary conditions stated in eqn (118),

we get the following expressions:

$$\rho_{\mathbf{k}}(\tau=1)=\frac{1}{2}\cosh^{-1}\left[\frac{\Omega_{\mathbf{k}}^2+|\Sigma_{\mathbf{k}}|^2}{2\Omega_{\mathbf{k}}\mathrm{Re}(\Sigma_{\mathbf{k}})}\right], \tag{120}$$

$$\rho_{-\mathbf{k}}(\tau=1)=\frac{1}{2}\cosh^{-1}\left[\frac{\Omega_{-\mathbf{k}}^2+|\Sigma_{-\mathbf{k}}|^2}{2\Omega_{-\mathbf{k}}\mathrm{Re}(\Sigma_{-\mathbf{k}})}\right]. \tag{121}$$

Here the simple geodesic is represented by a straight line which is given by, $\rho(\tau)=\rho(\tau=1)\tau$. Finally, to get the expression for the quantum circuit complexity measure using covariance matrix method we have to sum both the contributions coming from two momentum modes in squeezed state formalism. Consequently, considering the contributions from both linear and quadratic cost functions the corresponding circuit complexity measures C_1 and C_2, can be computed as:

$$C_1=\rho_{\mathbf{k}}(\tau=1)+\rho_{-\mathbf{k}}(\tau=1)=\frac{1}{2}\left\{\cosh^{-1}\left(\frac{\Omega_{\mathbf{k}}^2+|\Sigma_{\mathbf{k}}|^2}{2\Omega_{\mathbf{k}}\mathrm{Re}(\Sigma_{\mathbf{k}})}\right)+\cosh^{-1}\left(\frac{\Omega_{-\mathbf{k}}^2+|\Sigma_{-\mathbf{k}}|^2}{2\Omega_{-\mathbf{k}}\mathrm{Re}(\Sigma_{-\mathbf{k}})}\right)\right\}, \tag{122}$$

$$C_2=\sqrt{\rho_{\mathbf{k}}(\tau=1)^2+\rho_{-\mathbf{k}}(\tau=1)^2}=\frac{1}{2}\sqrt{\left(\cosh^{-1}\left[\frac{\Omega_{\mathbf{k}}^2+|\Sigma_{\mathbf{k}}|^2}{2\Omega_{\mathbf{k}}\mathrm{Re}(\Sigma_{\mathbf{k}})}\right]\right)^2+\left(\cosh^{-1}\left[\frac{\Omega_{-\mathbf{k}}^2+|\Sigma_{-\mathbf{k}}|^2}{2\Omega_{-\mathbf{k}}\mathrm{Re}(\Sigma_{-\mathbf{k}})}\right]\right)^2}. \tag{123}$$

Further substituting the explicit form of the functions $\Sigma_{\mathbf{k}}$ and $\Sigma_{-\mathbf{k}}$ the quantum circuit complexity measure reduces to following simplified form:

$$C_1=4\ r_{\mathbf{k}} \tag{124}$$

$$C_2=2\sqrt{2}\ r_{\mathbf{k}}. \tag{125}$$

These results completely independent of the squeezing angle $\phi_{\mathbf{k}}$ and only characterized by $r_{\mathbf{k}}$. These two circuit complexities computed from linear and quadratic cost functions are related to each other by the following expression:

$$C_1=\sqrt{2}\ C_2. \tag{126}$$

For small squeezing amplitude $r_{\mathbf{k}}\to 0$, $C_1, C_2\approx 0$ which implies the reference and target states are same.

B. Complexity via Nielsen's wave-function method

Nielsen's approach deals with wave functions which finally give rise to expression for quantum circuit complexity for the two mode squeezed states, which we will show that dependent on both the squeezing parameters, the squeezing amplitude $r_{\mathbf{k}}$ and the squeezing angle $\phi_{\mathbf{k}}$. However, the general philosophical construction of computing the quantum circuit complexity measure of Nielsen's approach is similar to that of covariance matrix method. In this approach, we will directly quantify the quantum circuit complexity measure using the initial reference and target states using the two-mode squeezed state formalism.

Let us start with the exponent of the target two-mode squeezed states as stated in eqn (96), which after diagonalization can be expressed as:

$$\Psi_{\mathrm{T}}=\mathcal{N}\exp\left(-\frac{1}{2}q_a\tilde{\mathcal{W}}^{ab}q_b\right) \tag{127}$$

where, $\mathcal{N}$ is the normalization constant of the wave function , which is represented by the denominator in eqn (96) and, the matrix $\tilde{\mathcal{W}}$ can be expressed as:

$$\tilde{\mathcal{W}}=\begin{pmatrix}-2A+B & 0\\ 0 & -2A-B\end{pmatrix}=\begin{pmatrix}\Sigma_{\mathbf{k}} & 0\\ 0 & \Sigma_{-\mathbf{k}}\end{pmatrix}. \tag{128}$$

Also, the initial reference state is represented by the following expression:

$$\Psi_{\mathrm{R}}=\mathcal{N}\exp\left(-\frac{\Omega_{\mathbf{k}}}{2}\left(q_{\mathbf{k}}^2+q_{-\mathbf{k}}^2\right)\right). \tag{129}$$

In this context the reference and target wave functions of the Gaussian structure, can be generally written as:

$$\Psi^{\tau}=\mathcal{N}\exp\left(-\frac{1}{2}\left(v_a.\mathcal{K}_{ab}^{\tau}.v_b\right)\right) \tag{130}$$

where, the basis vector $v=(q_{\mathbf{k}},q_{-\mathbf{k}})$ and $\mathcal{K}^{\tau}$ is a diagonal matrix. For the target state eq: (127) and reference state (129) the structure of this matrix is given by:

$$\mathcal{K}^{\tau=1}=\begin{pmatrix}\Sigma_{\mathbf{k}} & 0\\ 0 & \Sigma_{-\mathbf{k}}\end{pmatrix}=\mathcal{W}, \tag{131}$$

$$\mathcal{K}^{\tau=0}=\begin{pmatrix}\Omega_{\mathbf{k}} & 0\\ 0 & \Omega_{-\mathbf{k}}\end{pmatrix}. \tag{132}$$

Further, the unitary transformation on the matrix $\mathcal{K}$ is given by:

$$\mathcal{K}^{\tau}=\mathcal{U}(\tau).\mathcal{K}^{\tau=0}.\mathcal{U}^T(\tau), \tag{133}$$

and corresponding two boundary conditions are given by the following expressions:

$$\mathcal{K}^{\tau=1}=\mathcal{U}(\tau=1).\mathcal{K}^{\tau=0}.\mathcal{U}^T(\tau=1), \tag{134}$$

$$\mathcal{K}^{\tau=0}=\mathcal{U}(\tau=0).\mathcal{K}^{\tau=0}.\mathcal{U}^T(\tau=0). \tag{135}$$

Here the unitary operator $\mathcal{U}$ can be parametrized as in terms of eqn (74) provided at $\tau = 1$, the target state is reached. In principle the elements of $\mathcal{K}^{\tau=1}$ and $\mathcal{K}^{\tau=0}$ matrices are complex and for this reason the logical gates are parametrized by $GL(2,\mathbb{C})$ group. In this context, the tangent vectors Y^I as appearing in eqn (77) are complex and can be expressed in terms of its generators $\mathcal{O}_I$ by the following expression:

$$Y^I = \mathrm{Tr}(\partial_\tau U(\tau) U^{-1}(\tau) \mathcal{O}_I^T) \tag{136}$$

where, the generators satisfy the following constraint:

$$\mathcal{O}_I^T = \mathcal{O}_I^{-1}, \quad \mathrm{Tr}(\mathcal{O}_I.\mathcal{O}_J^T) = \delta^{IJ}, \forall I, J = 0, 1, 2, 3 \tag{137}$$

The corresponding $GL(2,\mathbb{C})$ metric is then given by the following expression:

$$ds^2 = G_{IJ} dY^I dY^{*J}. \tag{138}$$

Now for our computation we fix $G_{IJ} = \delta^{IJ}$ which will further fix the penalty factors. Further, we fix the off-diagonal elements in $GL(2,\mathbb{C})$ matrix to zero as they increase the distance between the initial reference and target states. In this diagonal basis then the unitary operator $U(\tau)$ takes the following simplified form:

$$U(\tau) = \exp\left(\sum_{i\in(\mathbf{k},-\mathbf{k})} \mu^i(\tau) \mathcal{O}_i^{diag}\right), \tag{139}$$

where, $\mu^i(\tau)$ represent complex parameters and $\mathcal{O}_i^{diag}$ are the corresponding generators in the diagonal basis having i number of identity at the position of the diagonal elements. In the new diagonal basis the corresponding metric takes the following simplified structure:

$$ds^2 = \sum_{i\in(\mathbf{k},-\mathbf{k})} (d\mu^{i,\mathrm{Re}})^2 + \sum_{i\in(\mathbf{k},-\mathbf{k})} (d\mu^{i,\mathrm{Im}})^2, \tag{140}$$

where the notation Re and Im signify the real and imaginary part of the complex parameters μ. Here the geodesic is straight line, which is given by:

$$\mu^{i,p}(\tau) = \mu^{i,p}(\tau = 1) + \mu^{i,p}(\tau = 0) \forall i \in (\mathbf{k}, -\mathbf{k}) \tag{141}$$

for $p = \mathrm{Re}$ and Im. Using these boundary conditions we get:

$$\mu^{i,\mathrm{Re}}(\tau = 0) = \mu^{i,\mathrm{Im}}(\tau = 0) = 0, \tag{142}$$

$$\mu^{i,\mathrm{Re}}(\tau = 1) = \frac{1}{2} \ln \left|\frac{\Sigma_i}{\omega_i}\right|, \tag{143}$$

$$\mu^{i,\mathrm{Im}}(\tau = 1) = \frac{1}{2} \tan^{-1} \frac{\mathrm{Im}(\Sigma_i)}{\mathrm{Re}(\Sigma_i)}. \tag{144}$$

where the index $i \in (\mathbf{k}, -\mathbf{k})$ correspond two-modes in the squeezed state formalism.

Finally, the quantum circuit complexity measure for linear and quadratic cost functions, $C_1(\Omega_\mathbf{k})$ and $C_2(\Omega_\mathbf{k})$ are given by:

$$\begin{aligned} C_1 &= \mu^{\mathbf{k},\mathrm{Re}}(\tau = 1) + \mu^{-\mathbf{k},\mathrm{Re}}(\tau = 1) + \mu^{\mathbf{k},\mathrm{Im}}(\tau = 1) + \mu^{-\mathbf{k},\mathrm{Im}}(\tau = 1) \\ &= \frac{1}{2}\left\{ \ln\left|\frac{\Sigma_\mathbf{k}}{\omega_\mathbf{k}}\right| + \ln\left|\frac{\Sigma_{-\mathbf{k}}}{\omega_{-\mathbf{k}}}\right| + \tan^{-1}\frac{\mathrm{Im}(\Sigma_\mathbf{k})}{\mathrm{Re}(\Sigma_\mathbf{k})} + \tan^{-1}\frac{\mathrm{Im}(\Sigma_{-\mathbf{k}})}{\mathrm{Re}(\Sigma_{-\mathbf{k}})} \right\}, \end{aligned} \tag{145}$$

$$\begin{aligned} C_2 &= \sqrt{(\mu^{\mathbf{k},\mathrm{Re}}(\tau = 1))^2 + (\mu^{-\mathbf{k},\mathrm{Re}}(\tau = 1))^2 + (\mu^{\mathbf{k},\mathrm{Im}}(\tau = 1))^2 + (\mu^{-\mathbf{k},\mathrm{Im}}(\tau = 1))^2} \\ &= \frac{1}{2}\sqrt{\left(\ln\left|\frac{\Sigma_\mathbf{k}}{\omega_\mathbf{k}}\right|\right)^2 + \left(\ln\left|\frac{\Sigma_{-\mathbf{k}}}{\omega_{-\mathbf{k}}}\right|\right)^2 + \left(\tan^{-1}\frac{\mathrm{Im}(\Sigma_\mathbf{k})}{\mathrm{Re}(\Sigma_\mathbf{k})}\right)^2 + \left(\tan^{-1}\frac{\mathrm{Im}(\Sigma_{-\mathbf{k}})}{\mathrm{Re}(\Sigma_{-\mathbf{k}})}\right)^2} \end{aligned} \tag{146}$$

Further using eqns. (100), (101) and (102) the quantum circuit complexity measure can be expressed in terms of the squeezing parameters, $r_\mathbf{k}$ and $\phi_\mathbf{k}$ as:

$$C_1 = \left|\ln\left|\frac{1 + \exp(-2i\phi_\mathbf{k})\tanh r_\mathbf{k}}{1 - \exp(-2i\phi_\mathbf{k})\tanh r_\mathbf{k}}\right|\right| + \left|\tanh^{-1}(\sin(2\phi_\mathbf{k})\sinh(2r_\mathbf{k}))\right|, \tag{147}$$

$$C_2 = \frac{1}{\sqrt{2}}\sqrt{\left(\ln\left|\frac{1 + \exp(-2i\phi_\mathbf{k})\tanh r_\mathbf{k}}{1 - \exp(-2i\phi_\mathbf{k})\tanh r_\mathbf{k}}\right|\right)^2 + \left(\tanh^{-1}(\sin(2\phi_\mathbf{k})\sinh(2r_\mathbf{k}))\right)^2}. \tag{148}$$

As a special case if we set $\phi_\mathbf{k} \to -\pi/2$ and consider large squeezing amplitude $r_\mathbf{k}$ then we get back the same

connecting relationship between the different measure of circuit complexities obtained from linear and quadratic cost functions in the case of covariance matrix method. This is given by:

$$C_1 \approx \sqrt{2}C_2 \approx \left|\ln\left(\frac{1-\tanh r_{\mathbf{k}}}{1+\tanh r_{\mathbf{k}}}\right)\right| \approx r_{\mathbf{k}} \tag{149}$$

VIII. Entanglement entropy of two mode squeezed states

In this section, our prime objective is to derive the expression for entanglement entropy using the previously mentioned two-mode squeezed state formalism. This will going to help us to make a comparison between quantum circuit complexity measure and entanglement entropy, which we will discuss in the next section in detail. Here its important to note that before doing the further computation that the two squeezed modes having momenta $\mathbf{k}$ and $-\mathbf{k}$ are (1) entangled with each other, (2) strongly correlated with each other. and (3) they are highly symmetric with each other. In this description the target state which is a squeezed state $|\psi_{\rm T}\rangle_{\mathbf{k},-\mathbf{k}} = |\psi_{\rm sq}\rangle_{\mathbf{k},-\mathbf{k}}$ is the eigenket of the operator $(\hat{n}_{\mathbf{k}} - \hat{n}_{-\mathbf{k}})$ having zero eigenvalue, where $\hat{n}_{\mathbf{k}} = \hat{c}^{\dagger}_{\mathbf{k}}\hat{c}_{-\mathbf{k}}$ and $\hat{n}_{-\mathbf{k}} = \hat{c}^{\dagger}_{-\mathbf{k}}\hat{c}_{\mathbf{k}}$ are physically interpreted as the number operators of the two squeezed modes having momenta $\mathbf{k}$ and $-\mathbf{k}$ respectively. In this setup if we compute the average number density for photons corresponding to each momentum modes then it will turn out to be exactly same and it is given by the following expression:

$$\langle \hat{n_{\mathbf{k}}}\rangle = \langle \hat{n}_{-\mathbf{k}}\rangle = \sinh^2 r_{\mathbf{k}}, \tag{150}$$

which will tend to trivially zero when the squeezing amplitude parameter $r_{\mathbf{k}}$ is extremely small. Here the $\langle\cdots\rangle$ operation is performed using the $|\psi_{\rm sq}\rangle_{\mathbf{k},-\mathbf{k}}$ state.

The reduced density operators for the two squeezed modes having momenta $\mathbf{k}$ and $-\mathbf{k}$ are given by the following expressions:

$$\hat{\rho}_{\mathbf{k}} = \sum_{n=0}^{\infty} \frac{1}{(\cosh\, r_{\mathbf{k}})^2}(\tanh\, r_k)^{2n}\langle n_{\mathbf{k}}|n_{\mathbf{k}}\rangle, \tag{151}$$

$$\hat{\rho}_{-\mathbf{k}} = \sum_{n=0}^{\infty} \frac{1}{(\cosh\, r_{-\mathbf{k}})^2}(\tanh\, r_{-\mathbf{k}})^{2n}\langle n_{-\mathbf{k}}|n_{-\mathbf{k}}\rangle. \tag{152}$$

Using these reduced density operators for two-mode squeezed states one can now compute entanglement entropy which is derived using the von-Neumann and Rényi entropy measure.

In general von-Neumann entropy measure can be expressed by the following expression in terms of the reduced density operator:

$$S(\hat{\rho}) = -{\rm Tr}\,(\hat{\rho}\,\ln\hat{\rho}) \tag{153}$$

For pure state von-Neumann entropy is zero and for mixed state it is no-zero, in which case it is meaningful to compute this expression for a given quantum mechanical system.

For the basis where the form of the reduced density operator is diagonal, entanglement entropy can be computed as:

$$S(\hat{\rho}) = -\sum_j \rho_{jj}\,\ln\rho_{jj}. \tag{154}$$

In this diagonal basis the entanglement entropy of two-mode squeezed states using the von-Neumann entropy measure can be written as:

$$S(\hat{\rho_{\mathbf{k}}}) = -{\rm Tr}\,(\hat{\rho_{\mathbf{k}}}\,\ln\hat{\rho_{\mathbf{k}}})\,, \tag{155}$$

$$S(\hat{\rho}_{-\mathbf{k}}) = -{\rm Tr}\,(\hat{\rho}_{-\mathbf{k}}\,\ln\hat{\rho}_{-\mathbf{k}})\,. \tag{156}$$

After doing simplification we found that the von-Neumann entropy measure from both the squeezed modes turns out to be same and given by the following expression:

$$\begin{aligned} S(\hat{\rho_{\mathbf{k}}}) &= S(\hat{\rho}_{-\mathbf{k}}) \\ &= -\sum_{n=0}^{\infty}\frac{\tanh^{2n} r_{\mathbf{k}}}{\cosh^2 r_{\mathbf{k}}}\ln\left(\frac{\tanh^{2n} r_{\mathbf{k}}}{\cosh^2 r_{\mathbf{k}}}\right) \\ &= \left\{\ln(\cosh^2 r_{\mathbf{k}})\cosh^2 r_{\mathbf{k}} - \ln(\sinh^2 r_{\mathbf{k}})\sinh^2 r_{\mathbf{k}}\right\}. \end{aligned} \tag{157}$$

This result shows that the von-Neumann entropy measure is completely independent of the squeezing angle $\phi_{\mathbf{k}}$ and solely determined by the squeezing amplitude $r_{\mathbf{k}}$.

Further one can generalize von-Neumann entropy to get the expression for Rényi entropy from the reduced density operator:

$$S_\alpha(\hat{\rho}) = \frac{1}{1-\alpha}\ln\left({\rm Tr}\;\hat{\rho}^{\alpha}\right) \tag{158}$$

which can be written for the two-mode squeezed states as:

$$S_\alpha(\hat{\rho_{\mathbf{k}}}) = \frac{1}{1-\alpha}\ln\left({\rm Tr}\;\hat{\rho}^{\alpha}_{\mathbf{k}}\right), \tag{159}$$

$$S_\alpha(\hat{\rho}_{-\mathbf{k}}) = \frac{1}{1-\alpha}\ln\left({\rm Tr}\;\hat{\rho}^{\alpha}_{-\mathbf{k}}\right). \tag{160}$$

After doing simplification we found that the Rényi entropy measure from both the squeezed modes turns out to be same and given by the following expression:

$$\begin{aligned} S_\alpha(\hat{\rho_{\mathbf{k}}}) &= S_\alpha(\hat{\rho}_{-\mathbf{k}}) \\ &= \frac{2\alpha\,\ln\,{\rm cosh}r_{\mathbf{k}} + \ln(1-\tanh^{2\alpha} r_{\mathbf{k}})}{\alpha-1}. \end{aligned} \tag{161}$$

where the Rényi Parameter $\alpha \geq 0$ and $\alpha \to 1$ limit gives back the expression for von-Neumann entropy measure. Again, we can clearly observe that the Rényi entropy measure solely described by the squeezing amplitude parameter $r_{\mathbf{k}}$ just like von-Neumann entropy measure. One can consider a limiting situation where we set the squeezing amplitude parameter $r_{\mathbf{k}}$ is very large, which will give

rise to following simplified expression for the Rényi entropy measure:

$$\lim_{r_{\mathbf{k}}\to\infty} S_\alpha \approx \frac{2\alpha}{(\alpha-1)}\, r_{\mathbf{k}}. \tag{162}$$

This discussion can be further extended to explicitly compute the expression for the effective temperature of photon. The average photon number of the thermal field is given by:

$$\langle \hat{n}_i \rangle = \frac{1}{(\exp(\omega_i/T)-1)} = \sinh^2 r_{\mathbf{k}} \quad \forall\, i \in (\mathbf{k}, -\mathbf{k}). \tag{163}$$

where we set $\hbar = 1$ and Boltzmann constant $k_B = 1$ in natural unit system. Then, one can compute the effective temperature as:

$$T = \omega_i \, \ln\left(\frac{\langle \hat{n}_i \rangle}{\langle \hat{n}_i \rangle + 1}\right) = \frac{\omega_i}{2\ln(\coth\, r_{\mathbf{k}})} \quad \forall\, i \in (\mathbf{k}, -\mathbf{k}). \tag{164}$$

Here the frequency parameter $\omega_i \forall\, i \in (\mathbf{k}, -\mathbf{k})$ is represented by eqn. (102).

IX. Quantum Circuit Complexity Vs Entanglement

In this section our prime objective is to make a comparison between quantum circuit complexity measure and entanglement entropy measure, particularly when both of them are described by two-mode squeezed state formalism. Recently in ref. [68] the authors showed that there exist some underlying relationship between the entangling power and quantum circuit complexity measure.

It seems from the origin of quantum circuit complexity measure and entanglement entropy that they are appearing from different physical origin. But in ref. [68] the authors showed that it is possible to use entanglement entropy measure to bound the circuit complexity for (1) small value of circuit cost and (2) small value of entanglement. Additionally, in ref. [68] there are few interesting arguments presented which will help us to know about such connections for the system under consideration in this paper. Some of them are appended below point-wise:

- Quantum gates which are close to the identity gates performs little entanglement from entangled states.
- If the entanglement entropy measure grows linearly with respect to time then the quantum circuit complexity measure also grows linearly. Particularly this linear growth of quantum entanglement is a generic feature of various quenched many-body quantum systems.

The interesting point is that the circuit complexity measure and entanglement entropy computed from two-mode squeezed states in this paper perfectly goes well with the results obtained in ref. [68]. From the discussion in the previous sections it is quite clear that the expressions for the quantum circuit complexity measure and entanglement entropy measure from covariance matrix method are solely determined by the squeezing amplitude parameter $r_{\mathbf{k}}$ and completely independent of the squeezing angle $\phi_{\mathbf{k}}$. On the other hand, the same results obtained from the Nielsen's wave function method is described by both squeezing amplitude parameter $r_{\mathbf{k}}$ and squeezing angle $\phi_{\mathbf{k}}$. For this reason one can use very easily the results obtained from covariance matrix method to show the underlying connection between circuit complexity measure and entanglement entropy in the present context of discussion.

Using the covariance matrix method the circuit complexity measure from linear and quadratic cost functions can be expressed as:

$$C_1 = 4\, r_{\mathbf{k}} \tag{165}$$

$$C_2 = 2\sqrt{2}\, r_{\mathbf{k}}. \tag{166}$$

On the other hand the expression for the Rényi entropy measure can be given by for different physical situations as:

$$S_\alpha(r_{\mathbf{k}} \to \infty) := \lim_{r_{\mathbf{k}}\to\infty} S_\alpha \approx \frac{2\alpha}{(\alpha-1)}\, r_{\mathbf{k}}, \tag{167}$$

$$S_\infty(r_{\mathbf{k}} \to 0) := \lim_{r_{\mathbf{k}}\to 0} S_\infty \approx\ r_{\mathbf{k}}^2. \tag{168}$$

Now comparing all these results one can easily write the following connecting relationship for the large and small squeezing amplitude parameter $r_{\mathbf{k}}$, which is given by:

$$C_1 = \sqrt{2}\, C_2 = \frac{2(\alpha-1)}{\alpha} S_\alpha(r_{\mathbf{k}} \to \infty), \tag{169}$$

$$C_1 = \sqrt{2}\, C_2 = 4\sqrt{S_\infty(r_{\mathbf{k}} \to 0)}. \tag{170}$$

Here we have derived two types of connecting relationship between the quantum circuit complexity measure and entanglement entropy measure for the large and small values of the squeezing amplitude parameter $r_{\mathbf{k}}$. For $\alpha = 2$ we further get:

$$C_1 = \sqrt{2}\, C_2 = S_2(r_{\mathbf{k}} \to \infty), \tag{171}$$

which means both of them are exactly same in terms of measure. For $\alpha > 2$ we have:

$$C_1 = \sqrt{2}\, C_2 > S_{\alpha>2}(r_{\mathbf{k}} \to \infty), \tag{172}$$

in this context. On the other hand hand, for small squeezing amplitude parameter $r_{\mathbf{k}}$ if we take the $\alpha \to 1$ limit then we get the following expression for the entanglement entropy:

$$S_{\rm vN}(r_{\mathbf{k}} \to 0) := \lim_{r_{\mathbf{k}}\to 0} S_{\alpha\to 1} \approx\ r_{\mathbf{k}}^2\,(1 - 2\ln\ r_{\mathbf{k}})\,. \tag{173}$$

Using this expression further we get the following connecting relationship between all of these measure in the

present context:

$$S_{\rm vN}(r_{\mathbf{k}}\to 0)=\frac{C_1^2}{16}\left[1-2\ln\left(\frac{C_1}{4}\right)\right]$$
$$=\frac{C_2^2}{8}\left[1-2\ln\left(\frac{C_2}{2\sqrt{2}}\right)\right]. \quad (174)$$

For small squeezing amplitude parameter $r_{\mathbf{k}}$ one can write down the following relation:

$$r_{\mathbf{k}}\approx\exp\left(-\frac{\omega_{\mathbf{k}}}{2T}\right)$$
$$=\exp\left(-\frac{\omega_{-\mathbf{k}}}{2T}\right)$$
$$=\exp\left(-\frac{\Omega_{\mathbf{k}}}{4T}\right), \quad (175)$$

where we have:

$$\omega_{\mathbf{k}}=\omega_{-\mathbf{k}}=\Omega_{\mathbf{k}}/2., \quad (176)$$

where the expression for $\Omega_{\mathbf{k}}$ is defined before. Using this expression one can can further write down the expressions for different circuit complexity measure and entanglement measure as:

$$C_1=4\ \exp\left(-\frac{\Omega_{\mathbf{k}}}{4T}\right), \quad (177)$$

$$C_2=2\sqrt{2}\ \exp\left(-\frac{\Omega_{\mathbf{k}}}{4T}\right), \quad (178)$$

$$S(r_{\mathbf{k}}\to 0)=\exp\left(-\frac{\Omega_{\mathbf{k}}}{2T}\right), \quad (179)$$

$$S_{\rm vN}(r_{\mathbf{k}}\to 0)=\left(1+\frac{\Omega_{\mathbf{k}}}{2T}\right)\exp\left(-\frac{\Omega_{\mathbf{k}}}{2T}\right). \quad (180)$$

On the other hand, for large squeezing amplitude parameter $r_{\mathbf{k}}$ one can write down the following relation:

$$r_{\mathbf{k}}=\mathrm{Coth}^{-1}\left[\exp\left(\frac{\Omega_{\mathbf{k}}}{4T}\right)\right]. \quad (181)$$

Using this expression one can can further write down the expressions for different circuit complexity measure and entanglement measure as:

$$C_1=4\ \mathrm{Coth}^{-1}\left[\exp\left(\frac{\Omega_{\mathbf{k}}}{4T}\right)\right], \quad (182)$$

$$C_2=2\sqrt{2}\ \mathrm{Coth}^{-1}\left[\exp\left(\frac{\Omega_{\mathbf{k}}}{4T}\right)\right], \quad (183)$$

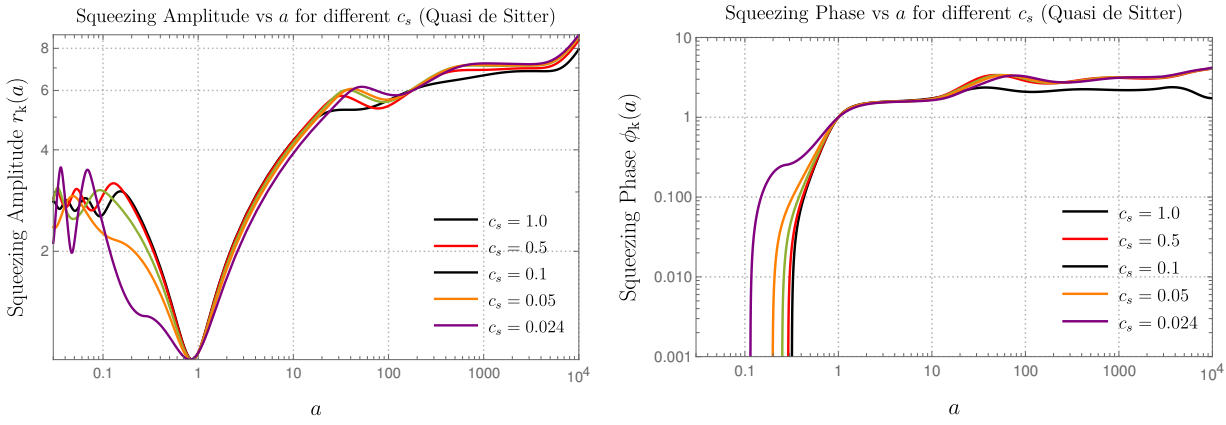
$$S_\alpha(r_{\mathbf{k}}\to\infty)=\frac{2\alpha}{(\alpha-1)}\ \mathrm{Coth}^{-1}\left[\exp\left(\frac{\Omega_{\mathbf{k}}}{4T}\right)\right]. \quad (184)$$

Till now we have used the results of the circuit complexity measure from covariance matrix method to establish this connecting relation analytically. In the next section we are going to use the Nielsen's wave function method to understand such connections numerically.

x. Numerical Results

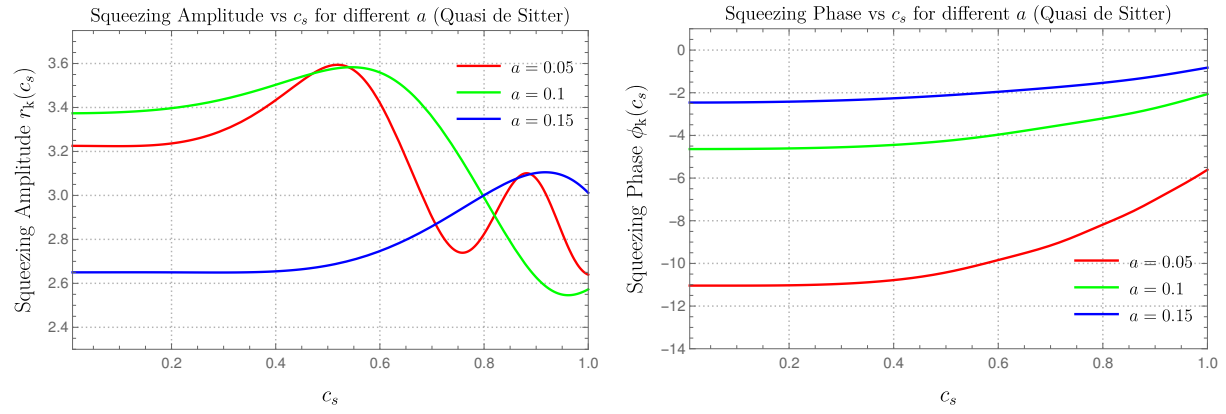
In this section we perform numerical analysis of quantum circuit complexity measure and entanglement entropy measure computed from the present COSMOEFT setup using two-mode squeezed state formalism. As we have mentioned before that instead we study the evolution with respect to the conformal time scale we here perform the numerical analysis using the cosmological scale factor for quasi de Sitter space in $3+1$ dimensions. This is because numerically solving the evolution of the squeezing amplitude $r_{\mathbf{k}}$ and the squeezing angle $\phi_{\mathbf{k}}$ from the previously mentioned coupled differential eqns. (65) and (66) for a given boundary condition is very complicated to solve due to having numerical instabilities with respect to time scale. Here we fix the boundary condition at the late time scale $\tau=\tau_0$ which fix the squeezing amplitude and the squeezing angle at $r_{\mathbf{k}}(\tau_0)=1$ and $\phi_{\mathbf{k}}(\tau_0)=1$. On the other hand, numerically solving the evolution of the squeezing amplitude $r_{\mathbf{k}}$ and the squeezing angle $\phi_{\mathbf{k}}$ with respect to scale factor as stated in eqns. (69) and (70) comparatively simpler for a given boundary condition. Here the boundary condition is fixed here at the late time scale $\tau=\tau_0$ where $a(\tau_0)=a_0=1$ which fix the squeezing amplitude and the squeezing angle at $r_{\mathbf{k}}(a_0=1)=1$ and $\phi_{\mathbf{k}}(a_0=1)=1$. This boundary condition help us to know the cosmological evolution of squeezing amplitude $r_{\mathbf{k}}(a)$ and squeezing angle $\phi_{\mathbf{k}}(a)$ at any value of the scale factor. Solving these coupled differential equations with respect to the cosmological scale factor comparatively simpler than solving with respect to the conformal time scale as the numerical instabilities are less and capturing more information in the cosmological scale. To solve the problem with respect to the cosmological scale factor we have used change of variable from the the conformal time scale to the scale factor by using eqn. (68).

In figure (1(a)) and figure (1(b)), we have plotted the behaviour of the squeezing amplitude parameter $r_{\mathbf{k}}$ and the squeezing angle parameter $\phi_{\mathbf{k}}$ with respect to the dynamical cosmological scale factor a of the quasi de Sitter space from our present COSMOEFT setup. In these plots we have considered different values of the scale factors lying within the window $0.024\leq c_s\leq 1$. Here $c_s=1$ corresponds to the single scalar field canonical models and rest of the values describe a wide class of non-canonical scalar fields models in the present framework. From figure (1(a)) now it is possible to get the information about the behaviour of the squeezing amplitude parameter $r_{\mathbf{k}}$ at different cosmological scales ranging from very small to large values of a, which describe past, present and future of our quasi de Sitter universe. For very early universe (extremely small values of a) squeezing amplitude parameter $r_{\mathbf{k}}$ shows irregular oscillatory behaviour and the amplitude of oscillators larger for smaller values of effective sound speed c_s and for $c_s=1$ the amplitude becomes smaller compared to all other oscillation for different sound speeds c_s. No smaller am-



(a) $r_{\mathbf{k}}$ vs a (b) $\phi_{\mathbf{k}}$ vs a

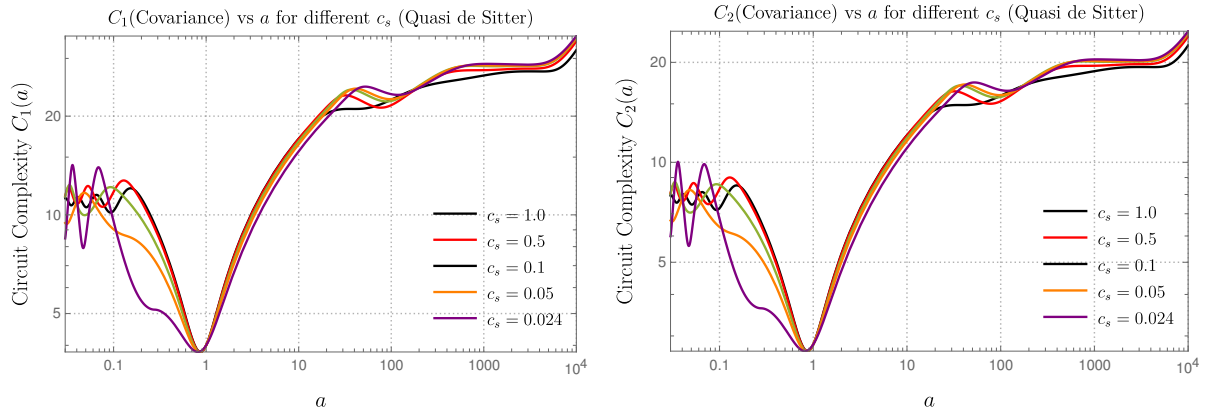
FIG. 1. Behaviour of the squeezing amplitude ($r_{\mathbf{k}}$) and squeezing angle/phase ($\phi_{\mathbf{k}}$) with respect to cosmological scale factor a for the different values of the effective sound speed parameter c_s lying within the window $0.024 \leq c_s \leq 1$.



(a) $r_{\mathbf{k}}$ vs c_s (b) $\phi_{\mathbf{k}}$ vs c_s

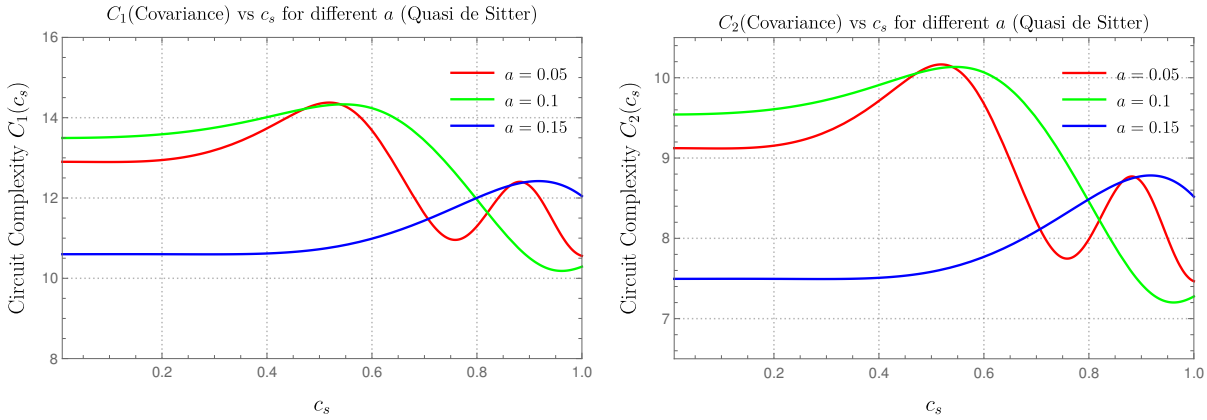
FIG. 2. Behaviour of the squeezing amplitude ($r_{\mathbf{k}}$) and squeezing angle/phase ($\phi_{\mathbf{k}}$) with respect to the effective sound speed parameter c_s for the different values of the cosmological scale factor a fixed at early time scale of our universe.

plitude is allowed due to the restriction from causality which tells us that $c_s \leq 1$ from the COSMOEFT theories that we have considered in this paper. For the value of the scale factor lying within the range $0.1 < a < 1$ squeezing amplitude parameter $r_{\mathbf{k}}$ decrease considerably for different values of c_s. However, it is clearly visible from this plot that in this region the behaviour of $r_{\mathbf{k}}$ is clearly distinguishable for the different values of c_s. Particularly the features show sufficient distinguishablity for smaller values of the effective sound speed, such as $c_s = 0.024$ and $c_s = 0.05$. Then at the present time scale where the corresponding value of the scale factor $a = 1$ the squeezing amplitude parameter $r_{\mathbf{k}}$ drawn for different values of c_s reach at coincident minimum value which small but no-zero. As we have pointed before we fix the boundary condition at this point which can be also visible from this plot clearly. After crossing this point within the range $1 < a < 10^4$ which corresponds to the future universe the squeezing amplitude parameter $r_{\mathbf{k}}$ sufficiently increase by following almost similar behaviour for different values of c_s, which are less distinguishable. Though in between $30 < a < 10^4$ some small distinguishable features can be clearly observed from the figure (1(a)). This behavioural study of the squeezing amplitude parameter $r_{\mathbf{k}}$ is extremely useful with respect to the dynamical scale a, which tells us that in which region of study the effect of the sound speed parameter c_s is quite prominent using which one can make distinction between various types of theories which is actually described by a single COSMOEFT setup. From figure (1(b)) the behaviour of the squeezing amplitude parameter $\phi_{\mathbf{k}}$ at different cosmological scales ranging from very small to large values of a, one can now able to interpret the underlying physics. For very early universe (extremely small values of a) squeezing amplitude angle $\phi_{\mathbf{k}}$ shows sharp and faster increment in the behaviour and



(a) C_1 vs a (b) C_2 vs a

FIG. 3. Behaviour of the circuit complexity from linear cost function (C_1) and quadratic cost function (C_2) with respect to cosmological scale factor a for the different values of the effective sound speed parameter c_s lying within the window $0.024 \leq c_s \leq 1$. Here we have used covariance matrix method.

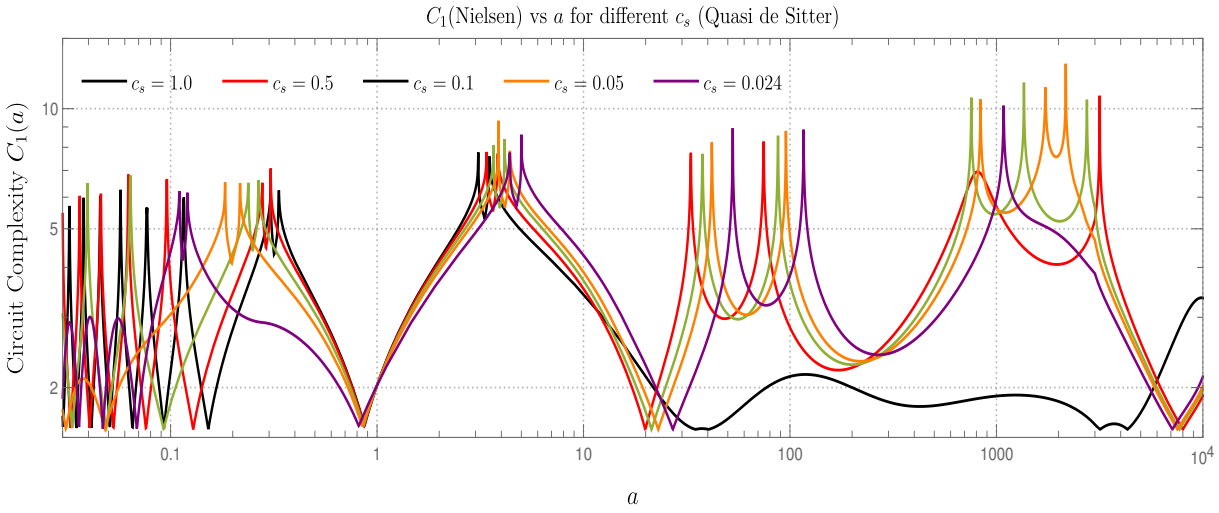


(a) C_1 vs c_s (b) C_2 vs c_s

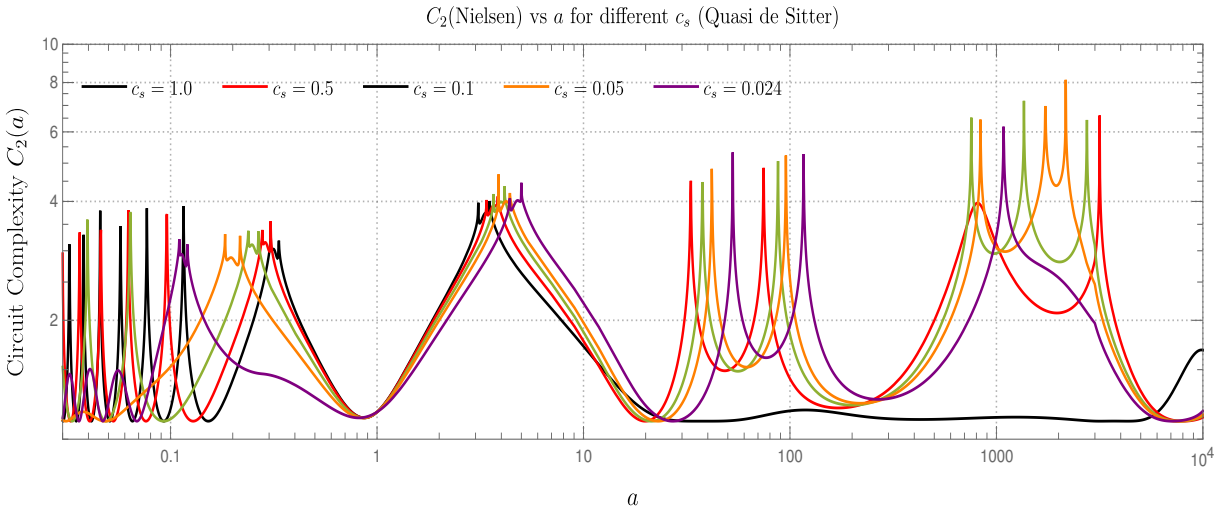
FIG. 4. Behaviour of the circuit complexity from linear cost function (C_1) and quadratic cost function (C_2) with respect to the effective sound speed parameter c_s for the different values of the cosmological scale factor a fixed at early time scale of our universe. Here we have used covariance matrix method.

each of increment is quite distinguishable for all values of the sound speed c_s lying within the previously mentioned window. Distinguishability become more prominent for the smaller values of c_s compared to larger values of c_s. This feature is observed within the range of the scale factor $0.1 < a < 1$. At exactly $a = 1$ all of the behaviour observed for different values of c_s suddenly become indistinguishable. In between $1 < a < 30$ it shows again the incremental behaviour but as we have said in this region all the effects of the sound speed become subdominant. After this, within the range $30 < a < 10^4$ some small distinguishable features of the squeezing amplitude parameter $\phi_{\mathbf{k}}$ can be observed from plot. This behavioural study is extremely useful with respect the dynamical scale a, which tells us that in which region of study the effect of the sound speed parameter c_s is quite prominent. Further using these features one can able to make distinction between various types of theories which is actually described by a single COSMOEFT setup and able to physically constraint some of them.

In figure (2(a)) and figure (2(b)), we have plotted the behaviour of the squeezing amplitude parameter $r_{\mathbf{k}}$ and the squeezing angle parameter $\phi_{\mathbf{k}}$ with respect to the effective sound speed parameter c_s lying within the range $0.024 \leq c_s \leq 1$, which is restricted by cosmological observation from Planck and causality constraint. Here we have plotted the behaviour for three different values of the scale factor a which is fixed at the early time scale in the cosmological evolutionary scale. From figure (2(a)) the dynamical behaviour of the squeezing amplitude parameter $r_{\mathbf{k}}$ is depicted for smaller and larger values of the sound speed parameter c_s. For the smaller values of c_s we



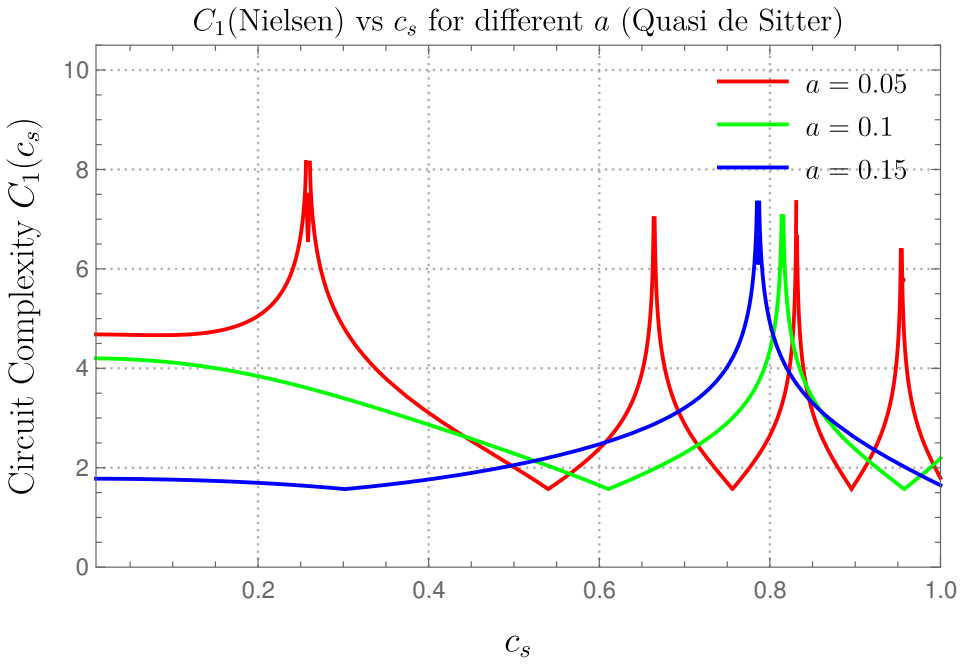
(a) C_1 vs a



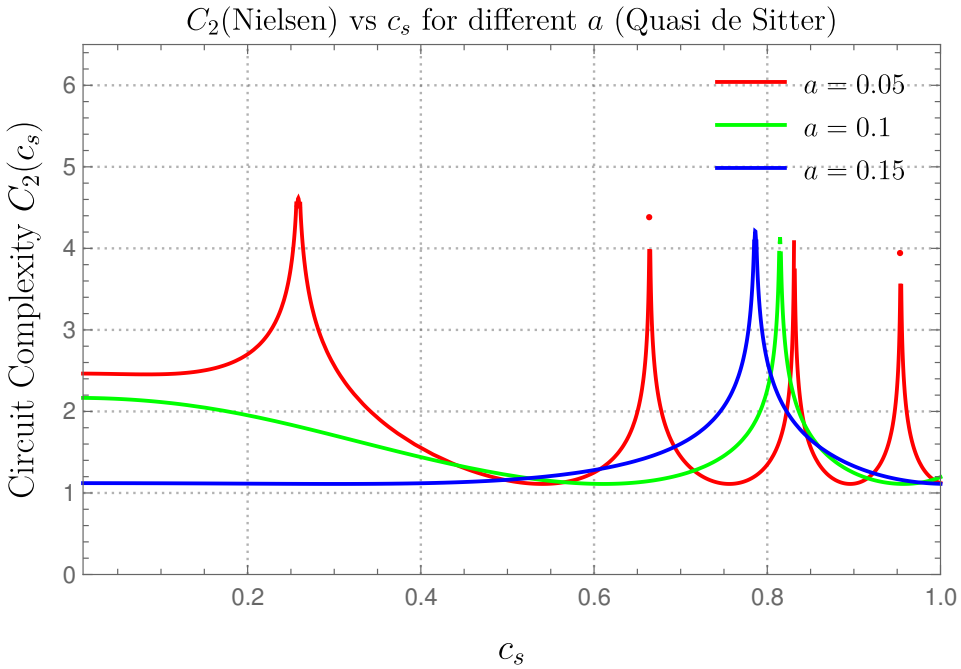
(b) C_2 vs a

FIG. 5. Behaviour of the circuit complexity from linear cost function (C_1) and quadratic cost function (C_2) with respect to cosmological scale factor a for the different values of the effective sound speed parameter c_s lying within the window $0.024 \leq c_s \leq 1$. Here we have used Nielsen's wave-function method.

can clearly observe that the squeezing amplitude parameter $r_{\mathbf{k}}$ is constant and for the large value of c_s it shows significant feature, initially it will increase to maximum value and then it will decrease. Due to having causality constraint we have not plotted the behaviour more than the value of the sound speed parameter $c_s = 1$. Now one more interesting feature we need to point here that for the large values of the fixed value of the scale factor we observe only one peak in this plot. For the smaller values of the scale factor (such as $a = 0.05$) we can observe two peaks, though the amplitude of the peaks are not same. Here one of them have very larger amplitude and the other has lesser amplitude. Such features are very interesting for our study in this paper, as it gives a lot of unknown information regarding the system under study. Also another important feature we can observe from the plot, which tells us that there are phase lagging between the plots for the different values of the scale factor a. From figure (2(b)) the dynamical behaviour of the squeezing angle parameter $\phi_{\mathbf{k}}$ is depicted for smaller and larger values of the sound speed parameter c_s. For the smaller values of c_s we can clearly observe that the squeezing angle parameter $\phi_{\mathbf{k}}$ is constant and for the large value of c_s it shows significant feature, it will increase as we increase c_s. It is important to note that within the range $0.024 \leq c_s \leq 1$ the squeezing angle parameter $\phi_{\mathbf{k}}$ become always negative. The maximum saturation amplitude of the negative value of $\phi_{\mathbf{k}}$ are quite



(a) C_1 vs c_s

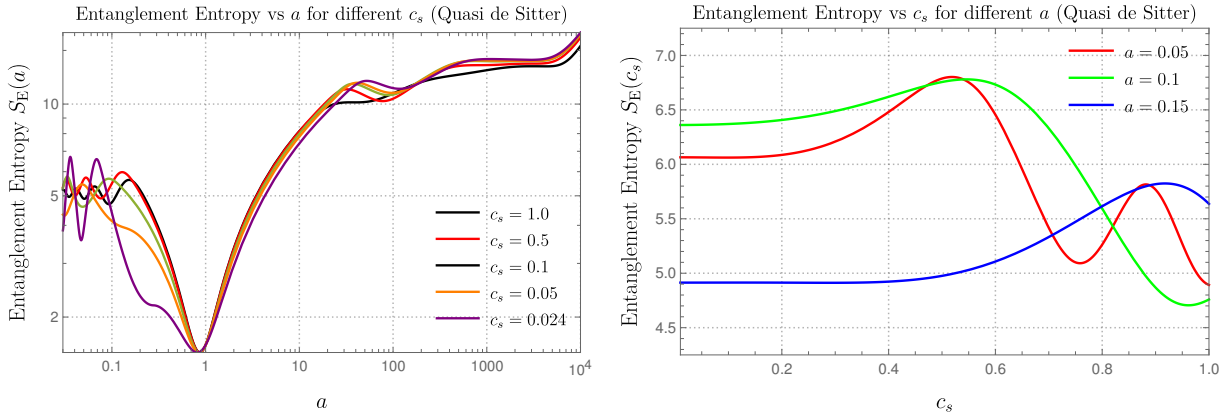


(b) C_2 vs c_s

FIG. 6. Behaviour of the circuit complexity from linear cost function (C_1) and quadratic cost function (C_2) with respect to the effective sound speed parameter c_s for the different values of the cosmological scale factor a fixed at early time scale of our universe. Here we have used Nielsen's wave-function method.

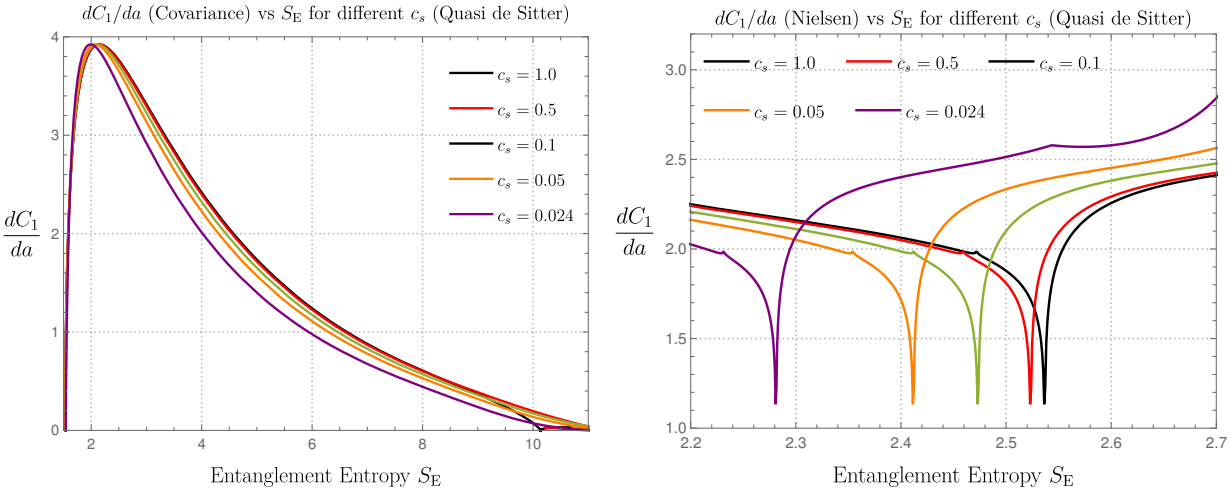
distinguishable for all fixed values of the scale factor.

In figure (3(a)) and (3(b)), we have plotted the behaviour of the circuit complexity from linear C_1 and quadratic cost function C_2 computed using the covariance matrix method with respect to the dynamical cosmological scale factor a of the quasi de Sitter space from our present COSMOEFT setup. In these plots we have considered different values of the scale factors lying



(a) S_E vs a (b) S_E vs c_s

FIG. 7. Behaviour of the entanglement entropy (S_E) with respect to cosmological scale factor a and effective sound speed c_s respectively.



(a) dC_1/da vs a (b) dC_1/da vs c_s

FIG. 8. Behaviour of the rate of change of complexity from linear cost function measure with respect to entanglement entropy (S_E) computed from the covariance matrix method and Nielsen's wave-function method for the different effective sound speed c_s respectively.

within the window $0.024 \leq c_s \leq 1$. The behaviour of both the measure for circuit complexity with respect the scale factor a is exactly similar to the squeezing amplitude parameter $r_{\mathbf{k}}$ as depicted in figure (1(a)). This is because of the fact that $C_1 = \sqrt{2}\ C_2 = 4\ r_{\mathbf{k}}$. Only the difference will be observed in the overall amplitude of the measure of quantum circuit complexity. Also both of the measure gives similar feature in the present context. The physical explanation that we have provided for the figure (1(a)) also holds perfectly well in the present context of discussion.

In figure (4(a)) and figure (4(b)), we have plotted the behaviour of the circuit complexity from linear C_1 and quadratic cost function C_2 computed using the covariance matrix method with respect to the effective sound speed parameter c_s lying within the range $0.024 \leq c_s \leq 1$, which is restricted by cosmological observation from Planck and causality constraint. The behaviour of both the measure for circuit complexity with respect the sound speed c_s is exactly similar to the squeezing amplitude parameter $r_{\mathbf{k}}$ as depicted in figure (2(a)). Since $C_1 = \sqrt{2}\ C_2 = 4\ r_{\mathbf{k}}$, only the difference will be observed in the overall amplitude for both of the plots. The physical explanation that we have given for the figure (2(a)) also holds perfectly well in this context.

In figure (5(a)) and (5(b)), we have plotted the behaviour of the circuit complexity from linear C_1 and quadratic cost function C_2 computed using the Nielsen's wave-function method with respect to the dynamical cosmological scale factor a of the quasi de Sitter space

from our present COSMOEFT setup. In these plots we have considered different values of the scale factors lying within the window $0.024 \leq c_s \leq 1$. For the smaller values of the scale factor (for $a < 1$) the behaviour of both the measure C_1 and C_2 computed using the Nielsen's wave-function method shows similar behaviour. Only the differences found in the smoothness at the transition points where the behaviour changes from increasing to decreasing or decreasing to increasing behaviour. For the smaller values of a the frequency of the increasing and decreasing behaviour is very high for both of the measures and it is hard to distinguish the individual effects appearing from the different values of the sound speed c_s. From both of these plots it is observed that at the point $a = 1$ all the behaviour obtained from different values of c_s coincide with each other. This is due to the fact that, we have fixed the boundary conditions at this point. As we approach for the larger values of scale factor a in both of these plots we see that the features obtained from the different values of the sound speed parameter c_s become more and more distinguishable. Another more important feature is observed in both of these plots that within the range $20 < a < 10^4$ for the sound speed value $c_s = 0.024$ we get quite impressive distinguishable effects. Also the rising and falling behaviour in both of these plots appearing in an aperiodic fashion. These distinguishable features are extremely helpful to probe the models having $c_s = 1$ and $c_s < 1$.

In figure (6(a)) and figure (6(b)), we have plotted the behaviour of the circuit complexity from linear C_1 and quadratic cost function C_2 computed using the Nielsen's wave-function method with respect to the effective sound speed parameter c_s lying within the range $0.024 \leq c_s \leq 1$, which is restricted by cosmological observation from Planck and causality constraint. For the smaller values of the sound speed the behaviour of both the measure C_1 and C_2 computed using the Nielsen's wave-function method shows similar behaviour. Only the differences can be observed in terms of the smoothness of the plots at the transition points where the circuit complexity from both of the measure changes its behaviour from increasing to decreasing or from decreasing to increasing feature. As we change the value of the sound speed parameter we observe multiple peaks appearing in irregular interval having different smoothness for C_1 and C_2. For the larger values of the sound speed these multiple peaks appear very frequently and from the phase shifts between the plots obtained for the different values of the scale factor which we have fixed at the early scale of the evolutionary time scale of our universe.

In figure (7(a)) we have plotted the behaviour of entanglement entropy with respect to the dynamical cosmological scale factor for the different values of the effective sound speed c_s. On the other hand, in figure (7(b)) we have plotted the behaviour of entanglement entropy with respect to the effective sound speed c_s for fixed values of the cosmological scale factor which we fixed at early time scale. Here the entanglement entropy is evaluated from the von-Neumann measure. The behaviour of entanglement entropy is exactly similar to that of the squeezing amplitude parameter $\phi_{\mathbf{k}}$, circuit complexity measure (C_1, C_2) computed from covariance matrix method in the present context. Only the difference that we found that for both of these plots the amplitude of the entanglement entropy computed from von-Neumann measure larger than $r_{\mathbf{k}}$ but also smaller than both circuit complexities, C_1 and C_2 computed from covariance matrix method. From the analytical expression for the entanglement entropy from the von-Neumann measure its extremely difficult to understand that why it is behaving like similar fashion to two complexity measure computed from covariance matrix method. The possible explanation lies within the dependency of both of them on the squeezing amplitude parameter $r_{\mathbf{k}}$. Though from the analytical expressions it seems like that entanglement entropy from von-Neumann measure and two measures of circuit complexity computed from covariance matrix method behave completely different way, but the numerical study from the present setup tells us that the overall features of both of them are exactly similar apart from having different amplitudes.

Finally, in figure (8(a)) and figure (8(b)) we have shown the behaviour of the rate of change of complexity function from linear cost function with respect to entanglement entropy from the COSMOEFT setup that we have constructed in this paper. Here circuit complexity and entanglement entropy is computed using (1) covariance matrix method and von-Neumann measure for figure (8(a)) and (2) Nielsen's wave-function method and von-Neumann measure for figure (8(b)). Both of them are the parametric plots where we have treated the effective sound speed c_s as a parameter i.e. we have plotted the behaviour with respect to different sound speeds which is lying within the window $0.024 \leq c_s \leq 1$. In the present context both of these plots are important because it will help us to know for our system the underlying connection between quantum circuit complexity and entanglement entropy, which is impossible to guess complexity from liner a from the first hand without performing any numerical calculation. First of all we need to point here that computing the rate of change of the circuit complexity function with respect to the scale factor is extremely difficult as we have seen from figure (3(a)) and figure (5(a)) that the behaviour of the circuit complexity itself with respect the scale factor is not at all simple. Due to having multiple increment and decrement in both of this plots computing the derivatives with respect to the scale factor at the transition point is numerically challenging. Instead of looking for a large range of the scale factor for this computation we have concentrated on a specific range for each plots within which the numerical instabilities are very less. For figure (8(a)) we have considered the range $0.3 < a < 100$ and for the figure (8(b)) we have taken $0.3 < a < 2$. From the mentioned range it is also clear that for the plot using covariance matrix method one can cover a larger interval of a compared to the plot

obtained using Nielsen's wave-function method. Particularly in this context it seems like that the covariance matrix method gives the better result compared to the Nielsen's wave-function method as numerical instabilities can be handled easily for the first one. Though we would rather say it is exactly opposite and this can be justified from a very crucial observation from the analytically derived expressions for the quantum circuit complexity from linear and quadratic cost function using both the methods. The result obtained from the covariance matrix method one captures the contribution of the squeezing amplitude parameter $r_{\mathbf{k}}$. On the other hand, the result obtained from Nielsen's wave-function method capture both the information from squeezing amplitude parameter $r_{\mathbf{k}}$ and squeezing angle/phase parameter $\phi_{\mathbf{k}}$. Due to having the contribution from this phase factor in the later one aperiodic oscillations, frequency of transitions from increment to decrement or vice verse, numerical instabilities at the transition point is larger. But if we find a region which is our point of interest, this means a region where the circuit complexity function only grow with the scale factor and not encounters any transition, then using Nielsen's wave-function method one can able to compute the derivative of the circuit complexity function with respect to the scale factor, which in the present context is identified as the growth rate of the circuit complexity. Now from figure (8(a)) it is clearly observed that initially within a very small value of the entanglement entropy the rate of change of the circuit complexity with respect to the scale factor sharply increase to a maximum value where the individual effects from different values of the effective sound speed c_s is not at all distinguishable. After that it shows a decaying feature within a wide range of entanglement entropy where one can explicitly identify the distinguishable effects from different values of the sound speed c_s. On the other hand from the figure (8(b)) throughout all the range of the entanglement entropy one can able to distinctly observe the individual effects from the sound speed c_s. Here one important feature one can also observe that as the sound speed decrease then each of the features obtained for different values of c_s shifted towards left. Also in each plots in this case we see initially a sharp decreasing feature and then after reaching a particular value transition to a incremental behaviour. This feature is observed for all values of c_s. Due to having intermediate features from figure (8(a)) and figure (8(b)) it is expected to have a non linear relationship between circuit complexity and entanglement entropy computed from both covariance matrix method and Nielsen's wave-function method. The prime motivation of finding a relationship between the circuit complexity and entanglement entropy lies in the conjecture proposed by Susskind in the context of black hole physics. In the introduction of this paper we have already pointed that the rate of growth rate of complexity is equal to the product of entropy of the black hole and the Hawking temperature of the black hole, which was proposed by Susskind a long time ago. In black hole physics the growth rate complexity is computed with respect to the underlying time scale. However, in the present context instead of computing this rate with respect to the cosmological time scale we have used the cosmological scale factor in terms of which we have studied the evolution of the complexity function. For the small value of the squeezing amplitude parameter $r_{\mathbf{k}} \to 0$ from the covariance matrix method one can write the following non-linear relations among equilibrium temperature, entanglement entropy and circuit complexity measure:

$$\underline{\text{Results for small } r_{\mathbf{k}}:}$$

$$C_1 = \sqrt{2}\, C_2 = 4\, \exp\left(-\frac{\Omega_{\mathbf{k}}}{4T}\right), \tag{185}$$

$$\begin{aligned} S &= \left(1+\frac{\Omega_{\mathbf{k}}}{2T}\right)\, \exp\left(-\frac{\Omega_{\mathbf{k}}}{2T}\right) \\ &= \frac{C_1^2}{16}\left[1-2\ln\left(\frac{C_1}{4}\right)\right] \\ &= \frac{C_2^2}{8}\left[1-2\ln\left(\frac{C_2}{2\sqrt{2}}\right)\right]. \end{aligned} \tag{186}$$

On the other hand, for the large value of the squeezing amplitude parameter $r_{\mathbf{k}} \to \infty$ from the covariance matrix method one can write the following non-linear relations among equilibrium temperature, entanglement entropy and circuit complexity measure:

$$\underline{\text{Results for large } r_{\mathbf{k}}:}$$

$$C_1 = \sqrt{2}\, C_2 = 4\, \text{Coth}^{-1}\left[\exp\left(\frac{\Omega_{\mathbf{k}}}{4T}\right)\right], \tag{187}$$

$$\begin{aligned} S_\alpha &= \frac{2\alpha}{(\alpha-1)}\, \text{Coth}^{-1}\left[\exp\left(\frac{\Omega_{\mathbf{k}}}{4T}\right)\right] \\ &= \frac{\alpha}{2(\alpha-1)}\, C_1 \\ &= \frac{\alpha}{\sqrt{2}(\alpha-1)}\, C_2. \end{aligned} \tag{188}$$

These relations written for the small and large values of squeezing amplitude parameter are extremely helpful to analytically understand the underlying connection between equilibrium temperature of the system, quantum circuit complexity and entanglement entropy measure for our EFT setup constructed in this paper.

XI. Covariance Vs Nielsen's method in COSMOEFT

In this section our objective is to discuss a comparison between the results obtained from covariance matrix method and the Nielsen's wave function method in the context of the present COSMOEFT setup. These comparisons are appended below point-wise:

- The results obtained in covariance method not depend on the squeezing angle/phase $\phi_{\mathbf{k}}$ and for Nielsen's method it will explicitly depend on this parameter.

- In the case of covariance method circuit complexity measure linearly depend on the squeezing amplitude parameter $r_{\mathbf{k}}$. On the other hand, in Nielsen's method circuit complexity measure non-linearly depend on the squeezing amplitude parameter $r_{\mathbf{k}}$.
- For covariance method the explicit details of the wave function of the squeezed state is not appearing at the final expression for the circuit complexity measure. On the other hand, for Nielsen's method circuit complexity measure explicitly depend on the details of the wave function. After detailed calculation it turns out that the circuit complexity in Nielsen's method is parametrized by both squeezing parameters $r_{\mathbf{k}}$ and $\phi_{\mathbf{k}}$.
- In covariance method circuit complexities computed from the linear and quadratic measure are approximately same as $C_1 = \sqrt{2}\, C_2$, which is very simple. On the other hand in Nielsen's method circuit complexities are related via the relation, $C_1^2 - C_2^2 = f(r_{\mathbf{k}}, \phi_{\mathbf{k}})$, where in general $f(r_{\mathbf{k}}, \phi_{\mathbf{k}})$ are the non-linear functions of both squeezing parameters $r_{\mathbf{k}}$ and $\phi_{\mathbf{k}}$.
- The features obtained from the computed circuit complexities from covariance and Nielsen's method have been performed numerically in the previous section. This study shows that in the case of covariance method the circuit complexity exactly behaving like the squeezing amplitude $r_{\mathbf{k}}$ apart from having a small change in amplitude. On the other hand, just by seeing the behaviour of the squeezed parameters $r_{\mathbf{k}}$ and $\phi_{\mathbf{k}}$ it is almost impossible to guess the behaviour of the circuit complexity computed from Nielsen's method. Frequency of oscillation of the aperiodic behaviour in the circuit complexity computed from both the methods are large initially and the individual effects for the different sound speed c_s is not clearly distinguishable. On the other hand, it turns out that that this frequency will decrease a lot and one can able to distinguish the effects of the different sound speed contribution explicitly from the behaviour at larger value of the scale factor. However, in the case of covariance method we get growing behaviour in the plot and also the aperiodic feature is complexity lost.
- Causality constraint on the circuit complexity computed from both of these methods show different behaviour in terms of smoothness of each of the plots at the transition point. Results obtained from the covariance method show high smoothness. On the other hand, the smoothness at the transition points are less and it is showing very sharp feature for the results obtained from Nielsen's method. Apart from that, the outcome of both the methods show phase shifts of each plot obtained for different values of the scale factor in a similar fashion. Such constraint restrict us to only consider the region $c_s \leq 1$ and from the cosmological observation from Planck we have fixed the lower limit at $c_s \geq 0.024$.
- Entanglement entropy is solely determined by the squeezing amplitude $r_{\mathbf{k}}$ in the present context. Since the computed results of the circuit complexity measure from covariance method also solely determined in terms of the squeezing amplitude $r_{\mathbf{k}}$, very easily one can able to develop a connecting relationship between the circuit complexity measure and entanglement entropy measure. On the other hand, the computed results of the circuit complexity measure from Nielsen's method parametrized by both the squeezing parameters $r_{\mathbf{k}}$ and $\phi_{\mathbf{k}}$, it is impossible to find an analytical relationship between the circuit complexity measure and entanglement entropy measure in this case.
- The rate of change of the circuit complexity measure with respect to the scale factor vary smoothly with respect to the entanglement entropy for different values of the sound speed c_s in the case of covariance method. The individual effects of the different sound speeds are not visible for the lower values of the entanglement entropy, and such effects become prominent for the large values of the entanglement entropy. On the other hand, the rate of change of the circuit complexity measure with respect to the scale factor vary sharply at the transition point with respect to the entanglement entropy for different values of the sound speed c_s in the case of Nielsen's method. The individual effects of the different sound speeds are clearly visible for the all values of the entanglement entropy.
- For the large and small values of the squeezed amplitude parameter $r_{\mathbf{k}}$ one can able to determine analytically the connecting relationship among equilibrium temperature, entanglement entropy measure and quantum circuit complexity measure using the covariance matrix method. On the other hand establishing such analytical connection is almost impossible analytically using the Nielsen's eave-function method. Only it is expected that, it would be highly complicated due to having non-linear dependencies on the both the squeezing parameters $r_{\mathbf{k}}$ and $\phi_{\mathbf{k}}$ in the expressions for two measures of quantum circuit complexity. Otherwise, the entanglement entropy and the equilibrium temperature solely parametrized by the squeezing amplitude parameter $r_{\mathbf{k}}$.

XII. Conclusion

The concluding remarks of this work are appended below point-wise:

- First we have developed an effective field theory framework of cosmological perturbation, which we

identified as COSMOEFT using Stückelberg trick which actually generates the scalar modes from Goldstone modes in this framework. In this formalism we have restricted upto the two derivative terms in the metric. Here the Goldstone modes are generated from the broken time diffeomorphism symmetry. This EFT framework at the level of cosmological perturbation is described by an effective fluid whose sound speed is c_s. Also the causality constraint demands that for such an effective fluid the sound speed has to satisfy $c_s \leq 1$. On the other hand, from the cosmological observation from Planck we have restricted the lower bound of the effective sound speed at $c_s \geq 0.024$. For this reason in our paper we have considered only the range $0.024 \leq c_s \leq 1$ for rest of the analysis. In this analysis, $c_s = 1$ corresponds to the single scalar field canonical models and $c_s < 1$ signify a class of non-canonical scalar field models.

- Next we have computed the quantized Hamiltonian of this EFT setup and found that the interacting part of the Hamiltonian can be easily parametrized by two parameters, amplitude $r_{\mathbf{k}}$ and phase $\phi_{\mathbf{k}}$ in the framework of squeezed state formalism. We have used this framework as the underlying working physical principle in this paper.

- We have computed the expression for the circuit complexity from two different measures- using linear and quadratic cost function from covariance matrix method and Nielsen's wave function method using two-mode squeezed state formalism. Also we have computed the entanglement entropy from this setup.

- From our analysis we have found that the results obtained by Nielsen's wave function method provides better understanding of quantum circuit complexity compared to the results obtained from covariance matrix method. Nielsen's wave function method capturing more information than covariance matrix method. This is because of the fact that circuit complexity in Nielsen's wave function method is parametrized in terms of both the squeezing parameters $r_{\mathbf{k}}$ and phase $\phi_{\mathbf{k}}$. On the other hand, in the case of covariance matrix method the result is parametrized in terms of squeezing amplitude $r_{\mathbf{k}}$.

- For the large and small values of squeezing amplitude $r_{\mathbf{k}}$ we have derived a connecting analytical relationship among different measures of quantum circuit complexity, entanglement entropy and equilibrium temperature using the results obtained from covariance matrix method. We have found non-trivial non-linear temperature dependence in this case. Though using the results obtained from Nielsen's wave function method we are incapable to make it possible due to having non-linear dependencies of both the squeezing parameters on the final expression for the circuit complexity computed from both linear and quadratic cost function measure.

- Though in support of our analysis to establish some connection among different measures of quantum circuit complexity, entanglement entropy and equilibrium temperature using the Nielsen's wave function method we have computed the rate of change of the circuit complexity function with respect to the scale factor numerically and plotted with respect to entanglement entropy in a wide range.

- The present analysis helps us to know the underlying features of the present EFT setup which describes a wide class of canonical and non-canonical scalar field models. Using the usual cosmological analysis, such as studying the cosmological correlation functions, amplitudes of the spectrum of the correlators, its scale dependent or invariant features sometimes it is very hard to distinguish various models available in this literature. The main reason all of these cosmological sometimes provide degenerate predictions from the cosmological observations. To break this degeneracy the higher point correlators has to be constrained from observation, which is not yet quite efficiently possible due to not having appropriate statistical accuracy in these measurements and also not having possible resources to do this job. Sidewise, one can look for some additional physics and try to probe them with the available tools. If one can make it possible then one can break such degeneracies and rule out several available models. Here in this work our motivation is to provide such theoretical tools, particularly the quantum information theoretic tools which helps us to explore some unknown facts from the present setup and helps us to distinguish the effects for the different values of the sound speed parameter c_s.

Future Prospects:

- In the present paper, we have analyzed the causality constraints on circuit complexity in the COSMOEFT framework and we have restricted the initial quantum vacuum state as the well known Bunch Davies or Euclidean state but one can also study the Causality constraint by considering other choices of initial quantum vacuum state such as $\alpha-$vacuum and $\alpha-\gamma$ vacuum (also called Motta-Allen vacuum).

- Also in the present context, our study was mainly focused on the analysis of causality constraint on circuit complexity with background space as de sitter space but it would be interesting to study other kinds of models having different scale factors such

as Radiation, Matter, Inflation, Bouncing, Cyclic and Blackhole gas model etc, because each of the scale factors carries some information of significant physics of the early universe.

- In this article we have done the analysis in a model independent fashion. But using the explicit details from the models like DBI [96, 117–120], Tachyon [96, 119, 120], Galileon [121, 122] type of theories one can explore the similar type of analysis in a model independent fashion. This type of analysis will help us to know about the role of UV completion in the context of circuit complexity.

- Circuit Complexity has been studied as deformations in the euclidean path integrals for CFT's. Similar deformations also appear in the context of cosmological perturbations one could also study these path integral optimization in de Sitter space.

- Recently in [123–125] Circuit Complexity has been explored in continuous variable system and open quantum systems, similarly techniques and analysis can be done with study of Circuit Complexity in context of cosmological perturbations.

Acknowledgement: SC would like to thank Institute of Physics, Bhubaneswar for supporting with Visiting Scientist position. SC and AM would also like to thank School of Physical Sciences, National Institute for Science Education and Research (NISER), Bhubaneswar where part of the work was done. SC also thank all the members of our newly formed virtual international non-profit consortium Quantum Structures of the Space-Time & Matter (QASTM) for elaborative discussions. NP and AR would like to thank the members of the QASTM Forum for useful discussions. Last but not least, we would like to acknowledge our debt to the people belonging to the various part of the world for their generous and steady support for research in natural sciences.

[1] S. Chapman and G. Policastro, "Quantum Computational Complexity – From Quantum Information to Black Holes and Back," arXiv:2110.14672 [hep-th].
[2] S. Chapman, M. P. Heller, H. Marrochio, and F. Pastawski, "Toward a Definition of Complexity for Quantum Field Theory States," *Phys. Rev. Lett.* **120** no. 12, (2018) 121602, arXiv:1707.08582 [hep-th].
[3] S. Chapman, H. Marrochio, and R. C. Myers, "Holographic complexity in Vaidya spacetimes. Part I," *JHEP* **06** (2018) 046, arXiv:1804.07410 [hep-th].
[4] S. Chapman, H. Marrochio, and R. C. Myers, "Holographic complexity in Vaidya spacetimes. Part II," *JHEP* **06** (2018) 114, arXiv:1805.07262 [hep-th].
[5] P. A. Cano, R. A. Hennigar, and H. Marrochio, "Complexity Growth Rate in Lovelock Gravity," *Phys. Rev. Lett.* **121** no. 12, (2018) 121602, arXiv:1803.02795 [hep-th].
[6] J. L. Barbon and J. Martin-Garcia, "Terminal Holographic Complexity," *JHEP* **06** (2018) 132, arXiv:1805.05291 [hep-th].
[7] M. Flory and N. Miekley, "Complexity change under conformal transformations in AdS_3/CFT_2," *JHEP* **05** (2019) 003, arXiv:1806.08376 [hep-th].
[8] S. Chapman, D. Ge, and G. Policastro, "Holographic Complexity for Defects Distinguishes Action from Volume," *JHEP* **05** (2019) 049, arXiv:1811.12549 [hep-th].
[9] C. A. Agón, M. Headrick, and B. Swingle, "Subsystem Complexity and Holography," *JHEP* **02** (2019) 145, arXiv:1804.01561 [hep-th].
[10] K. Goto, H. Marrochio, R. C. Myers, L. Queimada, and B. Yoshida, "Holographic Complexity Equals Which Action?," *JHEP* **02** (2019) 160, arXiv:1901.00014 [hep-th].
[11] M. Guo, J. Hernandez, R. C. Myers, and S.-M. Ruan, "Circuit complexity for coherent states," *Journal of High Energy Physics* **2018** no. 10, (Oct, 2018) . http://dx.doi.org/10.1007/JHEP10(2018)011.
[12] A. Bernamonti, F. Galli, J. Hernandez, R. C. Myers, S.-M. Ruan, and J. Simón, "First law of holographic complexity," *Physical Review Letters* **123** no. 8, (Aug, 2019) . http://dx.doi.org/10.1103/PhysRevLett.123.081601.
[13] K. Goto, H. Marrochio, R. C. Myers, L. Queimada, and B. Yoshida, "Holographic complexity equals which action?" *Journal of High Energy Physics* **2019** no. 2, (Feb, 2019) . http://dx.doi.org/10.1007/JHEP02(2019)160.
[14] A. Bhattacharyya, A. Shekar, and A. Sinha, "Circuit complexity in interacting QFTs and RG flows," *JHEP* **10** (2018) 140, arXiv:1808.03105 [hep-th].
[15] R. Khan, C. Krishnan, and S. Sharma, "Circuit Complexity in Fermionic Field Theory," *Phys. Rev. D* **98** no. 12, (2018) 126001, arXiv:1801.07620 [hep-th].
[16] L. Hackl and R. C. Myers, "Circuit complexity for free fermions," *JHEP* **07** (2018) 139, arXiv:1803.10638 [hep-th].
[17] D. W. Alves and G. Camilo, "Evolution of complexity following a quantum quench in free field theory," *JHEP* **06** (2018) 029, arXiv:1804.00107 [hep-th].
[18] H. A. Camargo, P. Caputa, D. Das, M. P. Heller, and R. Jefferson, "Complexity as a novel probe of quantum quenches: universal scalings and purifications," *Phys. Rev. Lett.* **122** no. 8, (2019) 081601, arXiv:1807.07075 [hep-th].
[19] H. A. Camargo, M. P. Heller, R. Jefferson, and J. Knaute, "Path integral optimization as circuit complexity," *Phys. Rev. Lett.* **123** no. 1, (2019) 011601, arXiv:1904.02713 [hep-th].

[20] S. Chapman, J. Eisert, L. Hackl, M. P. Heller, R. Jefferson, H. Marrochio, and R. C. Myers, "Complexity and entanglement for thermofield double states," *SciPost Phys.* **6** no. 3, (2019) 034, arXiv:1810.05151 [hep-th].

[21] M. Doroudiani, A. Naseh, and R. Pirmoradian, "Complexity for Charged Thermofield Double States," *JHEP* **01** (2020) 120, arXiv:1910.08806 [hep-th].

[22] S. S. Hashemi, G. Jafari, and A. Naseh, "First law of holographic complexity," *Phys. Rev. D* **102** no. 10, (2020) 106008, arXiv:1912.10436 [hep-th].

[23] A. Bernamonti, F. Galli, J. Hernandez, R. C. Myers, S.-M. Ruan, and J. Simón, "Aspects of the first law of complexity," *Journal of Physics A: Mathematical and Theoretical* **53** no. 29, (Jul, 2020) 294002. http://dx.doi.org/10.1088/1751-8121/ab8e66.

[24] E. Caceres, S. Chapman, J. D. Couch, J. P. Hernandez, R. C. Myers, and S.-M. Ruan, "Complexity of mixed states in qft and holography," *Journal of High Energy Physics* **2020** no. 3, (Mar, 2020) . http://dx.doi.org/10.1007/JHEP03(2020)012.

[25] S. Choudhury, A. Dutta, and D. Ray, "Chaos and Complexity from Quantum Neural Network: A study with Diffusion Metric in Machine Learning," *JHEP* **04** (2021) 138, arXiv:2011.07145 [hep-th].

[26] A. R. Brown, L. Susskind, and Y. Zhao, "Quantum Complexity and Negative Curvature," *Phys. Rev. D* **95** no. 4, (2017) 045010, arXiv:1608.02612 [hep-th].

[27] J. Couch, W. Fischler, and P. H. Nguyen, "Noether charge, black hole volume, and complexity," *JHEP* **03** (2017) 119, arXiv:1610.02038 [hep-th].

[28] N. Chagnet, S. Chapman, J. de Boer, and C. Zukowski, "Complexity for Conformal Field Theories in General Dimensions," arXiv:2103.06920 [hep-th].

[29] R. d. M. Koch, M. Kim, and H. J. R. Van Zyl, "Complexity from Spinning Primaries," arXiv:2108.10669 [hep-th].

[30] J. L. F. Barbon and E. Rabinovici, "Holographic complexity and spacetime singularities," *JHEP* **01** (2016) 084, arXiv:1509.09291 [hep-th].

[31] M. Alishahiha, "Holographic Complexity," *Phys. Rev. D* **92** no. 12, (2015) 126009, arXiv:1509.06614 [hep-th].

[32] R.-Q. Yang, "Strong energy condition and complexity growth bound in holography," *Phys. Rev. D* **95** no. 8, (2017) 086017, arXiv:1610.05090 [gr-qc].

[33] S. Chapman, H. Marrochio, and R. C. Myers, "Complexity of Formation in Holography," *JHEP* **01** (2017) 062, arXiv:1610.08063 [hep-th].

[34] A. Reynolds and S. F. Ross, "Divergences in Holographic Complexity," *Class. Quant. Grav.* **34** no. 10, (2017) 105004, arXiv:1612.05439 [hep-th].

[35] Y. Zhao, "Complexity and Boost Symmetry," *Phys. Rev. D* **98** no. 8, (2018) 086011, arXiv:1702.03957 [hep-th].

[36] D. Carmi, S. Chapman, H. Marrochio, R. C. Myers, and S. Sugishita, "On the Time Dependence of Holographic Complexity," *JHEP* **11** (2017) 188, arXiv:1709.10184 [hep-th].

[37] A. Reynolds and S. F. Ross, "Complexity in de Sitter Space," *Class. Quant. Grav.* **34** no. 17, (2017) 175013, arXiv:1706.03788 [hep-th].

[38] B. Swingle and Y. Wang, "Holographic Complexity of Einstein-Maxwell-Dilaton Gravity," *JHEP* **09** (2018) 106, arXiv:1712.09826 [hep-th].

[39] Z. Fu, A. Maloney, D. Marolf, H. Maxfield, and Z. Wang, "Holographic complexity is nonlocal," *JHEP* **02** (2018) 072, arXiv:1801.01137 [hep-th].

[40] S. Bolognesi, E. Rabinovici, and S. R. Roy, "On Some Universal Features of the Holographic Quantum Complexity of Bulk Singularities," *JHEP* **06** (2018) 016, arXiv:1802.02045 [hep-th].

[41] B. Chen, W.-M. Li, R.-Q. Yang, C.-Y. Zhang, and S.-J. Zhang, "Holographic subregion complexity under a thermal quench," *JHEP* **07** (2018) 034, arXiv:1803.06680 [hep-th].

[42] R. Abt, J. Erdmenger, M. Gerbershagen, C. M. Melby-Thompson, and C. Northe, "Holographic Subregion Complexity from Kinematic Space," *JHEP* **01** (2019) 012, arXiv:1805.10298 [hep-th].

[43] K. Hashimoto, N. Iizuka, and S. Sugishita, "Thoughts on Holographic Complexity and its Basis-dependence," *Phys. Rev. D* **98** no. 4, (2018) 046002, arXiv:1805.04226 [hep-th].

[44] J. Couch, S. Eccles, T. Jacobson, and P. Nguyen, "Holographic Complexity and Volume," *JHEP* **11** (2018) 044, arXiv:1807.02186 [hep-th].

[45] O. Ben-Ami and D. Carmi, "On Volumes of Subregions in Holography and Complexity," *JHEP* **11** (2016) 129, arXiv:1609.02514 [hep-th].

[46] A. R. Brown and L. Susskind, "Second law of quantum complexity," *Phys. Rev. D* **97** no. 8, (2018) 086015, arXiv:1701.01107 [hep-th].

[47] K. Hashimoto, N. Iizuka, and S. Sugishita, "Time evolution of complexity in Abelian gauge theories," *Phys. Rev. D* **96** no. 12, (2017) 126001, arXiv:1707.03840 [hep-th].

[48] T. Ali, A. Bhattacharyya, S. S. Haque, E. H. Kim, N. Moynihan, and J. Murugan, "Chaos and Complexity in Quantum Mechanics," *Phys. Rev. D* **101** no. 2, (2020) 026021, arXiv:1905.13534 [hep-th].

[49] S. Chapman and H. Z. Chen, "Charged Complexity and the Thermofield Double State," *JHEP* **02** (2021) 187, arXiv:1910.07508 [hep-th].

[50] L. Susskind, "Three Lectures on Complexity and Black Holes," 10, 2018. arXiv:1810.11563 [hep-th].

[51] D. Stanford and L. Susskind, "Complexity and Shock Wave Geometries," *Phys. Rev. D* **90** no. 12, (2014) 126007, arXiv:1406.2678 [hep-th].

[52] L. Susskind, "Computational Complexity and Black Hole Horizons," *Fortsch. Phys.* **64** (2016) 24–43, arXiv:1403.5695 [hep-th]. [Addendum: Fortsch.Phys. 64, 44–48 (2016)].

[53] D. A. Roberts, D. Stanford, and L. Susskind, "Localized shocks," *JHEP* **03** (2015) 051, arXiv:1409.8180 [hep-th].

[54] L. Susskind and Y. Zhao, "Switchbacks and the Bridge to Nowhere," arXiv:1408.2823 [hep-th].

[55] R. Jefferson and R. C. Myers, "Circuit complexity in quantum field theory," *JHEP* **10** (2017) 107, arXiv:1707.08570 [hep-th].

[56] S. Choudhury, "The Cosmological OTOC: Formulating new cosmological micro-canonical correlation functions for random chaotic fluctuations in Out-of-Equilibrium Quantum Statistical Field Theory," *Symmetry* **12** no. 9, (2020) 1527, arXiv:2005.11750 [hep-th].

[57] K. Y. Bhagat, B. Bose, S. Choudhury, S. Chowdhury,

R. N. Das, S. G. Dastider, N. Gupta, A. Maji, G. D. Pasquino, and S. Paul, "The Generalized OTOC from Supersymmetric Quantum Mechanics—Study of Random Fluctuations from Eigenstate Representation of Correlation Functions," *Symmetry* **13** no. 1, (2020) 44, arXiv:2008.03280 [hep-th].

[58] S. Choudhury, "The Cosmological OTOC: A New Proposal for Quantifying Auto-correlated Random Non-chaotic Primordial Fluctuations," *Symmetry* **13** no. 4, (2021) 599, arXiv:2106.01305 [physics.gen-ph].

[59] K. Hashimoto, K. Murata, and R. Yoshii, "Out-of-time-order correlators in quantum mechanics," *JHEP* **10** (2017) 138, arXiv:1703.09435 [hep-th].

[60] K. Hashimoto, K.-B. Huh, K.-Y. Kim, and R. Watanabe, "Exponential growth of out-of-time-order correlator without chaos: inverted harmonic oscillator," *JHEP* **11** (2020) 068, arXiv:2007.04746 [hep-th].

[61] S. Choudhury, S. Chowdhury, N. Gupta, A. Mishara, S. P. Selvam, S. Panda, G. D. Pasquino, C. Singha, and A. Swain, "Circuit Complexity From Cosmological Islands," *Symmetry* **13** (2021) 1301, arXiv:2012.10234 [hep-th].

[62] A. Bhattacharyya, S. Das, S. Shajidul Haque, and B. Underwood, "Cosmological Complexity," *Phys. Rev. D* **101** no. 10, (2020) 106020, arXiv:2001.08664 [hep-th].

[63] A. Bhattacharyya, S. Das, S. S. Haque, and B. Underwood, "Rise of cosmological complexity: Saturation of growth and chaos," *Phys. Rev. Res.* **2** no. 3, (2020) 033273, arXiv:2005.10854 [hep-th].

[64] P. Bhargava, S. Choudhury, S. Chowdhury, A. Mishara, S. P. Selvam, S. Panda, and G. D. Pasquino, "Quantum aspects of chaos and complexity from bouncing cosmology: A study with two-mode single field squeezed state formalism," *SciPost Phys. Core* **4** (2021) 026, arXiv:2009.03893 [hep-th].

[65] K. Adhikari, S. Choudhury, H. N. Pandya, and R. Srivastava, "PGW Circuit Complexity," arXiv:2108.10334 [gr-qc].

[66] K. Adhikari, S. Choudhury, S. Chowdhury, K. Shirish, and A. Swain, "Circuit complexity as a novel probe of quantum entanglement: A study with black hole gas in arbitrary dimensions," *Phys. Rev. D* **104** no. 6, (2021) 065002, arXiv:2104.13940 [hep-th].

[67] S. Choudhury, S. P. Selvam, and K. Shirish, "Circuit Complexity From Supersymmetric Quantum Field Theory With Morse Function," arXiv:2101.12582 [hep-th].

[68] J. Eisert, "Entangling power and quantum circuit complexity," arXiv:2104.03332 [quant-ph].

[69] A. Pich, "Effective field theory: Course," in *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions*, pp. 949–1049. 6, 1998. arXiv:hep-ph/9806303.

[70] C. P. Burgess, "Introduction to Effective Field Theory," *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 329–362, arXiv:hep-th/0701053.

[71] R. Shankar, "Effective field theory in condensed matter physics," in *Conference on Historical Examination and Philosophical Reflections on the Foundations of Quantum Field Theory*, pp. 47–55. 3, 1996. arXiv:cond-mat/9703210.

[72] J. F. Donoghue, "Introduction to the effective field theory description of gravity," in *Advanced School on Effective Theories*. 6, 1995. arXiv:gr-qc/9512024.

[73] J. F. Donoghue, "The effective field theory treatment of quantum gravity," *AIP Conf. Proc.* **1483** no. 1, (2012) 73–94, arXiv:1209.3511 [gr-qc].

[74] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, "The Effective Field Theory of Inflation," *JHEP* **03** (2008) 014, arXiv:0709.0293 [hep-th].

[75] S. Weinberg, "Effective Field Theory for Inflation," *Phys. Rev. D* **77** (2008) 123541, arXiv:0804.4291 [hep-th].

[76] N. Agarwal, R. Holman, A. J. Tolley, and J. Lin, "Effective field theory and non-Gaussianity from general inflationary states," *JHEP* **05** (2013) 085, arXiv:1212.1172 [hep-th].

[77] O. Özsoy, J. T. Giblin, E. Nesbit, G. Şengör, and S. Watson, "Toward an Effective Field Theory Approach to Reheating," *Phys. Rev. D* **96** no. 12, (2017) 123524, arXiv:1701.01455 [hep-th].

[78] O. Özsoy, G. Sengor, K. Sinha, and S. Watson, "A Model Independent Approach to (p)Reheating," arXiv:1507.06651 [hep-th].

[79] C. P. Burgess, "Intro to Effective Field Theories and Inflation," arXiv:1711.10592 [hep-th].

[80] D. Baumann and L. McAllister, *Inflation and String Theory*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 5, 2015. arXiv:1404.2601 [hep-th].

[81] D. Baumann, "Inflation," in *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small*, pp. 523–686. 2011. arXiv:0907.5424 [hep-th].

[82] S. Choudhury, *Field Theoretic Approaches To Early Universe*. PhD thesis, Indian Statistical Inst., Calcutta, 2016. arXiv:1603.08306 [hep-th].

[83] S. Choudhury, *Quantum Field Theory approaches to Early Universe Cosmology*. LAP LAMBERT Academic Publishing, 5, 2018.

[84] S. Choudhury and A. Mazumdar, "An accurate bound on tensor-to-scalar ratio and the scale of inflation," *Nucl. Phys. B* **882** (2014) 386–396, arXiv:1306.4496 [hep-ph].

[85] S. Choudhury, "Can Effective Field Theory of inflation generate large tensor-to-scalar ratio within Randall–Sundrum single braneworld?," *Nucl. Phys. B* **894** (2015) 29–55, arXiv:1406.7618 [hep-th].

[86] L. V. Delacretaz, V. Gorbenko, and L. Senatore, "The Supersymmetric Effective Field Theory of Inflation," *JHEP* **03** (2017) 063, arXiv:1610.04227 [hep-th].

[87] L. V. Delacretaz, T. Noumi, and L. Senatore, "Boost Breaking in the EFT of Inflation," *JCAP* **02** (2017) 034, arXiv:1512.04100 [hep-th].

[88] D. Lopez Nacir, R. A. Porto, L. Senatore, and M. Zaldarriaga, "Dissipative effects in the Effective Field Theory of Inflation," *JHEP* **01** (2012) 075, arXiv:1109.4192 [hep-th].

[89] A. Naskar, S. Choudhury, A. Banerjee, and S. Pal, "Inflation to Structures: EFT all the way," arXiv:1706.08051 [astro-ph.CO].

[90] L. Senatore and M. Zaldarriaga, "The Effective Field Theory of Multifield Inflation," *JHEP* **04** (2012) 024,

arXiv:1009.2093 [hep-th].
[91] D. Baumann, D. Green, H. Lee, and R. A. Porto, "Signs of Analyticity in Single-Field Inflation," *Phys. Rev. D* **93** no. 2, (2016) 023523, arXiv:1502.07304 [hep-th].
[92] S. Dubovsky, L. Hui, A. Nicolis, and D. T. Son, "Effective field theory for hydrodynamics: thermodynamics, and the derivative expansion," *Phys. Rev. D* **85** (2012) 085029, arXiv:1107.0731 [hep-th].
[93] M. Crossley, P. Glorioso, and H. Liu, "Effective field theory of dissipative fluids," *JHEP* **09** (2017) 095, arXiv:1511.03646 [hep-th].
[94] K. Adhikari, S. Choudhury, S. Kumar, S. Mandal, N. Pandey, A. Roy, S. Sarkar, P. Sarker, and S. S. Shariff, "Circuit Complexity in $\mathcal{Z}_2\mathcal{EEFT}$," arXiv:2109.09759 [hep-th].
[95] S. Choudhury, "CMB from EFT," *Universe* **5** no. 6, (2019) 155, arXiv:1712.04766 [hep-th].
[96] S. Choudhury and S. Panda, "COSMOS-e'-GTachyon from string theory," *Eur. Phys. J. C* **76** no. 5, (2016) 278, arXiv:1511.05734 [hep-th].
[97] S. Choudhury, "COSMOS-e'- soft Higgsotic attractors," *Eur. Phys. J. C* **77** no. 7, (2017) 469, arXiv:1703.01750 [hep-th].
[98] K. Aoki, M. A. Gorji, S. Mukohyama, and K. Takahashi, "The Effective Field Theory of Vector-Tensor Theories," arXiv:2111.08119 [hep-th].
[99] S. Choudhury, A. Mazumdar, and E. Pukartas, "Constraining $\mathcal{N} = 1$ supergravity inflationary framework with non-minimal Kähler operators," *JHEP* **04** (2014) 077, arXiv:1402.1227 [hep-th].
[100] M. A. Gorji, H. Motohashi, and S. Mukohyama, "Inflation with $0 \leq c_s \leq 1$," arXiv:2110.10731 [hep-th].
[101] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, "Ghost inflation," *JCAP* **04** (2004) 001, arXiv:hep-th/0312100.
[102] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2015 results. XX. Constraints on inflation," *Astron. Astrophys.* **594** (2016) A20, arXiv:1502.02114 [astro-ph.CO].
[103] H. Ruegg and M. Ruiz-Altaba, "The Stueckelberg field," *Int. J. Mod. Phys. A* **19** (2004) 3265–3348, arXiv:hep-th/0304245.
[104] C. Grosse-Knetter and R. Kogerler, "Unitary gauge, Stuckelberg formalism and gauge invariant models for effective lagrangians," *Phys. Rev. D* **48** (1993) 2865–2876, arXiv:hep-ph/9212268.
[105] P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pp. 124–134. 1994.
[106] P. W. Shor, "Polynomial time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM J. Sci. Statist. Comput.* **26** (1997) 1484, arXiv:quant-ph/9508027.
[107] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.
[108] S. Arora and B. Barak, *Computational Complexity: A Modern Approach*. Cambridge University Press, 2011.
[109] C. Moore and S. Mertens, *The Nature of Computation*. Oxford University Press.
[110] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. Smolin, and H. Weinfurter, "Elementary gates for quantum computation," *Phys. Rev. A* **52** (1995) 3457, arXiv:quant-ph/9503016.
[111] S. Aaronson, "The Complexity of Quantum States and Transformations: From Quantum Money to Black Holes," 7, 2016. arXiv:1607.05256 [quant-ph].
[112] D. P. DiVincenzo, "The physical implementation of quantum computation," *Fortschritte der Physik* **48** no. 9-11, (Sep, 2000) 771–783. http://dx.doi.org/10.1002/1521-3978(200009)48:9/11<771::AID-PROP771>3.0.CO;2-E.
[113] M. A. Nielsen, "A geometric approach to quantum circuit lower bounds," arXiv:quant-ph/0502070 [quant-ph].
[114] M. A. Nielsen, "Quantum computation as geometry," *Science* **311** no. 5764, (Feb, 2006) 1133–1135.
[115] M. A. N. Mark R. Dowling, "The geometry of quantum computation," arXiv:arXiv:quant-ph/0701004 [quant-ph].
[116] C. Gerry and P. Knight, *Introductory Quantum Optics*. Cambridge University Press, 2004.
[117] M. Alishahiha, E. Silverstein, and D. Tong, "DBI in the sky," *Phys. Rev. D* **70** (2004) 123505, arXiv:hep-th/0404084.
[118] S. Choudhury and S. Pal, "DBI Galileon inflation in background SUGRA," *Nucl. Phys. B* **874** (2013) 85–114, arXiv:1208.4433 [hep-th].
[119] A. Mazumdar, S. Panda, and A. Perez-Lorenzana, "Assisted inflation via tachyon condensation," *Nucl. Phys. B* **614** (2001) 101–116, arXiv:hep-ph/0107058.
[120] D. Choudhury, D. Ghoshal, D. P. Jatkar, and S. Panda, "On the cosmological relevance of the tachyon," *Phys. Lett. B* **544** (2002) 231–238, arXiv:hep-th/0204204.
[121] N. Chow and J. Khoury, "Galileon Cosmology," *Phys. Rev. D* **80** (2009) 024037, arXiv:0905.1325 [hep-th].
[122] A. De Felice and S. Tsujikawa, "Generalized Galileon cosmology," *Phys. Rev. D* **84** (2011) 124029, arXiv:1008.4236 [hep-th].
[123] A. Bhattacharyya, S. S. Haque, and E. H. Kim, "Complexity from the Reduced Density Matrix: a new Diagnostic for Chaos," *JHEP* **10** (2021) 028, arXiv:2011.04705 [hep-th].
[124] S. S. Haque, C. Jana, and B. Underwood, "Saturation of Thermal Complexity of Purification," arXiv:2107.08969 [hep-th].
[125] S. S. Haque, C. Jana, and B. Underwood, "Operator Complexity for Continuous Variable Systems," arXiv:2110.08356 [hep-th].

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/353904020

Constraints on holographic dark energy model with matter creation in Brans–Dicke theory and thermodynamic analysis

Article *in* Physics of the Dark Universe · August 2021

DOI: 10.1016/j.dark.2021.100869

CITATIONS
0

READS
24

2 authors:



Simran Kaur
Delhi Technological University
3 PUBLICATIONS **12** CITATIONS
SEE PROFILE



C.P. Singh
Delhi Technological University
94 PUBLICATIONS **1,322** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Dark energy models with recently proposed logarithmic form of Brans-Dicke scalar field View project

STUDY OF COSMOLOGICAL MODELS IN MODIFIED THEORIES OF GRAVITATION View project

All content following this page was uploaded by C.P. Singh on 06 November 2021.

The user has requested enhancement of the downloaded file.

Physics of the Dark Universe 33 (2021) 100869

Contents lists available at ScienceDirect

Physics of the Dark Universe

journal homepage: www.elsevier.com/locate/dark

Constraints on holographic dark energy model with matter creation in Brans–Dicke theory and thermodynamic analysis

Simran Kaur[1], C.P. Singh[1,*]

Department of Applied Mathematics, Delhi Technological University, Delhi 110 042, India

ARTICLE INFO

Article history:
Received 25 April 2021
Received in revised form 11 July 2021
Accepted 11 August 2021

Keywords:
Holographic dark energy
FRW model
Brans–Dicke theory
Thermodynamics

ABSTRACT

The adiabatic matter creation for a spatially flat, homogeneous and isotropic universe is discussed in holographic dark energy (HDE) in the context of Brans–Dicke theory. Since the HDE density is considered as a dynamical cosmological constant, it is natural to study it in a dynamical frame of Brans–Dicke theory. We show that HDE model with matter creation, whose infrared cutoff is set by the Hubble horizon, leads to solution that unify the early and late time acceleration. The latest observational data from supernovae (Pantheon sample), Hubble $H(z)$ and latest local H_0 by SH0ES are used for constraining the model parameters. Using the Markov chain Monte Carlo method, we obtain the best fit values of the parameters for different models. It is found that the HDE model without matter creation does not achieve transition phase when the Hubble horizon is taken as an IR cut-off. However, it achieves the phase transition from decelerated phase to accelerated phase with same IR cut-off with the inclusion of gravitationally induced matter creation. The best fit values of HDE model with matter creation are fitted very well in Hubble error data. Using AIC and BIC, and geometrical diagnostic parameters, we compare the models with ΛCDM model and discuss the viability of the model. A detailed thermodynamic analysis is also carried out.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, the discovery of late time accelerated expansion [1,2] of the universe is one of the most important development in modern cosmology. The physical origin of this acceleration remains a deep mystery. It has been observed that two-third of the energy density of the universe exists as a new unknown component, called dark energy (DE). This has been confirmed by the subsequent observations from type Ia supernovae (SNe) [3–7] and also by other probes like, cosmic microwave background (CMB) and baryon acoustic oscillations (BAO) [8,9].

Many theoretical approaches have been carried out to investigate this recent cosmic acceleration. The preferred candidate of DE is the Einstein's cosmological constant which fits the observations well, but it faces several strong problems such as the fine-tuning and the cosmic coincidence problems [10–12]. In general relativistic cosmology, the presence of negative pressure can be seen as a possible driving mechanism for the accelerated expansion of the universe. It can be observed through the accelerating Friedmann equations which require generally an effective negative pressure and thus violates the energy condition. Thus, a sufficient negative pressure is the key ingredient for the accelerating universe.

The negative pressure occurs naturally in many different contexts when the physical systems depart from thermodynamic equilibrium. One such example is the matter creation induced by the time-varying gravitational field in an expanding universe [13–18]. It has been shown that the negative pressure accompanying matter creation can be derived using relativistic non-equilibrium thermodynamics [15–18]. Later on, it has been demonstrated that matter creation is an irreversible process completely different from bulk viscosity mechanism [19,20]. Many phenomenological models of matter creation have been proposed in literature [21–29]. In recent years, it has been shown that the matter creation can explain not only the present era of cosmic acceleration but also provides a viable alternative to the concordance ΛCDM model [30–45].

On the other hand, there is another concept in which the DE is assumed a dynamically evolving component. In fact, several dynamical models of DE have been proposed and studied in detail. One of such dynamical models is the holographic dark energy (HDE) which is basically based on the holographic principle [46–48] of quantum gravity. Bekenstein's entropy bound suggests that the conventional local quantum field theory breaks down due to the too many degrees of freedom. This leads to the formation of black hole. According to the holographic principle, for a system with size L and ultra-violet (UV) cut-off Λ, the vacuum energy

* Corresponding author.
E-mail addresses: skjuneja.maths@gmail.com (S. Kaur), cpsphd@rediffmail.com (C.P. Singh).

[1] Both the authors contributed in writng the paper

https://doi.org/10.1016/j.dark.2021.100869
2212-6864/© 2021 Elsevier B.V. All rights reserved.

density should not exceed the mass of the black hole of the same size, i.e., $L^3\rho_\Lambda \leq LM_{Pl}^2$, where L is the size of the region of black hole, ρ_Λ is the vacuum energy density, M_{Pl} is the reduced Planck mass $M_{Pl}^{-2} = 8\pi G$. At the equilibrium, the HDE density is defined as $\rho_\Lambda = 3b^2M_{Pl}^2L^{-2}$, where b is a numerical constant. The region of size L, i.e., infra-red (IR) cut-off is related to the large scale structure of the universe, i.e., Hubble horizon, particle horizon, event horizon, etc. Hsu [49] studied the HDE model with Hubble horizon as an IR cut-off. However, it could not explain the accelerated expansion of the universe. Later on, Li [50] studied the HDE model with event horizon and found that it could explain the acceleration phenomena. But it is also not compatible with the age of some old high redshift objects [51]. In a paper, Xu et al. [52] have studied HDE model with Hubble horizon as an IR cut-off in BD theory and found that there is no accelerated expansion of the universe. However, an accelerated expansion is obtained once the event horizon is considered as an IR cut-off. In literature, many authors [53–55] have studied the HDE models by assuming different IR cut-offs to discuss the evolution of the universe.

The Brans–Dicke(BD) theory [56] is the most natural choice as the scalar tensor generalization of general relativity which was formulated in order to incorporate the Mach's Principle. In this theory, the gravitational constant G is replaced with the inverse of a time-dependent BD scalar field ϕ, which couples to gravity with a coupling parameter ω. It explains the recent accelerated expansion of the universe and accommodates the observational data as well [57–59].

In this paper, we consider HDE model with matter creation in a more dynamical framework, i.e., in BD cosmology. It is to be noted that the HDE density belongs to a dynamical cosmological constant, therefore, we need a dynamical frame to accommodate it instead of Einstein gravity. Thus, the study of HDE model in the framework of BD theory is well motivated. A number of authors [60–67] have studied HDE models with different IR cut-offs in the framework of BD theory to explain the recent acceleration and to alleviate the cosmic coincidence problem. Xu and Li [68] have constrained the HDE model in BD theory by cosmic observational data.

The purpose of this paper is to study a cosmological model of late acceleration in BD theory with the assumption that the pressureless dark matter with matter creation and HDE conserve separately. Instead of assuming the other IR cut-off, we include matter creation mechanism in HDE model with the same Hubble horizon as an IR cut-off with the possibility that this model would be consistent with the current observations and show the phase transition. We consider HDE models with and without matter creation. We discuss the evolution and dynamics of the model by constraining the model parameters through combined observational data of Supernova(Pantheon), Hubble data and local Hubble constant. It is found that the HDE model with matter creation shows phase transition from decelerated phase to accelerated phase. The model behaves quintessence like and approaches to ΛCDM model in late times.

The paper is organized as follows. In Section 2, the basic of matter creation cosmology is described. In Section 3, we present the field equations of HDE model with Hubble horizon as an IR cut-off in BD theory when adiabatic matter creation is allowed. Section 4 is devoted to the analytical solution of the different HDE models with and without matter creation. Section 5 studies the constraints on cosmological model by using data of Supernova(pantheon), Hubble data and local Hubble constant. In Section 6, we present the result of the best-fit values of different model parameters to discuss the evolution of the universe. Section 7 discusses the geometrical diagnostic parameters to distinguish HDE models with ΛCDM model. We present the thermodynamic analysis of the model based on generalized second law of thermodynamics in Section 8. Finally, we summarize our findings in Section 9.

2. A brief note on matter creation cosmology

As stated in the introduction that Prigogine et al. [15–17] proposed a theoretical approach to matter creation thermodynamics in the framework of cosmology which includes large-scale entropy production. They have shown how the thermodynamic of open systems in cosmology leads naturally to a modification in the energy–momentum tensor by adding an additional pressure term through a covariant formulation. In what follows, we briefly review the thermodynamics of matter creation in cosmology.

For a closed system where the total number of particle remains constant, the first law of thermodynamics can be expressed as [16,17]

$$d(\rho_m V) = dQ - p_m dV, \tag{1}$$

where ρ_m is the energy density, p_m the thermodynamical pressure and dQ is the heat received by the system during time dt.

However, if we consider the universe as an open system where the total number of particles is not invariant, the first law of thermodynamics is given by

$$d(\rho_m V) + p_m dV = \frac{h}{n}\, d(n\, V), \tag{2}$$

where $n = N/V$ is the particle number per unit volume, $h = \rho_m + p_m$ is the enthalpy per unit volume of the system. This particle creation corresponds to irreversible energy flow from gravitational field to matter constituents. Hence, the creation of matter acts as a source of internal energy.

Using $V = \frac{4\pi}{3}a^3(t)$, where a is the scale factor, the right hand side of (2) can be written as

$$\frac{h}{n}\, d(n\, V) = (\rho_m + p_m)\, V \left(\frac{\dot{n}}{n} + 3H\right), \tag{3}$$

where $H = \dot{a}/a$ is the Hubble parameter and an overdot denotes derivative with respect to the cosmic time t.

The particle flux vector is assumed to have the form $N^\mu = nu^\mu$, where N^μ satisfies the balance equation $N^\mu_{;\mu} = 0$. However, in the gravitationally induced particle creation mechanism, this balance equation is modified as $N^\mu_{;\mu} = n\,\Gamma$, which can be rewritten as

$$\dot{n} + 3Hn = n\Gamma, \tag{4}$$

where Γ is the particle production rate. If $\Gamma > 0$, we have particle creation, $\Gamma < 0$ gives the particle destruction and $\Gamma = 0$ means particle number is conserved.

Now, inserting (3) and (4) into (2), we obtain

$$\dot{\rho_m} + 3H(\rho_m + p_m + p_c) = 0, \tag{5}$$

where p_c is the effect of particle pressure which is expressed as [30,31]

$$p_c = -\frac{(\rho_m + p_m)}{3H}\Gamma. \tag{6}$$

Thus, the negative pressure p_c is a supplementary pressure to the thermodynamic pressure p_m. Hence, the energy–momentum tensor of the cosmic fluid in the presence of creation of matter includes the creation pressure term p_c and is given by

$$T_{\mu\nu} = (\rho_m + p_m + p_c)u_\mu u_\nu - (p_m + p_c)g_{\mu\nu} \tag{7}$$

From (7), it can be seen that the energy–momentum tensor of the matter creation part is given by

$$T^{crea}_{\mu\nu} = p_c(u_\mu u_\nu - g_{\mu\nu}), \tag{8}$$

where p_c is given by (6). The negative pressure caused by particle creation can explain the current accelerated expansion of the universe. Thus, the energy transfer from space–time curvature to matter appeared to be an irreversible process leading to a burst of entropy associated with the creation of matter whose change is given by $dS = \frac{s}{n}d(nV)$, where $s = S/V$ is the specific entropy.

3. HDE model with matter creation in BD theory

The Brans–Dicke (BD) theory is an attempt to properly incorporate both the Mach's principle [69] and the Dirac's Large Number Hypothesis [69] in which gravitational constant is allowed to vary with space–time. In Jordan frame, the action of BD theory of gravity in the presence of matter with Lagrangian $\mathcal{L}_m$ is described by [70]

$$S = \int d^4x\sqrt{-g}\left[\frac{1}{16\pi}\left(\phi R - \frac{\omega}{\phi}\nabla_\alpha\phi\nabla^\alpha\phi\right) + \mathcal{L}_m\right], \tag{9}$$

where R is the Ricci scalar, ϕ is the BD scalar field representing the inverse of Newton's constant which is allowed to vary with space and time, ω is the generic dimensionless BD coupling parameter which measures how strongly ϕ couples to matter content and $\mathcal{L}_m$ represents the matter Lagrangian density excluding the scalar field.

Variation of the action (9) with respect to the metric $g_{\mu\nu}$ and the BD scalar field ϕ, yield the following field equations.

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi}{\phi}\tilde{T}_{\mu\nu} + 8\pi T^{BD}_{\mu\nu}, \tag{10}$$

and

$$\nabla_\alpha\nabla^\alpha\phi = \frac{8\pi}{(2\omega+3)}\tilde{T}^\lambda_\lambda. \tag{11}$$

In Eq. (10), $\tilde{T}_{\mu\nu}$ is the total energy–momentum tensor which is the sum of the matter fluid including creation and HDE parts, and adopts the perfect fluid form:

$$\tilde{T}_{\mu\nu} = (\rho_t + p_t)u_\mu u_\nu + p_t g_{\mu\nu}, \tag{12}$$

where $\rho_t = \rho_m + \rho_h$ and $p_t = p_m + p_c + p_h$. Also, $T^{BD}_{\mu\nu}$ is the energy–momentum for the BD scalar which is defined by

$$T^{BD}_{\mu\nu} = \frac{1}{8\pi}\left[\frac{\omega}{\phi^2}\left(\nabla_\mu\phi\nabla_\nu\phi - \frac{1}{2}g_{\mu\nu}\nabla_\alpha\phi\nabla^\alpha\phi\right) + \frac{1}{\phi}\left(\nabla_\mu\nabla_\nu\phi - g_{\mu\nu}\nabla_\alpha\nabla^\alpha\phi\right)\right]. \tag{13}$$

We assume the background as a spatially homogeneous and isotropic flat Friedmann–Robertson–Walker (FRW) space–time, which is given by

$$ds^2 = -dt^2 + a^2(t)\left[dr^2 + r^2(d\theta^2 + sin^2\theta d\phi^2)\right], \tag{14}$$

where $a(t)$ is the scale factor of the universe. Throughout we use units such that the speed of light, $c = 1$.

The gravitational field equations for a universe dominated by dark matter (DM) and HDE in which a process of DM creation from the gravitational field, governed by (7), yield

$$3H^2 + 3H\frac{\dot{\phi}}{\phi} - \frac{\omega}{2}\frac{\dot{\phi}^2}{\phi^2} = \frac{8\pi}{\phi}\rho_t, \tag{15}$$

$$2\dot{H} + 3H^2 + \frac{\ddot{\phi}}{\phi} + 2H\frac{\dot{\phi}}{\phi} + \frac{\omega}{2}\frac{\dot{\phi}^2}{\phi^2} = -\frac{8\pi}{\phi}p_t \tag{16}$$

where an over dot denotes derivative with respect to cosmic time t and $H = \dot{a}/a$ is the usual Hubble parameter. The dynamical equation for the scalar field is given by

$$\frac{\ddot{\phi}}{\phi} + 3H\frac{\dot{\phi}}{\phi} = \frac{8\pi}{(2\omega+3)\phi}(\rho_t - 3p_t), \tag{17}$$

Ignoring the inhomogeneities arising from the (linear) field perturbations, the BD field can be treated as a perfect fluid with energy and pressure are respectively given by

$$\rho_{BD} = \frac{1}{8\pi G}\left[\frac{\omega}{2}\left(\frac{\dot{\phi}}{\phi}\right)^2 - 3H\frac{\dot{\phi}}{\phi}\right], \tag{18}$$

$$p_{BD} = \frac{1}{8\pi G}\left[\frac{\omega}{2}\left(\frac{\dot{\phi}}{\phi}\right)^2 + 2H\frac{\dot{\phi}}{\phi} + \frac{\ddot{\phi}}{\phi}\right]. \tag{19}$$

Finally, the geometric Bianchi identity of $\nabla_\nu G^{\mu\nu} = 0$ in Eq. (10), which plays a role of the consistency relation, leads to

$$\nabla_\nu\left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R\right) = 0 = \nabla_\nu\left(\frac{8\pi}{\phi}\tilde{T}^{\mu\nu} + 8\pi T^{\mu\nu}_{BD}\right). \tag{20}$$

In this paper, we study the model dominated by pressureless dark matter ($p_m = 0$) in the presence of matter creation and HDE in BD theory. Further, we assume that dark components – the pressureless dark matter with matter creation and the HDE – do not interact with each other. Therefore, the energy conservation equation for each conserves separately. It means that $\tilde{T}^{\mu\nu}$ obeys the usual conservation law, $\nabla_\nu\tilde{T}^{\mu\nu} = 0$, which takes the forms

$$\dot{\rho_m} + 3H(\rho_m + p_c) = 0, \tag{21}$$

and

$$\dot{\rho_h} + 3H(\rho_h + p_h) = 0. \tag{22}$$

One interesting thing about working in Jordan frame is that the conservation equation holds for matter and BD scalar field separately, i.e., equations of motion of matter do not enter into the BD scalar field. Therefore, from (20), we are left

$$\left(\nabla_\nu\{\frac{8\pi}{\phi}\}\tilde{T}^{\mu\nu} + \nabla_\nu\{8\pi T^{\mu\nu}_{BD}\}\right) = 0, \tag{23}$$

which gives

$$\dot{\rho}_{BD} + 3H(\rho_{BD} + p_{BD}) = \left(\frac{\dot{\phi}}{\phi^2}\right)\rho_t, \tag{24}$$

which can be treated as an additional conservation law. Eq. (24) is indeed a consistency condition originating from the Bianchi identity $\nabla_\nu G^{\mu\nu} = 0$ [58,70].

It can indeed be checked that the covariant conservation laws (21), (22) and (24) can also be obtained upon lengthy but straightforward computation by combining Eqs. (15)–(17), which are identical to that of general relativity. Although the calculation is more involved than in general relativity, the final result turns out to be the same. Thus, we shall use (21), (22) and (15) to obtain the solution of the model and finally we use (24) to get the consistency condition, which will be used to constraint the model parameters.

The general analytical solutions to the system (15)–(17) are not known. However, physical sense guides us into searching for possible solutions in which the BD scalar field ϕ evolves very slowly. We assume that BD scalar field can be described as a power-law of the scale factor, namely $\phi = \phi_0\ a^\alpha$, where ϕ_0 and α are constants [60–63]. A case of particular interest is that when α is small whereas ω is high so that the product $\alpha\omega$ results of order unity [60]. This is interesting because local experiments set a very high lower bound on ω. This choice with small α can lead to consistent results which may justify this specific choice among other possible choices [60]. We note that in Ref. [71] this kind of power-law solution is used to show that BD cosmology with a cosmological term can mimic the running vacuum model, which is very convenient in order to improve the fitting of the cosmological data [72]. With this power-law assumption, the Friedmann equation (15) can be rewritten as

$$\left(1 + \alpha - \frac{\omega}{6}\alpha^2\right)H^2 = \frac{8\pi}{3\phi}(\rho_m + \rho_h). \tag{25}$$

It is to be noted that the experiments of solar-system provide the lower bound of ω as $|\omega| > 40\,000$ [73]. But, when it is probed on a large scale, the value of ω is less than this result [74].

In the introduction section, we have discussed that there are various choices of IR cut-off for the cosmological length scale available in literature. The Hubble horizon, which is given by $\rho_h = 3b^2(8\pi GL^2)^{-1}$, where $L = H^{-1}$ is the IR cut-off, is the simplest choice of IR cut-off but this IR cut-off does not describe the cosmic acceleration [49]. In the framework of BD theory, the scalar field ϕ plays the role of Newton's constant ($\phi \sim 1/G$). Therefore, the modify HDE in BD theory is given by $\rho_h = \frac{3b^2}{8\pi}\phi H^2$ [70].

Now, using this form of ρ_h with (21) and (22), the evolution equation (25) for cold dark matter can be written as

$$\dot{H}+\left(\frac{3+\alpha}{2}+\frac{9b^2w_h}{6+6\alpha-\omega\alpha^2}\right)H^2=\frac{(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)}\,\Gamma H, \tag{26}$$

where $w_h = p_h/\rho_h$ is the equation of state (EoS) parameter of HDE. In what follows, we shall solve the above equation for different choices of particle creation rate Γ and perform the observational constraints on the model parameters.

4. Solution of HDE model

Now, one can observe from (26) that the dynamics of the universe can only be observed if the matter creation rate Γ is known. In literature, some phenomenological but general choice of Γ have been proposed to study the evolution of the universe. It has been shown that $\Gamma \propto H^2$ [75] gives inflationary solution, $\Gamma \propto H$ can explain the intermediate decelerating era [37], Γ =constant can explain the evolution of the universe from big-bang to the present accelerating phase [39]. Steigman et al. [31] proposed a linear combination of the above rates to describe the early and late time evolution of the universe. In what follows, we present different models depending on the choices of the matter creation rate Γ.

4.1. Solution with $\Gamma = 0$

Let us first find the solution of HDE model with Hubble horizon as an IR cut-off without matter creation (hereafter *HDE 1*). Taking $\Gamma = 0$, Eq. (26) reduces to

$$\dot{H}+\left(\frac{3+\alpha}{2}+\frac{9b^2w_h}{6+6\alpha-\omega\alpha^2}\right)H^2=0. \tag{27}$$

The solution of (27) is given by

$$H = H_0\left(\frac{a}{a_0}\right)^{-\left(\frac{3+\alpha}{2}+\frac{9b^2w_h}{6+6\alpha-\omega\alpha^2}\right)}, \tag{28}$$

where H_0 and a_0 are the present value of Hubble parameter and the scale factor at present cosmic time $t = t_0$, respectively. The solution of the scale factor, normalized to unity at present epoch, in terms of cosmic time t can be obtained as

$$a=\left[1+\left(\frac{3+\alpha}{2}+\frac{9b^2w_h}{6+6\alpha-\omega\alpha^2}\right)H_0\,(t-t_0)\right]^{\frac{1}{\left(\frac{3+\alpha}{2}+\frac{9b^2w_h}{6+6\alpha-\omega\alpha^2}\right)}}. \tag{29}$$

We shall investigate the evolution of the deceleration parameter, which is defined as

$$q=-\frac{a\ddot{a}}{\dot{a}^2}=-1-\frac{a}{H}\frac{dH}{da}. \tag{30}$$

Using (28) into (30), we obtain

$$q=\left(\frac{3+\alpha}{2}+\frac{9b^2w_h}{6+6\alpha-\omega\alpha^2}\right)-1. \tag{31}$$

For sake of completeness, we derive the effective EoS parameter, which is defined as $w_{eff} = -1-\frac{2a}{3H}\frac{dH}{da}$. Using (28), we obtain

$$w_{eff}=-1+\frac{2}{3}\left(\frac{3+\alpha}{2}+\frac{9b^2w_h}{6+6\alpha-\omega\alpha^2}\right). \tag{32}$$

From (29), (31) and (32), we observe that the expansion of the model is of power-law form, and the deceleration parameter and effective EoS parameter are constant. Therefore, the HDE model with Hubble horizon as an IR cut-off in BD theory does not show the phase transition. The HDE model will decelerate or accelerate through out the evolution depending on the value of parameters. Thus, the HDE model with Hubble horizon as an IR cut-off in BD theory is ruled out to fit the observational data (see Section 5 for details). In Refs. [49,50,76], the authors have also found such solutions. Many authors [77–80] proposed different IR cut-off like particle horizon, event horizon, Ricci scalar, etc., to discuss the accelerating universe. However, it is to be noted that the Hubble horizon is a natural choice of cosmological length scale. Therefore, instead of considering any other IR cut-off or interaction between the fluids as did by many authors, we restrict ourselves to Hubble horizon as an IR cut-off but in the presence of matter creation to realize an accelerated expansion of the universe.

4.2. Solution with $\Gamma \neq 0$

The purpose of review of the power-law solution of HDE model in previous section is to present our motivation to work with matter creation. In this subsection, let us study the evolution equation (26) of HDE model with matter creation. We assume the functional form of matter creation rate of a more general form [30,31,41] (hereafter, *HDE 2*)

$$\Gamma = 3\gamma H_0 + 3\beta H, \tag{33}$$

where the parameters γ and β lie on the interval [0, 1]. Inserting (33) into (26), one finds

$$H' + \frac{k_1}{a}H = \frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)}\frac{H_0}{a}, \tag{34}$$

where $k_1 = \frac{(\alpha+3(1-\beta))(6+6\alpha-\omega\alpha^2)+18(w_h+\beta)b^2}{2(6+6\alpha-\omega\alpha^2)}$ and a prime denotes derivative with respect to the scale factor a. The solution of (34) in terms of redshift z reads

$$H = H_0\left[\frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)k_1} + \left(1-\frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)k_1}\right)(1+z)^{k_1}\right]. \tag{35}$$

Using $H = \dot{a}/a$ and $1 + z = a^{-1}$, integration of the above expression, in normalized unit of the scale factor at present time, gives

$$a=\left[1+\frac{2(6+6\alpha-\omega\alpha^2)k_1}{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)} \times\left(e^{\frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)}H_0(t-t_0)}-1\right)\right]^{1/k_1}. \tag{36}$$

We observe that as $(t - t_0) \to 0$, $a \to [1 + k_1H_0(t - t_0)]^{1/k_1}$, which shows the power-law expansion in early time. As $(t-t_0) \to \infty$, the scale factor tends to $a \to e^{\frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)}H_0(t-t_0)}$, which corresponds to the expansion like de Sitter. Thus, the model has early deceleration phase followed by an acceleration phase in the later stage of the evolution.

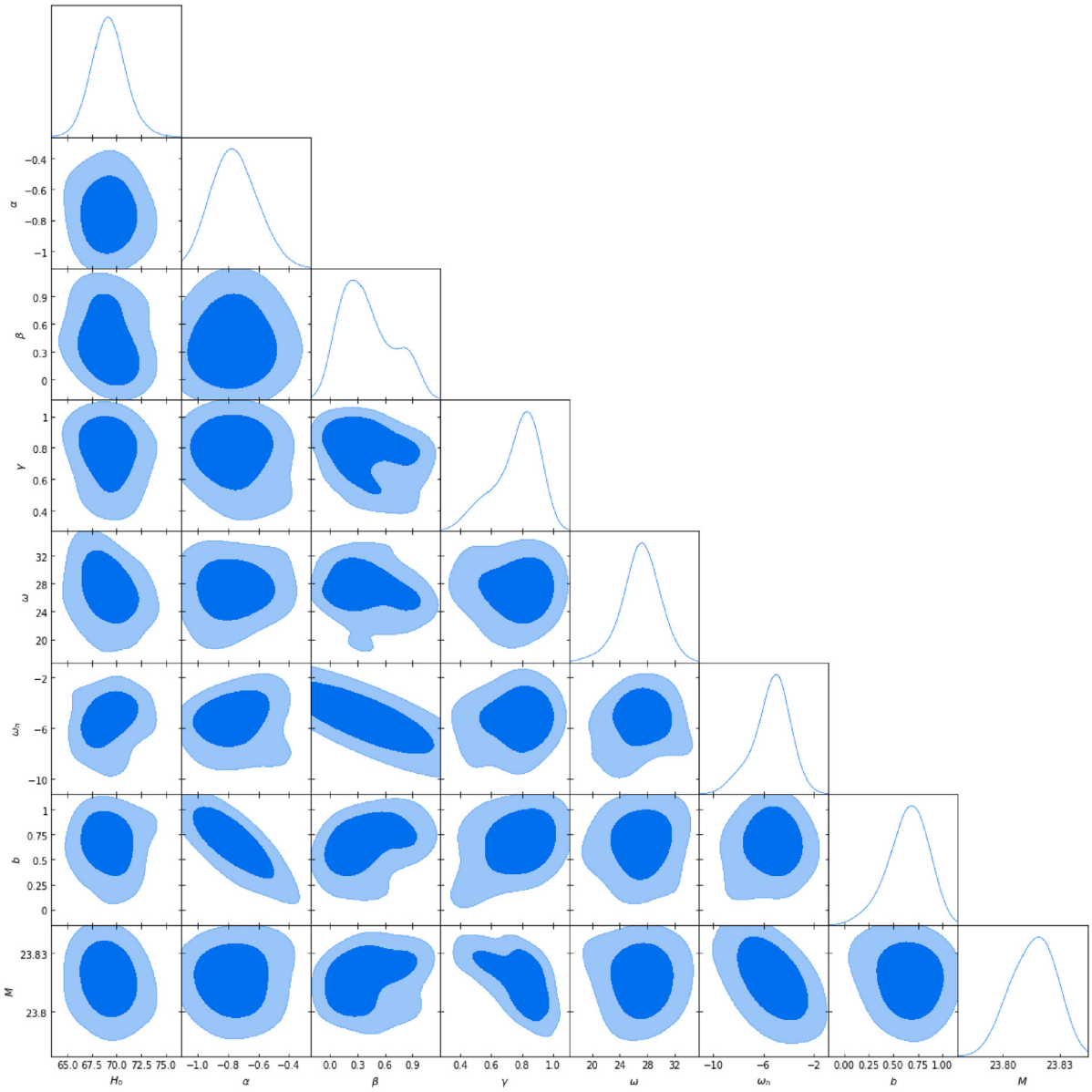


Fig. 1. The contour map of HDE2 model using DS1 data set with marginalized probability for the parameters.

The cosmic time $t(a)$ in terms of the scale factor is given by

$$t(a) - t_0 = \frac{2(6+6\alpha-\omega\alpha^2)H_0^{-1}}{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}ln \times \left(1+\frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)k_1}(a^{k_1}-1)\right) \quad (37)$$

Thus, the cosmic time t_{BB} when the Big-Bang $a(t_{BB}) = 0$ happens, is given by

$$t_{BB} = t_0 + \frac{2(6+6\alpha-\omega\alpha^2)H_0^{-1}}{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}ln \times \left(1-\frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)k_1}\right) \quad (38)$$

The age of the universe, i.e., the elapsed time between the time t_{BB} until the present time t_0, is given by

$$|t_0 - t_{BB}| = \left|\frac{2(6+6\alpha-\omega\alpha^2)H_0^{-1}}{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}ln \times \left(1-\frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)k_1}\right)\right|. \quad (39)$$

Using (35) into (30), the deceleration parameter in terms of redshift is given by

$$q(z) = -1 + \frac{1}{2(6+6\alpha-\omega\alpha^2)} \times \frac{\left(2(6+6\alpha-\omega\alpha^2)k_1 - 3\gamma(6+6\alpha-\omega\alpha^2-6b^2)\right)(1+z)^{k_1}}{\frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)k_1} + \left(1-\frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)k_1}\right)(1+z)^{k_1}}. \quad (40)$$

Thus, the parameter q is time-dependent, which means that the transition of the universe from deceleration to late-time acceleration is possible. One can obtain transition redshift by taking $q = 0$ in Eq. (40), which implies that

$$z_{tr} = -1 + \left[\frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{(k_1-1)\left\{2(6+6\alpha-\omega\alpha^2)k_1 - 3\gamma(6+6\alpha-\omega\alpha^2-6b^2)\right\}} \right]^{\frac{1}{k_1}}. \tag{41}$$

The present value of q at $z = 0$ is obtained as

$$q_0 = \frac{2(6+6\alpha-\omega\alpha^2)(k_1-1) - 3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)}. \tag{42}$$

Thus, the transition from decelerated to accelerated phase occurs at present for $2(6+6\alpha-\omega\alpha^2)(k_1-1) = 3\gamma(6+6\alpha-\omega\alpha^2-6b^2)$. At this value, the transition scale factor $a_{tr} = 1$ which is the value of scale factor to be assume at present. For $2(6+6\alpha-\omega\alpha^2)(k_1-1) > 3\gamma(6+6\alpha-\omega\alpha^2-6b^2)$, we have decelerated universe in the present time and for $2(6+6\alpha-\omega\alpha^2)(k_1-1) < 3\gamma(6+6\alpha-\omega\alpha^2-6b^2)$, we have an accelerated one today.

For this model, we get the effective EoS, given by

$$w_{eff} = -1 + \frac{H_0}{3H}\frac{[(3+\alpha-3\gamma)(6+6\alpha-\omega\alpha^2)+18(w_h+\gamma)b^2](1+z)^{k_1}}{(6+6\alpha-\omega\alpha^2)}. \tag{43}$$

The present effective EoS parameter when $H = H_0$, is given by

$$w_{eff}(z=0) = -1 + \frac{1}{3} \times \frac{(3+\alpha-3\gamma)(6+6\alpha-\omega\alpha^2)+18(w_h+\gamma)b^2}{(6+6\alpha-\omega\alpha^2)}. \tag{44}$$

The condition of EoS parameter, $w_{eff} < -1/3$ for acceleration of the present universe can be obtained as

$$3w_{eff}(z=0)+1 = (1+\alpha-3\gamma)(6+6\alpha-\omega\alpha^2)+18(w_h+\gamma)b^2 < 0. \tag{45}$$

Now, it is interesting to discuss in what sense this HDE model with matter creation behaves like an irreversible process. Let us consider the entropy behaviour of the model. In adiabatic matter creation the specific entropy (per particle) $\sigma = S/N$, where S and N are, respectively, the entropy of the dominant component and the corresponding number of particles, remains constant, i.e., $\dot{\sigma} = 0$. This implies that $\frac{\dot{S}}{S} = \frac{\dot{N}}{N}$. Since $N = na^3$, it follows from (4) and (33) that N increases as $N = N_0\ a^{3\beta}\ e^{3\gamma H_0(t-t_0)}$, where a is obtained in Eq. (36). The corresponding S can be obtained as $S = S_0\ a^{3\beta}\ e^{3\gamma H_0(t-t_0)}$.

Using (18), (19) and (25), one can check that the consistency condition (24) can be written as

$$2(\omega\alpha-3)\dot{H} + (\omega\alpha^2+6\omega\alpha-12)H^2 = 0 \tag{46}$$

Putting (35) in (46), we obtain

$$2(\omega\alpha-3)\,(X-k_1)\,a^{-k_1} + (\omega\alpha^2+6\omega\alpha-12) \times \left[\frac{X}{k_1} + \left(1-\frac{X}{k_1}\right)a^{-k_1}\right] = 0, \tag{47}$$

where $X = \frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)}$. The above equation cannot be always satisfied. So, we assume this equation is satisfied nowadays, i.e., when $a = a_0 = 1$, which gives a relation between the constants as

$$2(\omega\alpha-3)\,(X-k_1) + (\omega\alpha^2+6\omega\alpha-12) = 0. \tag{48}$$

5. Observational data and method

In this section, we consider most recent observational data sets to constrain free parameters of our models (*HDE 1* and *HDE 2*). Accordingly, we are able to check the consistency of HDE model with matter creation.

5.1. Pantheon Type Ia supernovae

We use the Pantheon sample, the latest compilation of Type Ia supernovae (SNe), comprising of 1048 data points in the redshift region $0.01 < z < 2.3$ [81].

The χ^2 function of the Pantheon SNe data is given by

$$\chi^2_{Pantheon} = \mu\ C^{-1}\mu^T \tag{49}$$

in which $\mu = \mu_i^{obs} - \mu^{th}$, where μ_i^{obs} is the observed distance modulus defined in Ref. [82]. The theoretical distance modulus μ^{th} is given by

$$\mu^{th} = 5\ \log_{10} D_L + \mathcal{M}, \tag{50}$$

where $\mathcal{M}$ is the nuisance parameter and D_L is the luminosity distance, given by

$$D_L = (1+z_{hel})\int_0^{z_{cmb}} \frac{dz'}{H(z',\theta)}, \tag{51}$$

where θ represents the set of model parameters, z_{hel} is the heliocentric redshift and z_{cmb} is the redshift of the CMB rest frame.

It should be noted that the covariance matrix C in (49) is the sum of the systematic covariance C_{sys} and statistical matrix D_{stat} having a diagonal component [81,83].

5.2. Hubble data

The observational Hubble parameter data can be divided into two classes: the $H(z)$ data extracted from BAO surveys and cosmic chronometric (CC) data. In general, to obtain $H(z)$ data from BAO observations, one has to model the redshift space distortions and assume an acoustic scale, which requires the assumption of a specific model, i.e., model dependent which cannot be used for constraining a cosmological model. On the other hand the CC data determined by using the most massive and passively evolving galaxies based on the "galaxy differential age" method are model-independent [84,85]. Therefore, we use 31 CC data points in our analysis covering the range $z\ \epsilon$ [0.07, 1.97].

The χ^2 for 31 Hubble measurement is given by

$$\chi^2_{CC} = \sum_{i=1}^{31} \frac{[H_{obs}(z_i) - H_{th}(z_i)]^2}{\sigma_i^2}, \tag{52}$$

where $H_{th}(z_i)$ and $H_{obs}(z_i)$ represents theoretical and observed values respectively and σ_i^2 the standard deviation of each $H_{obs}(z_i)$.

5.3. Local Hubble constant

We also include the recently measured local H_0 given by $H_0 = 73.5 \pm 1.4$ km s^{-1} Mpc^{-1} by SH0ES as reported in [86].

5.4. Joint analysis

In order to constrain the model parameters with the above datasets we perform a Bayesian Markov Chain Monte Carlo (MCMC) method. This method is based on the publicly available EMCEE package [87] for analysing and plotting the contours. In our calculation, we have minimized the chi-square for two combinations of data set, which we believe are helpful for better fit values. The first one is labelled *DS*1 and contains chi-square function as $\chi^2_{DS1} = \chi^2_{Pantheon} + \chi^2_{CC} + H_0$ and the second one, *DS*2 as $\chi^2_{DS2} = \chi^2_{Pantheon} + \chi^2_{CC}$.

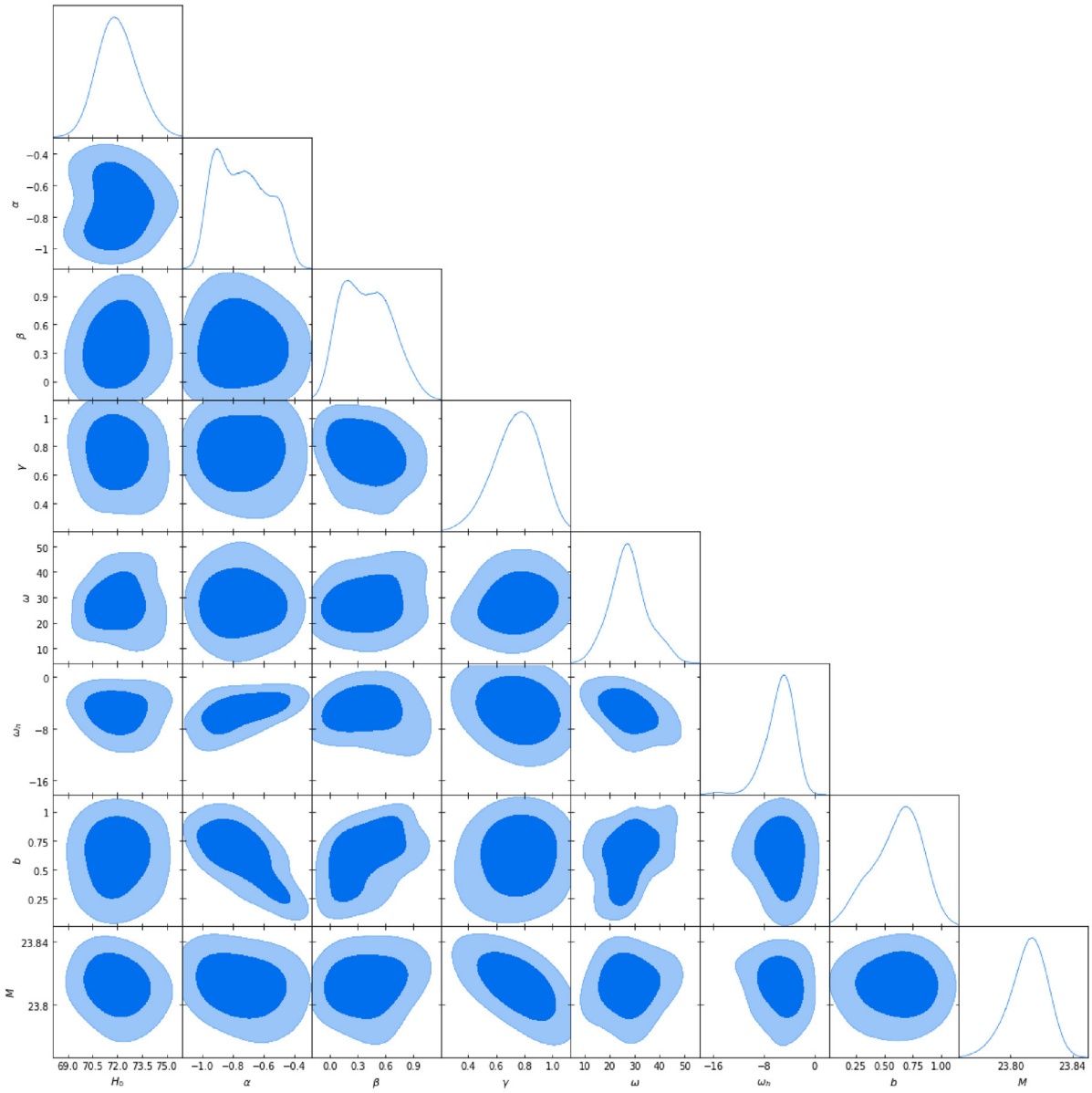


Fig. 2. The contour map of HDE2 model using DS2 data set with marginalized probability for the parameters.

Table 1
The best fit values for the considered models (HDE 1, HDE 2 and ΛCDM) using two datasets *DS*1, *i.e.*, *SNe*(*Pantheon*)+ $CC + H_0$ and *DS*2, *i.e.*, *SNe*(*Pantheon*) + *CC*.

Parameter	DS1			DS2		
	HDE1	HDE2	ΛCDM	HDE1	HDE2	ΛCDM
H_0	$71.903^{+1.360}_{-1.114}$	$71.926^{+1.222}_{-1.057}$	$72.525^{+1.274}_{-1.305}$	$68.964^{+2.645}_{-1.888}$	$68.957^{+1.540}_{-1.511}$	$69.030^{+2.196}_{-2.591}$
α	$0.032^{+0.397}_{-0.277}$	$-0.746^{+0.226}_{-0.189}$	–	$0.041^{+0.363}_{-0.316}$	$-0.763^{+0.137}_{-0.168}$	–
β	–	$0.380^{+0.266}_{-0.288}$	–	–	$0.347^{+0.460}_{-0.257}$	–
γ	–	$0.758^{+0.163}_{-0.167}$	–	–	$0.805^{+0.105}_{-0.207}$	–
ω	$12.967^{+12.398}_{-8.459}$	$27.495^{+7.058}_{-6.700}$	–	$13.212^{+6.218}_{-6.409}$	$27.136^{+2.537}_{-2.359}$	–
w_h	$-4.501^{+1.906}_{-2.547}$	$-5.153^{+1.804}_{-2.330}$	–	$-4.124^{+2.131}_{-3.346}$	$-5.101^{+1.057}_{-1.601}$	–
b	$0.345^{+0.153}_{-0.124}$	$0.648^{+0.171}_{-0.303}$	–	$0.333^{+0.183}_{-0.104}$	$0.667^{+0.204}_{-0.221}$	–
Ω_m	–	$0.279^{+0.032}_{-0.077}$	$0.284^{+0.021}_{-0.023}$	–	$0.255^{+0.1487}_{-0.1245}$	$0.300^{+0.032}_{-0.022}$
Ω_l	–	$0.720^{0.077}_{-0.032}$	$0.676^{+0.023}_{-0.018}$	–	$0.745^{0.124}_{-0.148}$	$0.697^{+0.041}_{-0.034}$

Table 2
Summary of the computed values of a_{tr}, z_{tr}, q_0, $w_{eff}(z=0)$ and t_0 coming from best fit results of HDE model with and without matter creation obtained from DS1 and DS2 data sets.

Model	DS1		DS2	
	HDE1	HDE2	HDE1	HDE2
a_{tr}	–	$0.557^{+0.201}_{-0.310}$	–	$0.549^{+0.047}_{-0.161}$
z_{tr}	–	$0.794^{+0.476}_{-1.09}$	–	$0.821^{+0.172}_{-2.402}$
q_0	$-0.264^{+0.456}_{-0.504}$	$-0.478^{+0.046}_{-0.103}$	$-0.140^{+0.066}_{-0.415}$	$-0.509^{+0.023}_{-0.185}$
$w_{eff}(z=0)$	$-0.509^{+0.304}_{-0.336}$	$-0.402^{+0.033}_{-0.023}$	$-0.427^{+0.44}_{-0.277}$	$-0.461^{+0.063}_{-0.029}$
t_0	$15.782^{+0.745}_{-0.406}$ Gyr	$13.67^{+0.558}_{-0.406}$ Gyr	$14.6018^{+0.676}_{-0.506}$ Gyr	$13.75^{+0.721}_{-0.036}$ Gyr

6. Results and discussion

In this section we present and discuss our results obtained in Bayesian analysis. Table 1 presents the mean fitting values obtained for HDE1, HDE2 and ΛCDM models using two data sets, DS1 and DS2. The reported uncertainties correspond to 1σ confidence level (CL). In our study, keeping due consideration of (48), we choose $-1 < \alpha < 1$, $0 < \beta < 1$, $0 < \gamma < 1$, $\omega > 0$, $-20 < \omega_h < 0$ and $0 < b < 1$. The contour maps of space parameters of HDE2 model for DS1 and DS2 data sets with $1\sigma(68.3\%)$ and $2\sigma(95.4\%)$ confidence level are shown in Figs. 1 and 2.

Using best-fit values, we find the matter density Ω_m and dark energy density Ω_l parameters for the HDE2 and ΛCDM models with DS1 and DS2 data sets which are listed in Table 1. The data set DS1 for HDE2 model yields the present density parameters $\Omega_m = 0.279^{+0.032}_{-0.077}$ and $\Omega_l = 0.720^{0.077}_{-0.032}$ where as these are $\Omega_m = 0.255^{+0.1487}_{-0.1245}$ and $\Omega_l = 0.745^{0.124}_{-0.148}$ with DS2 data set. It is observed that Ω_m is less and Ω_l is more than ΛCDM model for these two data sets.

In comparison with ΛCDM, we see that HDE1 and HDE2 models push the value of H_0 towards a smaller direction for DS1 and DS2. The H_0 tension between HDE models and local measurement value 73.5 ± 1.4 km s^{-1} Mpc^{-1} are mentioned in Table 1. We also notice the similar tension between ΛCDM model and local measurement value of the Hubble constant. Also, the value of the Hubble constant obtained from DS2 is close to the local value $H_0 = 69.8 \pm 0.8$ km s^{-1} Mpc^{-1} based on a calibration of the Tip of Red Giant Branch (TRGB) [88].

In what follows, we reconstruct the cosmological parameters, such as scale factor, deceleration parameter, effective equation of state parameter for both HDE models using best fit value obtained from two data sets, DS1 and DS2.

We calculate the corresponding deceleration–acceleration transition scale factor (redshift), the present value of deceleration parameter q_0 and effective EoS parameter $w_{eff}(z=0)$, and the present age of the universe t_0 for HDE1 and HDE2 models using DS1 and DS2 data sets which are listed in Table 2.

The evolutions of the scale factor of HDE1 and HDE2 models with the cosmic time are shown in Fig. 3. It is observed that the scale factor of HDE 1 model, for data sets DS1 and DS2, evolve from finite past big-bang singularity and expand with accelerated rate through out the evolution. Therefore, HDE1 model does not show phase transition. However, in case of HDE 2 model for DS1 and DS2, the scale factors start to evolve from finite past with decelerated rate and take the transition to accelerated phase at $a_{tr} = 0.557^{+0.201}_{-0.310}$ ($z_{tr} = 0.794^{+0.476}_{-1.099}$) and $a_{tr} = 0.549^{+0.047}_{-0.161}$ ($z_{tr} = 0.821^{+0.172}_{-2.402}$) respectively.

In Fig. 4, we plot the trajectory of deceleration parameter with redshift for HDE 1 and HDE2 models with their respective best-fit values obtained from DS1 and DS2. The HDE 1 has negative constant value of $q = -0.264^{+0.456}_{-0.504}$ and $q = -0.140^{+0.066}_{-0.415}$ through out the evolution for both DS1 and DS2, respectively.

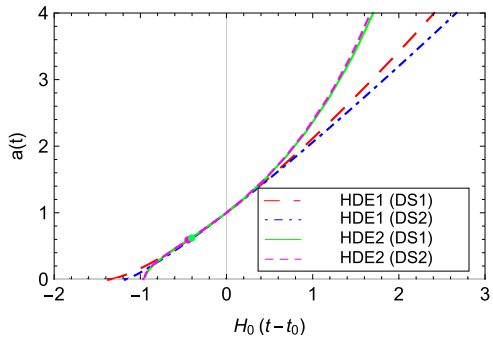


Fig. 3. Plot of the scale factor with respect to $H_0(t-t_0)$ for the best-fit value of the model parameters of HDE models using DS1 and DS2 data sets. A dot denotes the transition point where the transition from decelerated phase to accelerated phase occurs.

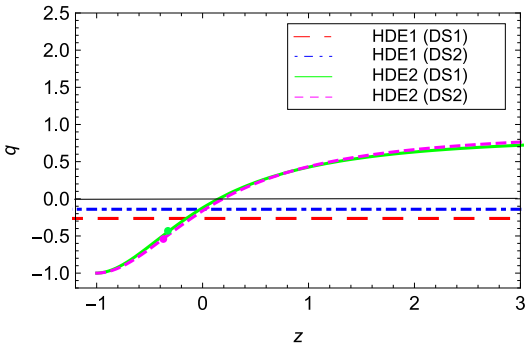


Fig. 4. The deceleration parameter as a function of redshift for best fit values of HDE model for DS1 and DS2 data sets. A dot denotes the current value of q (hence, q_0). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Hence, this model expands with accelerated rate through out the evolution. In HDE2 model, q shows a transition from decelerated phase (positive value of q) to accelerated phase (negative value of q) at transition redshifts $z_{tr} = 0.794^{+0.476}_{-1.099}$ and $z_{tr} = 0.821^{+0.172}_{-2.402}$, for DS1 and DS2, respectively. The corresponding current values of q are $q_0 = -0.478^{+0.046}_{-0.103}$ and $q_0 = -0.509^{+0.023}_{-0.185}$, respectively. It is to be noted that the latest observations indicate that $q_0 \simeq -0.63 \pm 0.12$ [89,90]. It is observed that $q \to -1$ in late time evolution of the universe with both the data sets.

The evolution of effective EoS parameter w_{eff} with redshift z is plotted in Fig. 5 for different models with their respective best fit values. The present values of ω_{eff} for HDE 1 and HDE 2 models are listed in Table 2. In HDE 1 model, w_{eff} is negative constant values, $w_{eff} = -0.509^{+0.304}_{-0.336}$ and $w_{eff} = -0.427^{+0.44}_{-0.277}$ for DS1 and DS2, respectively, which are less than $w_{eff} = -(1/3)$ showing the accelerated phase. The respective present values of w_{eff} for HDE2 at $z = 0$ are $w_{eff}(z=0) = -0.402^{+0.033}_{-0.023}$ and $w_{eff}(z =$

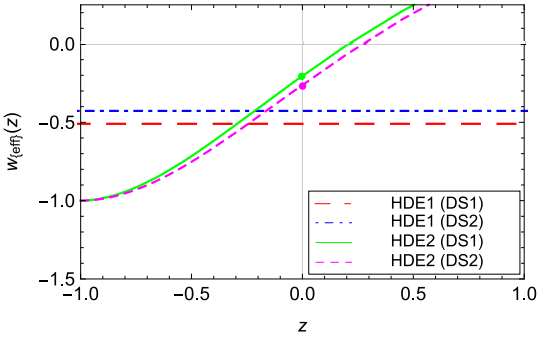


Fig. 5. The effective EoS parameter as a function of redshift for best fit values of HDE1 and HDE2 models using DS1 and DS2 data sets. A dot denotes the present value of w_{eff}. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

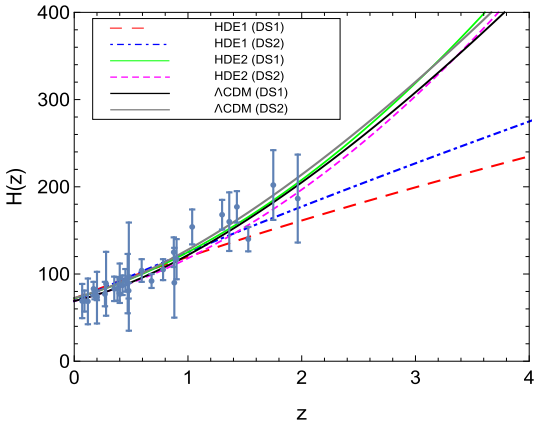


Fig. 6. The Hubble function as a function of the redshift z for the best fit values of HDE1 and HDE2 models. The observational 31 $H(z)$ data points are shown with error bars (grey colour). The variation of the Hubble function in the standard ΛCDM model is also represented as the Grey and black solid curves for DS1 and DS2 respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$0) = -0.461^{+0.063}_{-0.029}$. These values are comparatively higher than that is predicted by the joint analysis of $WMAP + BAO + H_0 + SNe$ data which is around $-0.93^{+0.13}_{-0.13}$ [91]. The data obtained in [91] is consistent with a flat universe dominated by a cosmological constant. In late time, w_{eff} of these two models approaches to -1 leading to Einstein–de-Sitter behaviour. It has been observed that these models do not cross the phantom-divide line $w_{eff} = -1$.

Fig. 6 plots the evolution of Hubble function versus redshift with the error bar of Hubble dataset in the range $z \in (0, 4)$ and compare HDE 1 and HDE 2 models with the ΛCDM model. It can be observed that for both the combination of data sets, HDE 1 do not fit with the error bar of Hubble data set at higher redshift. However, HDE 2 gives the best fit with ΛCDM model through out the evolution for both DS1 and DS2. The trajectories of these models pass through most of the data set of error bar of Hubble parameter.

In HDE 1, the current age of the universe are found to be $15.78^{+0.745}_{-0.406}$ Gyr and $14.60^{+0.676}_{-0.506}$ Gyr with DS1 and DS2, respectively, which are higher than the age predicted by ΛCDM model ($13.978^{+0.234}_{-0.140}$ Gyr for DS1 and $13.889^{+0.24}_{-0.103}$ Gyr for DS2) . However, in HDE 2, the current age of universe is calculated as $13.67^{+0.558}_{-0.406}$ Gyr and $13.75^{+0.721}_{-0.036}$ Gyr, for DS1 and DS2 respectively, which are agreeing with the age deduced from ΛCDM model.

The reduced χ^2, denoted by χ^2_{red}, is defined as $\chi^2_{red} = \chi^2_{min}/(N-d)$, where N and d are respectively the total number of data points used in the model and the number of free parameters of model, and their values are $N = 1080$ for DS1 (or 1079 for DS2) and $d = 3$ for ΛCDM, $d = 5$ for HDE1, $d = 7$ for HDE 2 models. The values of χ^2 and χ^2_{red} of ΛCDM and HDE models are given in Table 3. The HDE 1 and HDE 2 models have the values $\chi^2_{red} < 1$ for both the data sets DS1 and DS2. Thus, HDE 1 and HDE 2 are in good agreement with the observational data.

Table 3
Summary of χ^2, χ^2_{red}, *AIC* and *BIC* values and their differences from the reference model of ΛCDM obtained from DS1 and DS2 data sets.

Model	DS1			DS2		
	HDE1	HDE2	ΛCDM	HDE1	HDE2	ΛCDM
χ^2	50.45	30.19	27.67	64.88	28.31	27.43
χ^2_{red}	0.046	0.028	0.025	0.060	0.026	0.025
AIC	60.45	44.19	33.67	74.88	42.31	33.43
BIC	85.35	79.07	48.61	99.78	77.19	48.37
ΔAIC	26.78	10.52	0	41.45	8.88	0
ΔBIC	36.74	30.46	0	51.41	28.82	0

We shall now use Akaike information criterion (AIC) and Bayesian information criterion (BIC) as model selection criteria. The AIC [92] and BIC [93], for the large number of data points, are respectively defined as

$$AIC = \chi^2_{min} + 2d, \qquad BIC = \chi^2_{min} + d\ ln\ N, \tag{53}$$

where d is the number of free parameters and N is the number of data points used in the analysis. In these approaches, the model with lowest values of AIC (BIC) is preferred by data. We present the calculation results of AIC and BIC, and relative value between different models, ΔAIC and ΔBIC in Table 3. Since ΛCDM model is an important reference model for the study of DE models, we also include its results in Table 3. A model having $0 \leq \Delta AIC < 2$ and $0 \leq \Delta BIC < 2$ has "strong evidence in favour", for $2 < \Delta AIC < 4$ and for $2 \leq \Delta BIC < 6$, model has "average evidence in favour", for $4 < \Delta AIC < 7$ and $6 \leq \Delta BIC < 10$, model has "less evidence in favour", and for $\Delta AIC > 10$ or $\Delta BIC > 10$, the model has "no evidence in favour" [94].

It is observed from Table 3 that HDE 1 model is not supported over ΛCDM taking into account DS1 and DS2 data. However, HDE 2 is much better, though they are still less accurate than the ΛCDM model. Also, it has been found that the concordance ΛCDM model remains the best one to explain current data. Therefore, the HDE models are punished due to the large number of parameters, thus is not favoured by the Information criterion from a model selection point of view. On the other hand, the HDE models give good fit to the data from reduced chi-square point of view because it does not contain the information of the complexity as Information criterion have.

7. Diagnostic parameters

In this section, we discuss the geometric view of HDE model with matter creation through statefinder and *Om* parameters.

7.1. Statefinder parameter

Sahni et al. [95] proposed two dimensionless and model independent geometric parameters, known as statefinder parameters $\{r, s\}$, which are defined as

$$r = \frac{1}{aH^3}\frac{d^3a}{dt^3} = \frac{\ddot{H}}{H^3} - 3q - 2 \quad \text{and} \quad s = \frac{r-1}{3(q-1/2)}. \tag{54}$$

The above pair of parameters is used to discriminate the observationally supported DE models from other phenomenological DE models existing in literature. In a flat ΛCDM model, the

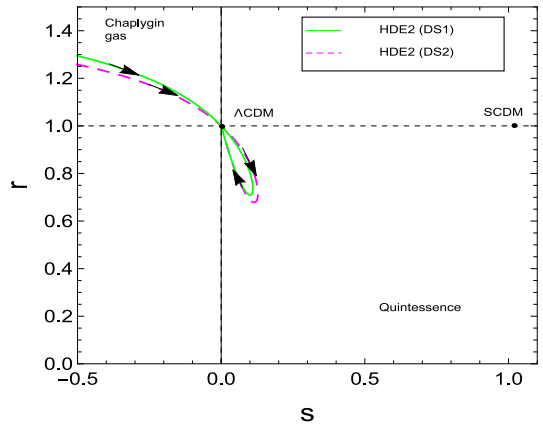


Fig. 7. The trajectory of $\{r, s\}$ in $s-r$ plane corresponds to best fitted parameters of HDE 2 model for DS1 and DS2. The arrow shows the direction of the evolution of the trajectory.

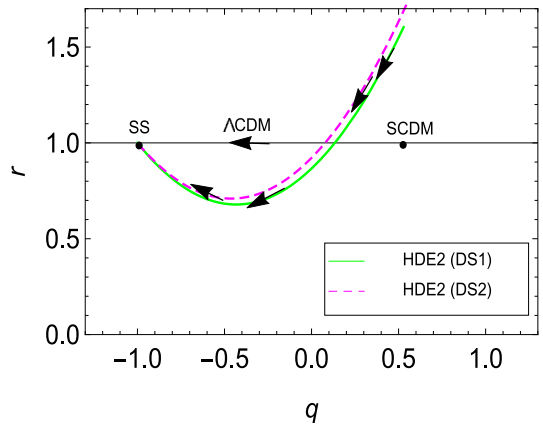


Fig. 8. The trajectory of $\{r, q\}$ in $q-r$ plane for the best fitted parameters of HDE 2 model for DS1 and DS2. The arrow shows the direction of the evolution of the trajectory.

statefinder pair has a fixed point value $\{r, s\} = \{1, 0\}$. One can plot the trajectories in $r - s$ and $r - q$ planes to discriminate various DE models.

The expressions for r and s can be obtained by using (35) and (40) into (54). It is observed that as $a \to \infty$, $\{r, s\} \to \{1, 0\}$ which coincide with the ΛCDM model. The $s - r$ plane trajectory of the HDE 2 models for their best fit values of parameters obtained by observational data sets are shown in Fig. 7. The direction of evolution of trajectory is shown by the arrows. The trajectories of both the models start from the Chaplygin gas region ($r > 1, s < 0$) at early time and in intermediate time pass through quintessence region ($r < 1, s > 0$) and then ultimately approach to ΛCDM in late time. The present position of the universe corresponds to $\{r_0, s_0\} = \{0.680, 0.108\}$ and $\{r_0, s_0\} = \{0.711, 0.095\}$ for HDE 2 model for DS1 and DS2, respectively, which are different from the ΛCDM model. Thus, our HDE model with matter creation is well discriminated from the ΛCDM model.

We have also plotted the $\{r, q\}$ trajectory of HDE 2 model for DS1 and DS2 as shown in Fig. 8. In Fig., the *SCDM* model and steady state (SS) model correspond to fixed points $\{r, q\} = \{1, 0.5\}$ and $\{r, q\} = \{1, -1\}$, respectively. The horizontal line at $r = 1$ corresponds to the time evolution of ΛCDM model. The trajectories show that both the models approach to SS model as the standard model like ΛCDM in late time [96].

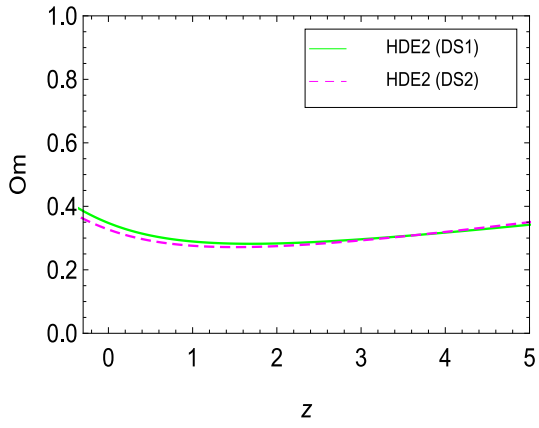


Fig. 9. The trajectory of $Om(z)$ for the best fit value of parameters of HDE 2 model for DS1 and DS2.

7.2. $Om(z)$ parameter

The Om is another diagnostic approach to distinguish different dark energy models. It is defined as [96]

$$Om(z) = \frac{h^2(z) - 1}{(1+z)^3 - 1}, \tag{55}$$

where $h = H/H_0$ is the dimensionless parameter. For ΛCDM model, the value of $Om(z)$ is a constant independent of the redshift. Therefore, if $Om(z)$ is variable, it possibly leads to an alternative dark energy or modified gravity model. The $Om(z)$ is sensitive to EoS of DE, namely, a positive slope of $Om(z)$ suggests a phase of phantom ($w < -1$) while a negative slope of $Om(z)$ represents quintessence ($w > -1$).

Using (35) into (55), we can write the expression of $Om(z)$ for HDE 2 as

$$Om(z) = \frac{\left[-1 + \left(\frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)k_1} + (1 - \frac{3\gamma(6+6\alpha-\omega\alpha^2-6b^2)}{2(6+6\alpha-\omega\alpha^2)k_1})(1+z)^{k_1}\right)^2\right]}{(1+z)^3 - 1}, \tag{56}$$

Fig. 9 shows the evolution of $Om(z)$ with respect to the redshift z for HDE 2 model for DS1 and DS2, corresponding to different best-fit values of model parameters. The negative slope of each trajectory shows that the model behaves like quintessence.

8. Thermodynamics analysis

This section provides the thermodynamics analysis of the models presented in the previous sections. The validity of a cosmological model is tested from its thermodynamical behaviour. Many authors [41,42,97–102] have studied the relation between gravity and thermodynamics. In Ref. [103], it has been demonstrated that cosmological apparent horizons are also endowed with thermodynamic properties. It can relate temperature and entropy to the apparent horizon like to the black hole event horizon.

Macroscopic systems tend towards a thermodynamical equilibrium where, according to the generalized second law (GSL) of thermodynamic, the total entropy $S = S_{hor} + S_m + S_h$ of an isolated system never decreases, i.e., $\dot{S}_{hor} + \dot{S}_m + \dot{S}_h \geq 0$. Here, S_{hor} is the entropy of the apparent horizon, S_m and S_h are the entropy of dark matter and HDE, respectively enclosed by the horizon. The entropy of apparent horizon is defined as $S_{hor} = \kappa_B A/4l_{Pl}^2$ [104], where κ_B is the Boltzmann's constant, $A = 4\pi r_{hor}^2$ is the horizon

area in which $r_{hor} = H^{-1}$ is the horizon radius for flat FRW universe and $l_{pl} = 1/\sqrt{8\pi}M_{Pl}$ is the Planck's length.

Using (35), we obtain

$$\dot{S_{hor}} = -\frac{2\pi\kappa_B}{l_{pl}^2}\frac{\dot{H}}{H^3} = \frac{2\pi\kappa_B H_0}{l_{pl}^2 H^2}\frac{a^{-k_1}}{2(6+6\alpha-\omega\alpha^2)}\times$$
$$[(3(1-\beta-\gamma)+\alpha)(6+6\alpha-\omega\alpha^2)+18b^2(w_h+\beta+\gamma)]. \quad (57)$$

Now, we need to study the entropy of fluids S_f (dark matter and HDE), which arises from Gibb's equation [105] for the fluid

$$TdS_f = d(\rho V) + pdV = Vd\rho + \rho dV + pdV, \quad (58)$$

where $V = 4\pi r_h^3/3$ is the spatial volume enclosed by the horizon and T is the fluid temperature.

The "gravity- thermodynamics" conjecture says that the temperature and the entropy of horizon can be interpreted by the corresponding quantities of black-hole thermodynamics, but with the apparent horizon in place of the black-hole one. According to this conjecture, the universe background geometry becomes FRW once the equilibrium establishes. At this position, all the fluids in the universe acquire the same temperature T [54], which is moreover equal to the temperature of the horizon T_{hor} [106], otherwise the energy flow would deform the geometry of flat FRW [107]. However, this happens when the universe fluids and the horizon will have interacted for a long time in late time. In general, two systems must interact for some length of time before they can attain thermal equilibrium. The assumption of equilibrium is widely accepted in the GSL literature [107–109]. We can see that this choice is also backed by the holographic principle [46–48]. Therefore, we follow this assumption keeping in mind that the results hold at late times of the evolution of the universe. Thus, under the thermal equilibrium condition between the fluids and horizon, we have $T = T_{hor}$, where $T_{hor} = 1/2\pi r_{hor} = H/2\pi$, which is equal to the Hawking temperature of the horizon. Using (35), the above equation for HDE gives

$$\dot{S_h} = \frac{24\pi^2\rho_h(1+w_h)H_0}{H^4}\Big[\frac{a^{-k_1}}{2(6+6\alpha-\omega\alpha^2)}$$
$$\times\Big\{(3(1-\beta-\gamma)+\alpha)(6+6\alpha-\omega\alpha^2)$$
$$+\ 18b^2(w_h+\beta+\gamma)\Big\} - \frac{H}{H_0}\Big] \quad (59)$$

From (58), the change of entropy for pressureless dark matter is given by

$$\dot{S_m} = \frac{24\pi^2\rho_m H_0}{3H^4}\Big[3\gamma + \frac{a^{-k_1}}{2(6+6\alpha-\omega\alpha^2)}$$
$$\times\Big\{(3(1-\beta-\gamma)+\alpha)(6+6\alpha-\omega\alpha^2)$$
$$+\ 18b^2(w_h+\beta+\gamma)\Big\} - (1-\beta)\frac{H}{H_0}\Big] \quad (60)$$

From (57), (59) and (60), it is observed that $\dot{S_{hor}}$, $\dot{S_h}$ and $\dot{S_h}$ would be positive for $0 < (\beta+\gamma) < 1$ and $(6+6\alpha-\omega\alpha^2) > 0$ in an expanding universe. Thus, the total change in entropy is always positive with this best fit value. This indicates that the GSL of thermodynamics is satisfied for HDE model with Hubble horizon as IR cutoff in the presence of matter creation in BD theory.

9. Conclusion

In this paper, we have carried out a study on matter creation dominated HDE model with Hubble horizon as an IR cut-off in BD theory. We have studied two different HDE models with $\Gamma = 0$ and $\Gamma = 3\gamma H_0 + 3\beta H$, respectively. We have presented the exact solution for these HDE models using a power-law relation between the BD scalar field and the scale factor. It can be observed that the HDE model gives the constant deceleration parameter in the absence of matter creation ($\Gamma = 0$). However, the HDE model, with suitable choice of Γ gives the time-dependent deceleration parameter, which describes the phase transition from decelerated phase to accelerated phase.

We have constrained the model parameters by using two latest data sets DS1: *SNe*(*Pantheon*)+*CC*+H_0 and DS2: *SNe*(*Pantheon*)+*CC*. The best fit values of parameters for both HDE models have been given in Table 1. We have presented a comparative study of these models by studying the evolution of cosmological parameters through plotting the graphs using the best fit parameters. We summarize the main results as follows.

It has been found that the HDE 1 ($\Gamma = 0$) model with Hubble horizon as an IR cut-off is not able to describe the transition of the universe. In this model, the scale factor varies as power-law form, and q and w_{eff} are constant. These show that the model may decelerate or accelerate through out the evolution depending upon the conditions applied. Using best fit values obtained from DS1 and DS2 in the analytical solutions, we have obtained $q = -0.264^{+0.456}_{-0.504}$ and $q = -0.140^{+0.066}_{-0.415}$, respectively where as the respective effective EoS are $w_{eff} = -0.509^{+0.304}_{-0.336}$ and $w_{eff} = -0.427^{+0.44}_{-0.277}$. These observations show that this model expands with accelerated rate for both the combination of data sets. The evolution of the Hubble functions of this model is far from evolution of ΛCDM model and Hubble data points. The age of the universe are found to be $15.78^{+0.745}_{-0.406}$ Gyr and $14.601^{+0.676}_{-0.506}$ Gyr, respectively which are very high from the predicted value. The respective reduced χ^2 are 0.046 and 0.06 which are less than $\chi^2_{red} = 0.025$ and $\chi^2_{red} = 0.025$ of ΛCDM model, respectively. We have also studied the model selection criterion for this model and have found the values ΔAIC = 26.78 and ΔBIC = 36.74 from DS1, and ΔAIC = 41.45 and ΔBIC = 51.41 from DS2. Thus, we do not have support for the HDE1 model over ΛCDM. The statefinder parameters $\{r, s\}$ of this model for both data sets are found to be constant,i.e., $\{r, s\} = \{-0.101, 0.572\}$ and $\{r, s\} = \{-0.124, 0.490\}$, respectively, which show the quintessence behaviour.

To consider the phase transition we have introduced matter creation in HDE model with the same IR cut-off in BD theory. We have assumed the functional form of matter creation rate as $\Gamma = 3\gamma H_0 + 3\beta H$. The best fit values of HDE 2 model obtained from data sets DS1 and DS2 have been given in Table 1. Using best fit values, we have plotted the trajectory of various cosmological parameters and have discussed the evolution accordingly. We have observed that the both the models describe a universe having Big-bang at finite past, then entering an early stage of decelerated expansion followed by acceleration in the later stage of the evolution as shown in Fig. 3. The transitions from the decelerated to accelerated expansion in HDE 2 model occur at $a_{tr} = 0.557^{+0.201}_{-0.310}$ (or $z_{tr} = 0.794^{+0.476}_{-1.099}$) and $a_{tr} = 0.549^{+0.047}_{-0.161}$ (or $z_{tr} = 0.821^{+0.172}_{-2.402}$), respectively which are higher than the ΛCDM value ($z_{tr} = 0.66$). We have found a time-dependent deceleration parameter, which shows that q transits from positive to negative values corresponds to the transition from deceleration to acceleration in the expansion of the universe (Fig. 4). From the observational results, the present values of q for HDE 2 using DS1 and DS2 are $q_0 = -0.478^{+0.046}_{-0.103}$ and $q_0 = -0.509^{+0.023}_{-0.185}$, respectively, which are higher than the current observational value $q_0 = -0.64 \pm 0.03$ [110].

The evolutions of effective EoS parameter of both the data sets of HDE2 model are shown in Fig. 5. It has been observed that as $z \to -1$ ($a \to \infty$), $w_{eff} \to -1$ in future, which corresponds to the de Sitter universe. Since it does not cross the phantom divide line,

both the model is free from Big-rip singularity. Using the best fit values obtained from DS1 and DS2, the present values of effective EoS for HDE2 model are found to be $w_{eff}(z=0) = -0.402^{+0.033}_{-0.023}$ and $w_{eff}(z=0) = -0.461^{+0.063}_{-0.029}$, respectively, which are higher than the current value $-0.93^{+0.13}_{-0.13}$ [91]. The evolutions of Hubble functions is shown in Fig. 6. The ages of universe for this model with data sets DS1 and DS2 have been found as $t_0 \sim 13.67$ Gyr and $t_0 \sim 13.75$ Gyr, respectively which are in very good agreement with the age predicted by ΛCDM model ~ 13.97 Gyr and ~ 13.88 Gyr, respectively.

We have calculated the χ^2_{red} of ΛCDM and HDE models using the best fit values (See, Table 2). It has been found that the χ^2_{red} of these two models are very close to ΛCDM model. We have also studied the AIC and BIC to test the stability of the models. It has been found that the concordance ΛCDM model remains the best one to explain current data. The HDE model with matter creation are punished by AIC and BIC mainly because of large number of parameters and thus is not favoured by the current joint data set from a model selection point of view. It is to be noted that AIC and BIC strongly penalize the models when they have large number of free parameters. However, the HDE models with matter creation give good fit to the data from reduced chi-square point of view as it does not contain the information of the complexity.

We have studied the geometrical diagnostic, like statefinder parameters and Om to compare the HDE 2 model with other existing DE models, like ΛCDM model. We have found that as $a \to \infty$ ($z \to -1$), $\{r, s\}$ and $\{r, q\}$ tend to $\{1, 0\}$ and $\{1, -1\}$, respectively, which are in good agreement with ΛCDM model. Using best fit values from DS1 and DS2, the current values of $\{r, s\}$ for this model are $\{r_0, s_0\} = \{0.680, 0.108\}$ and $\{r_0, s_0\} = \{0.711, 0.095\}$, respectively, which show deviation from the ΛCDM model, and thus, the models are well discriminated from ΛCDM model. The evolution of Om shows that the model behaves like quintessence. Using AIC and BIC with data obtained from both the combination of data sets, we compare the HDE model with ΛCDM model. It has been found that the model has less evidence in favour.

In the last section we have studied the thermodynamics of generalized second law in the present HDE2 model. It has been found that the generalized second law of thermodynamics is valid with apparent horizon as the boundary.

We therefore conclude that the present HDE model with suitable form of matter creation is capable of producing acceleration of the universe in BD theory with Hubble horizon as an IR cut-off.

CRediT authorship contribution statement

Simran Kaur: Observatons and figures, Investigation, Methodology. **C.P. Singh:** Conceptualization, Investigation, Methodology.

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.

Acknowledgments

The authors express their thank to the reviewer for fruitful comments and suggestions that have significantly improved our work in terms of research quality. Authors acknowledge fruitful discussions with Anil Kumar Yadav, Department of Physics, United College of Engineering and Research, Greater Noida, India. One of the authors, SK would like to thank CSIR, New Delhi for providing Senior Research Fellowship under CSIR(SRF) Scholarship.

References

[1] A.G. Riess, et al., Astron. J. 116 (1998) 1009.
[2] S. Perlmutter, et al., Astrophys. J. 517 (1999) 565.
[3] W. Wood-Vasey, et al., Astrophys. J. 666 (2007) 694, [astro-ph/0701041].
[4] M. Kowalski, et al., Astrophys. J. 686 (2008) 749, [arXiv:0804.4142].
[5] M. Hicken, et al., Astrophys. J. 700 (2009) 1097, [arXiv:0901.4804].
[6] A. Conley, et al., Astrophys. J. Suppl. 192 (2011) 1, [arXiv:1104.1443].
[7] R. Amanullah, et al., Astrophys. J. 716 (2010) 712, [arXiv:1004.1711].
[8] C. Blake, et al., Mon. Not. R. Astron. Soc. 418 (2011) 1707, [arXiv:1108.2635].
[9] W.J. Percival, et al., Mon. Not. R. Astron. Soc. 401 (2010) 2148, [arXiv:0907.1660].
[10] S. Weinberg, Rev. Modern Phys. 61 (1989) 1.
[11] P.J. Peebles, B. Ratra, Rev. Modern Phys. 75 (2003) 559.
[12] T. Padmanabhan, Phys. Rep. 380 (2003) 235.
[13] Y.B. Zeldovich, JETP Lett. 12 (1970) 307.
[14] L. Parker, Phys. Rev. D 3 (1971) 346.
[15] I. Prigogine, J. Geheniau, Proc. Natl. Acad. Sci. 83 (1986) 6245.
[16] I. Prigogine, J. Geheniau, E. Gunzig, P. Nardone, Proc. Natl. Acad. Sci. 85 (1988) 7428.
[17] I. Prigogine, et al., Gen. Relativ. Gravit. 21 (1989) 8.
[18] M.O. Calvao, J.A.S. Lima, I. Waga, Phys. Lett. A 162 (1992) 223.
[19] J.A.S. Lima, A.S. Germano, Phys. Lett. A 170 (1992) 373.
[20] R. Silva, J.A.S. Lima, M.O. Calvão, Gen. Relativ. Gravit. 34 (2002) 865.
[21] L.R.W. Abramo, J.A.S. Lima, Classical Quantum Gravity 13 (1996) 2953.
[22] J.A.S. Lima, A.S. Germano, L.R.W. Abramo, Phys. Rev. D 53 (1996) 4287.
[23] W. Zimdahl, J. Triginer, D. Pavón, Phys. Rev. D 54 (1996) 6101.
[24] V.B. Johri, K. Desikan, Astrophys. Lett. Commun. 33 (1996) 287.
[25] J.A.S. Lima, J.S. Alcaniz, Astron. Astrophys. 348 (1999) 1.
[26] J.S. Alcaniz, J.A.S. Lima, Astron. Astrophys. 349 (1999) 72.
[27] W. Zimdahl, D.J. Schwarz, A.B. Balakin, D. Pavón, Phys. Rev. D 64 (2001) 063501.
[28] C.P. Singh, A. Beesham, Astrophys. Space Sci. 336 (2011) 469.
[29] C.P. Singh, Astrophys. Space Sci. 338 (2012) 411.
[30] J.A.S. Lima, F.E. Silva, R.C. Santos, Classical Quantum Gravity 25 (2008) 205006.
[31] G. Steigman, R.C. Santos, J.A.S. Lima, J. Cosmol. Astropart. Phys. 06 (2009) 033.
[32] J.A.S. Lima, J.F. Jesus, F.A. Oliveira, J. Cosmol. Astropart. Phys. 11 (2010) 027.
[33] S. Basilakos, J.A.S. Lima, Phys. Rev. D 82 (2010) 023504.
[34] J.A.S. Lima, S. Basilakos, F.E.M. Costa, Phys. Rev. D 86 (2012) 103534.
[35] J.A.S. Lima, L.L. Graef, D. Pavón, S. Basilakos, J. Cosmol. Astropart. Phys. 10 (2014) 042.
[36] R.O. Ramos, M.V.D. Santos, I. Waga, Phys. Rev. D 89 (2014) 083524.
[37] S. Pan, S. Chakraborty, Adv. High Energy Phys. 201 (2015) 654025.
[38] R.C. Nunes, D. Pavón, Phys. Rev. D 91 (2015) 063526.
[39] J. de Haro, S. Pan, Classical Quantum Gravity 33 (2016) 165007.
[40] S. Pan, J. de Haro, A. Paliathanasis, R.J. Slagter, Mon. Not. R. Astron. Soc. 460 (2016) 1445.
[41] L.A. Salo, J. de Haro, Classical Quantum Gravity 34 (2017) 065001.
[42] S. Pan, B.K. Pal, S. Pramanik, Int. J. Geom. Methods Mod. Phys. 15 (2018) 1850042.
[43] C.P. Singh, S. Kaur, Phys. Rev. D 100 (2019) 084057.
[44] C.P. Singh, A. Kumar, Eur. Phys. J. C 80 (2020) 106.
[45] A. Kumar, C.P. Singh, Astrophys. Space Sci. 365 (2020) 84.
[46] G. t'Hooft, arXiv:gr-qc/9310026.
[47] L. Susskind, J. Math. Phys. 36 (1995) 6377.
[48] R. Bousso, Classical Quantum Gravity 17 (2000) 997.
[49] S.D. Hsu, Phys. Lett. B 594 (2004) 13.
[50] M. Li, Phys. Lett. B 603 (2004) 1.
[51] H. Wei, S.N. Zhang, Phys. Rev. D 76 (2007) 063003, arXiv:0707.2129.
[52] L. Xu, W. Li, J. Lu, Eur. Phys. J. C 60 (2009) 135.
[53] M.R. Setare, Phys. Lett. B 642 (2006) 421.
[54] M.R. Setare, Phys. Lett. B 654 (2007) 1.
[55] C.P. Singh, P. Kumar, Astrophys. Space Sci. 361 (2016) 157.
[56] C. Brans, R.H. Dicke, Phys. Rev. 124 (1961) 925.
[57] O. Bertolami, P.J. Martins, Phys. Rev. D 61 (2000) 064007.
[58] H. Kim, Mon. Not. R. Astron. Soc. Lett. 364 (2005) 813.
[59] T. Clifton, J.D. Barrow, Phys. Rev. D 73 (2006) 104022.
[60] N. Banerjee, D. Pavón, Phys. Lett. B 647 (2007) 477.
[61] X.L. Liu, X. Zhang, Commun. Theor. Phys. 52 (2009) 761.
[62] A. Sheykhi, Phys. Lett. B 681 (2009) 205.
[63] A. Sheykhi, Phys. Rev. D 81 (2010) 023525.
[64] P. Kumar, C.P. Singh, Astrophys. Space Sci. 362 (2017) 52.
[65] C.P. Singh, P. Kumar, Int. J. Thoer. Phys. 56 (2017) 3297.
[66] F. Felegary, F. Darabi, M.R. Setare, Internat. J. Modern Phys. D 27 (2018) 1850017.
[67] M. Srivastava, C.P. Singh, Int. J. Geom. Methods Mod. Phys. 15 (2018) 1850124.

[68] L. Xu, W. Li, Eur. Phys. J. C 60 (2009) 135.
[69] S. Weinberg, Gravitation and Cosmology, John Wiley and Sons, Inc., 1972, pp. 619–631, Ch.16.
[70] H. Kim, H.W. Lee, Y.S. Myung, Phys. Lett. B 628 (2005) 11.
[71] J. Solà Peracaula, Internat. J. Modern Phys. D 27 (2018) 1847029.
[72] J. de Cruz Pérez, J. Solà Peracaula, Modern Phys. Lett. A 33 (2018) 1850228.
[73] B. Bertotti, L. Iess, P. Tortora, Nature 425 (2003) 374.
[74] V. Acquaviva, L. Verde, J. Cosmol. Astropart. Phys. 0712 (2007) 001.
[75] M. Akbar, R.G. Cai, Phys. Rev. D 75 (2007) 084003.
[76] S. Basilakos, J. Solà, Phys. Rev. D 90 (2014) 023008, [arXiv:1402.6594].
[77] C. Gao, F. Wu, X. Chen, Phys. Rev. D 79 (2009) 043511.
[78] X. Zhang, F.Q. Wu, Phys. Rev. D 72 (2005) 043524.
[79] B. Chen, et al., Nuclear Phys. B 774 (2007) 256.
[80] M. Li, X.D. Li, S. Wang, X. Zhang, J. Cosmol. Astropart. Phys. 0906 (2009) 036.
[81] D.M. Scolnic, et al., Astrophys. J. 859 (2018) 101.
[82] M. Betoule, et al., Astron. Astrophys. 568 (2014) A22.
[83] A. Conley, et al., Astrophys. J. Suppl. Ser. 192 (2011) 1.
[84] M. Moresco, et al., J. Cosmol. Astropart. Phys. 05 (2016) 014.
[85] D. Wang, X. Meng, Phys. Rev. D 95 (2017) 023508.
[86] M.J. Reid, D.W. Pesce, A.G. Riess, Astrophys. J. 886 (2) (2019) L27.
[87] D. Foreman-Mackey, D. Hogg, D. Lang, J. Goodman, Publ. Astron. Soc. Pac. 125 (2012) 306.
[88] W.I. Freeman, et al., Astrophys. J. 882 (2019) 34.
[89] N. Aghanim, et al., Astron. Astrophys. 596 (2016) A107.
[90] R. Adam, et al., Astron. Astrophys. 596 (2016) A108.
[91] E. Komatsu, et al., WMAP Collaboration, Astrophys. J. Suppl. 192 (2011) 18.
[92] H. Akaike, IEEE Trans. Automat. Control 19 (1974) 716.
[93] G. Schwarz, Ann. Statist. 6 (1978) 461.
[94] A.R. Liddle, Mon. Not. R. Astron. Soc. 377 (2007) 74.
[95] V. Sahni, T.D. Saini, A.A. Starobinsky, U. Alam, JETP Lett. 77 (2003) 201.
[96] V. Sahni, A. Shafieloo, A.A. Starobinsky, Phys. Rev. D 78 (2008) 103502.
[97] J.D. Bekenstein, Phys. Rev. D 7 (1973) 2333.
[98] J.D. Bekenstein, Phys. Rev. D 9 (1974) 3292.
[99] S. Hawking, Comm. Math. Phys. 43 (1975) 199.
[100] G.W. Gibbons, S.W. Hawking, Phys. Rev. D 15 (1977) 2738.
[101] T. Jacobson, Phys. Rev. Lett. 75 (1995) 1260.
[102] T. Padmanabhan, Phys. Rep. 406 (2005) 49.
[103] R.G. Cai, L.M. Cao, Y.P. Hu, Classical Quantum Gravity 26 (2009) 155018.
[104] D. Bak, S.J. Rey, Classical Quantum Gravity 17 (2000) L83.
[105] G. Izquierdo, D. Pavón, Phys. Lett. B 633 (2006) 420.
[106] M. Akbar, R.G. Cai, Phys. Rev. D 75 (2007) 084003.
[107] G. Izquierdo, D. Pavón, Phys. Lett. B 639 (2006) 1.
[108] T. Padmanabhan, Rep. Progr. Phys. 73 (2010) 046901.
[109] M. Jamil, E.N. Saridakis, M.R. Setare, J. Cosmol. Astropart. Phys. 10 (2010) 032.
[110] M.G. Hu, X.H. Meng, Phys. Lett. B 635 (2006) 186.

View publication stats

Journal of Engg. Research ICARI Special Issue pp. 48-54 DOI : 10.36909/jer.ICARI.15319

CVD Diamond Coated Tungsten Carbide (WC) Tool Inserts

Ramakant Rana*, Lucky Krishnia**, R.S. Walia*** and Qasim Murtaza*

** Department of Mechanical Engineering, Delhi Technological University, Delhi, India*
*** Amity Center of Nanotechnology, Amity University Haryana -122413, India*
**** Department of Production and Industrial Engineering, Punjab Engineering College, Chandigarh, India*
*** Corresponding Author : luckykrishnia.research@gmail.com*

ABSTRACT

In this paper a self-developed polycrystalline diamond coating was done on tungsten carbide (WC) tool insets by using simple thermal chemical vapor deposition technique. The growth of these diamond films has been carried out at ~900 ºC temperature. The as-grown polycrystalline diamond films on the surface of tungsten carbide tool inserts have been characterized using Raman spectrometer and scanning electron microscope (SEM). The morphological studies reveal that the as-grown diamond films are of high crystalline quality. The as-grown diamond films possess compressive stress. The micro-hardness indentation test of the as-grown diamond films on WC tool inserts and bare have also been done and it has been found that the Vicker's hardness of the as-grown diamond WC tool inserts is found to be 1423.32 HV which is 29% better than the un-coated tools.

Keywords: Tungsten carbide; Diamond coating; Wear; Friction; Tool inserts.

INTRODUCTION

In wide range of machining applications tungsten carbide is used as tool inserts material. Tungsten carbide is highly suitable for very difficult to cut materials, alloys of aluminium-silicon, cast iron, and metal-matrix composites (Köpf et al., 2001). Diamond is known for its highest hardness among all the naturally known and available materials, but it also has plenty of other properties like Chemical inertness, low coefficient of friction, high resistance to the wear, along with high thermal conductivity. Because of all these properties diamond has been the center of attraction for cutting tools for the application in machining industries. The most commonly used material in machining industry for cutting tool is tungsten carbide (WC) (Jeyapandiarajan and Xavior, 2019, Rana et al., 2021).

Diamonds have unique set of properties which makes it ideal for protective coating on the cutting tools (Gäbler et al., 2000). The majority used cutting tool substrate material in the machining of plastics, aluminium composite and fiberglass is Tungsten Carbide (WC) (Dai et al., 2000, Ullen et al., 2020). Deposition of thin polycrystalline diamond films on WC tool insets would significantly improve the various properties viz. wear resistance, tool life, cutting performance and precision machining and will reduce machining cost, (Kuo et al., 1996, Endler et al., 1999).

Chemical Vapour Deposition (CVD) is the most commonly used approach for depositing the diamond coatings directly onto the substrates (Hintermann, 1996, Mallika and Komanduri, 1999). An effective and efficient way for improving the physical properties of tungsten carbide substrate WC + 6 wt.% Co (WC6) is to coat it with diamond surface layers using Chemical vapour deposition (CVD). Occasional poor adhesion of coating to the selected substrate

material and the development of cracks within either the interface or the coating, has always been the major obstacle while tapping the full potential of the chemical, mechanical and tribological properties of CVD diamond layers. Although the graphitization situated at the interface is induced by the binder phase Co which, has been the major reason behind the terrible untimely failure of adhesion of the developed diamond coatings on carbide tools (Chen and Narayan, 1993). Well defined compressive stresses compensate external thermal and mechanical loads during the service, which forestall the generation and propagation of cracks in the deposited diamond coatings. But then, the film delamination will occur if the compressive internal stresses become high (Nesladek et al., 1995, Gunnars and Alahelisten, 1996]. As mentioned by researchers, the chemical vapour deposition (CVD) method is used to produce diamond film coatings (DC) on the selected tungsten carbide substrate to increases the wear resistance as well as durability of the carbide tools. It also significantly reduces the coefficient of friction between the coated tool and produced chips. Also, DC prevents corrosion and erosion of tool, reduces the built-up edges as well as the heat produced, while improving the treated surface's quality (Dumpala et al., 2015). This is true, in the field of defense and aerospace applications, where high-precision processing of materials is used. Formation of two or multilayered DCs achieves the increase in the adhesion of the coating to the substrate while improving the cutting properties of the tool.

Previous researchers described the deposition of a nanocrystalline diamond layer from above (Catledge et al., 2000, Sun et al., 2009, Dumpala, et al., 2014), while writing on the properties and deposition of the two-layer micro/nanocrystalline composite diamond coatings, with this coating configuration, high elasticity and low roughness of nanocrystalline diamond layers can be used together with the good adhesion to the selected substrate material and high hardness of microcrystalline diamond layers (Dumpala et al., 2015). To increase the thermal conductivity and prevent the cracking one can use multilayer DCs additionally (Schäfer et al., 2006).

In the present work, we demonstrate a fresh approach to develop the polycrystalline diamond films on the WC tools inserts. The polycrystalline diamond films are of high adhesion and crystallinity. Good adhesion and crystallinity make these materials more suitable for the machine industry.

EXPERIMENTAL

In the present work, tungsten carbide tools were coated with polycrystalline diamond films at atmospheric pressure using thermal chemical vapor deposition method. The substrates were cleaned prior to diamond deposition and then loaded in the thermal CVD system. The experiment was carried out at atmospheric pressure and ~900ºC in the thermal CVD. Sugarcane bagasse was used as a carbon precursor for the growth of diamond on WC substrates with a mixture of Ar/H2to carry the effluent gases. The rich gas phase chemistry of the effluent gases produced by pyrolysis of sugarcane bagasse as mentioned in (Krishnia and Tyagi, 2018, Krishnia et al., 2018) is responsible for diamond growth. The morphology of both bare and after diamond growth WC substrates was studied using scanning electron microscope (Hitachi TM3000). The diamond growth was also confirmed by capturing the Raman spectra using Micro-Raman spectrometer Renishaw via model (514 nm). The micro-hardness of the samples was evaluated using indentation method by using Fischerscope HM2000s (Lal and Singh, 2018, Bansal et al., 2021).

RESULTS AND DISCUSSION

The morphology of the polycrystalline diamond films on WC tools inserts was determined using Scanning Electron Microscopy (SEM). The SEM micrographs were captured at different locations and magnifications as shown in Figure 1 (b & c). The quality of the films on the overall tool surface is almost same and the areas of highly crystalline quality with faceted growth have been shown in inset of Figure 1 (b) inset. It is clearly evident from Figure 1(a) that there absence of any faceted structure on the surface of the bare WC tool insert whereas this faceted growth can be seen in Figure 1 (b & c).

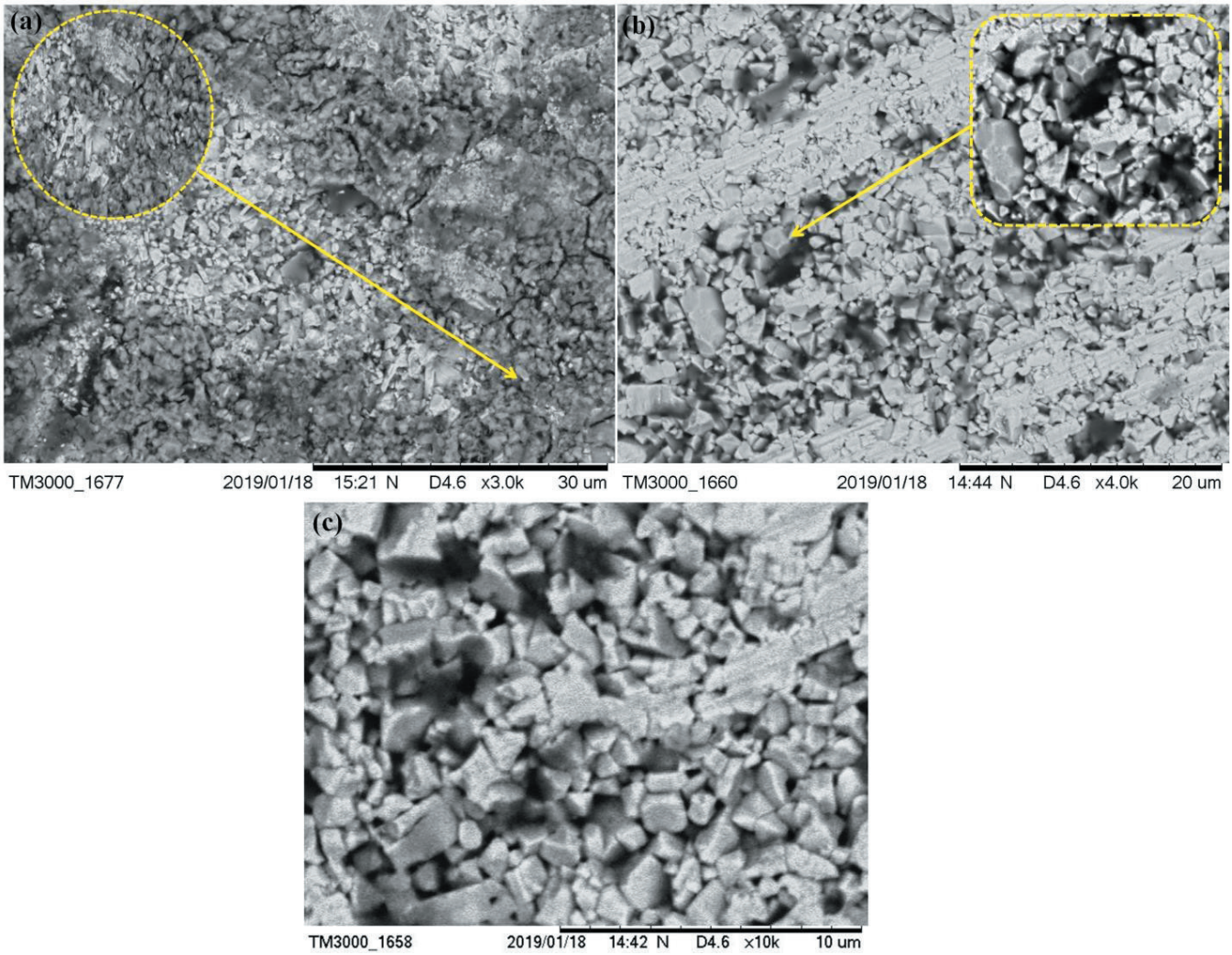


Figure 1. (a) Bare WC tool inert, the inset shows the absence of any faceted structure before diamond deposition, (b & c) diamond coated WC tool inserts where a clear faceted growth on surface can be seen.

Raman spectroscopy was used to investigate the quality of the diamond coatings. Figure 2 shows the Raman spectrum of the diamond coated WC tool insert. The D and G band lies at 1338 and 1576 cm-1 respectively. After deconvolution it has been found that the sp3 content in the films is more and ID/IG is found to be 1.11. The blue shift of the D band from its signature position i.e. 1332 cm-1 of natural diamond may be due to the compressive stress in the coating (Sein et al., 2003,Tang et al., 2001, Krishnia and Tyagi, 2018, Krishnia et al. 2018) which arises as a result of the tool geometry as well as mismatch in thermal coefficients of both diamond and tungsten carbide.

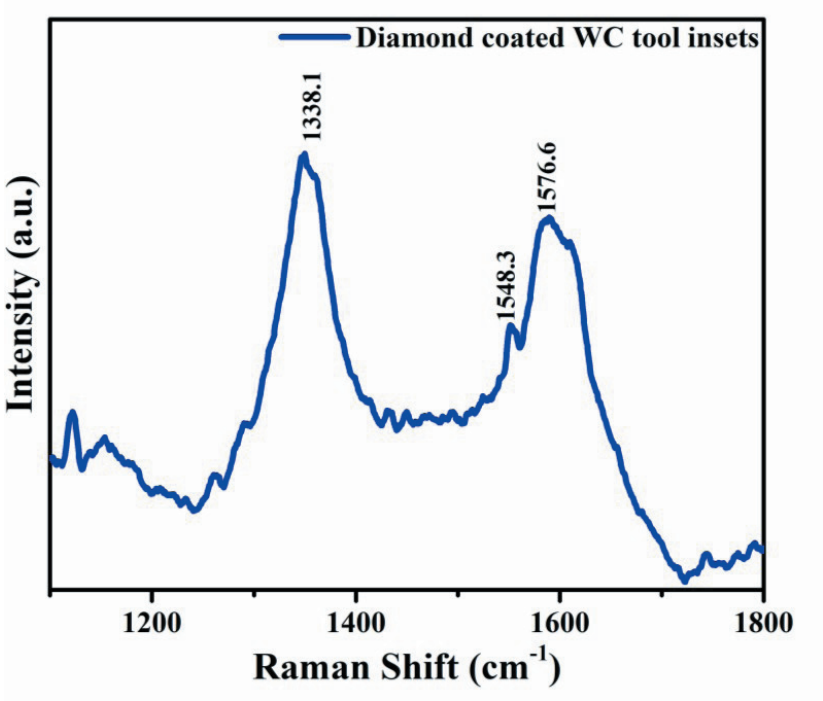


Figure 2 Raman spectra of diamond coated WC tool inserts

The micro-hardness testing has been done using indentation. During indentation the indenter (diamond coated/ uncoated tool) is allowed to indent into the surface and the surface profile after indentation is measured so as to test hardness. The indentation test allows the evaluation of the diamond coating film adhesion on the WC substrate surface (Polini et al., 2000). The Vickers Hardness evaluated after the indentation is shown in Figure 3: and the values are found to be 1103.1 and 1423.32 HV for both un-coated and diamond coated tool respectively. The loading and unloading profile diagrams obtained during the indentation testing with the indentation mechanism is shown in Figure 4 (a-c).

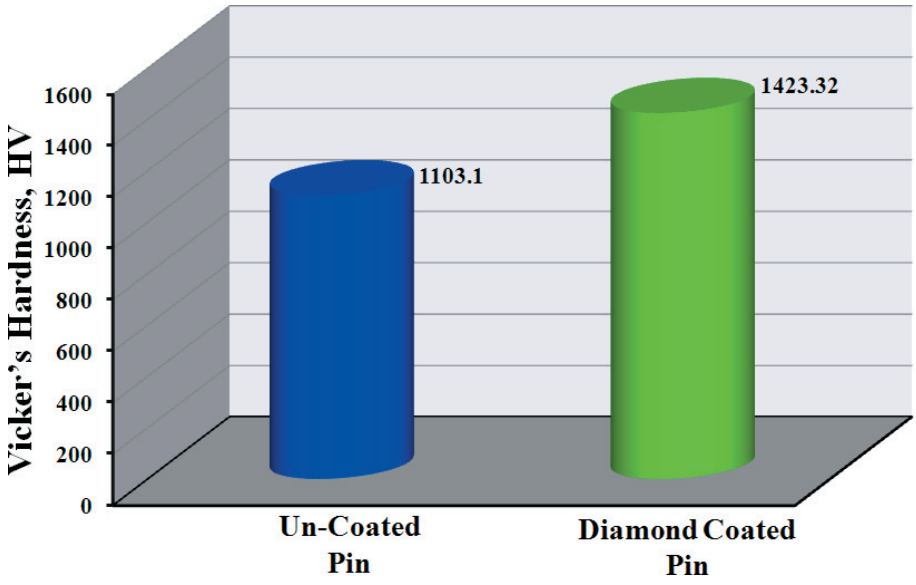


Figure 3 Vickers hardness values of uncoated and diamond coated WC tool inserts.

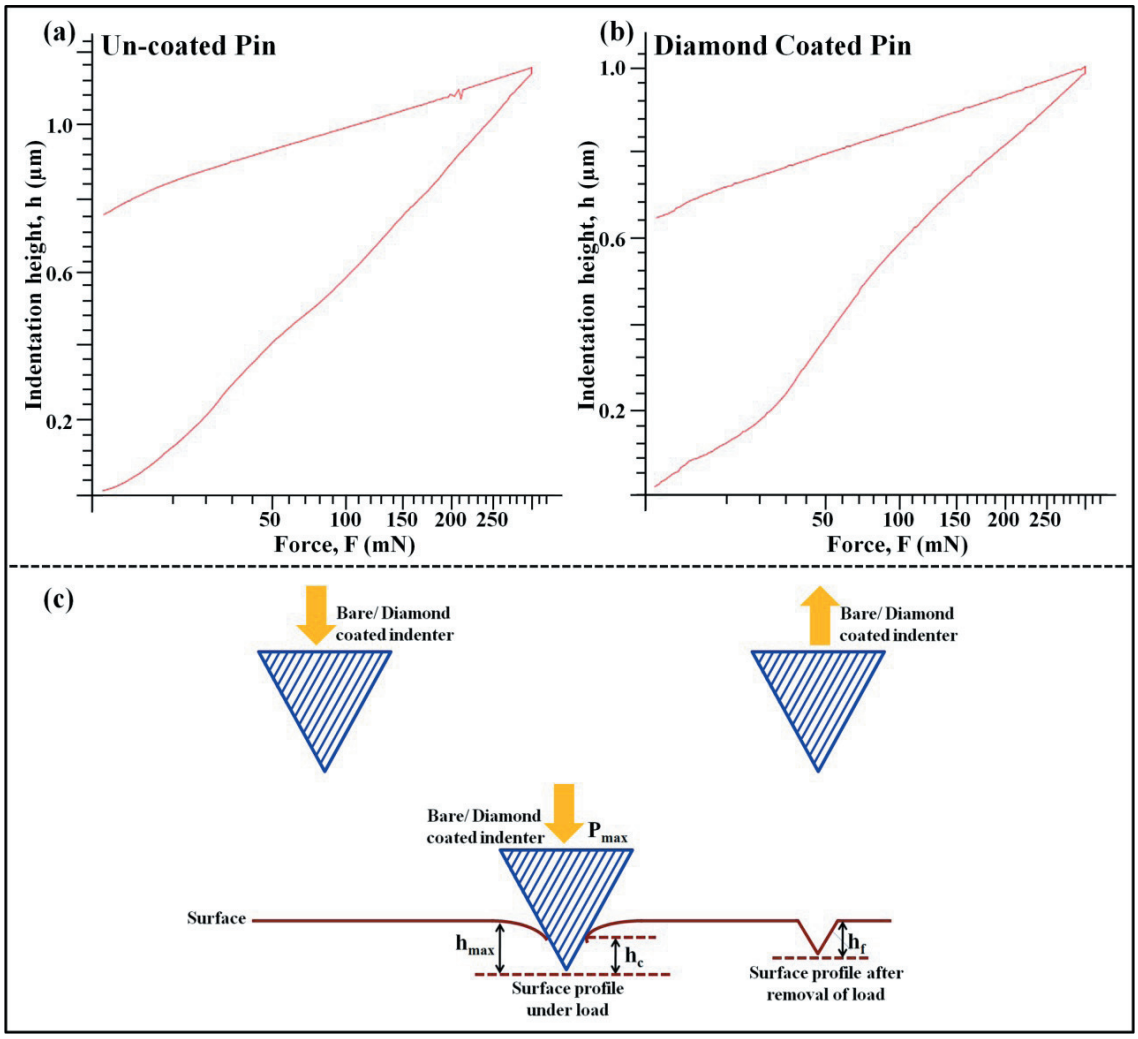


Figure 4 Loading and unloading profile diagrams for (a) uncoated and (b) diamond coated pins with (c) indentation mechanism for micro-hardness testing.

CONCLUSION

In this study, we have coated the WC tool inserts with diamond using thermal-CVD system. The coating has been achieved by using sugarcane bagasse. The surface morphological studies reveal growth of very good quality and crystallinity diamond films. The thicker and good quality of diamond coating over the tool surface is supposed to give the tool inserts better tool life as well as better machining quality. The Vickers hardness of the tool inserts is found to be 1423.32 HV for diamond coated tool inserts which is 29% better than the un-coated tools.

ACKNOWLEDGEMENT

Authors would like to acknowledge the support of Nano Research Facility of Indian Institute Technology, Delhi, India and Institute Instrumentation Centre of Indian Institute Technology, Roorkee, India for allowing us to use their facilities for material characterization. Authors also extend their regards to thank Centre for Advanced Production and Industrial Engineering Research (CAPIER) of Delhi Technological University, New Delhi, India, for providing the facilities for the completion of this work.

REFERENCES

Bansal S., Saraf A., Rana R. and Lal R., 2021. Effect of Picosecond Laser Texture Surface on Tribological Properties on High-Chromium Steel Under Non-lubricated Conditions. Advances in Manufacturing and Industrial Engineering, pp. 257-267. Springer, Singapore.

Catledge, S.A., Baker, P., Tarvin, J.T. and Vohra, Y.K., 2000. Multilayer nanocrystalline / microcrystalline diamond films studied by laser reflectance interferometry. Diamond and related materials, 9(8), pp.1512-1517.

Chen, X. and Narayan, J., 1993. Effect of the chemical nature of transition□metal substrates on chemical□vapor deposition of diamond. Journal of applied physics, 74(6), pp.4168-4173.

Dai, M., Zhou, K., Yuan, Z., Ding, Q. and Fu, Z., 2000. The cutting performance of diamond and DLC-coated cutting tools. Diamond and Related Materials, 9(9-10), pp.1753-1757.

Dumpala, R., Chandran, M. and Rao, M.R., 2015. Engineered CVD diamond coatings for machining and tribological applications. Jom, 67(7), pp.1565-1577.

Dumpala, R., Kumar, N., Kumaran, C.R., Dash, S., Ramamoorthy, B. and Rao, M.R., 2014. Adhesion characteristics of nano-and micro-crystalline diamond coatings: Raman stress mapping of the scratch tracks. Diamond and related materials, 44, pp.71-77.

Endler, I., Bartsch, K., Leonhardt, A., Scheibe, H.J., Ziegele, H., Fuchs, I. and Raatz, C., 1999. Preparation and wear behaviour of woodworking tools coated with superhard layers. Diamond and Related Materials, 8(2-5), pp.834-839.

Gäbler, J., Schäfer, L. and Westermann, H., 2000. Chemical vapour deposition diamond coated microtools for grinding, milling and drilling. Diamond and related materials, 9(3-6), pp.921-924.

Gunnars, J. and Alahelisten, A., 1996. Thermal stresses in diamond coatings and their influence on coating wear and failure. Surface and Coatings Technology, 80(3), pp.303-312.

Hintermann, H.E., 1996. Advances and development in CVD technology. Materials Science and Engineering: A, 209(1-2), pp.366-371.

Jeyapandiarajan, P. and Xavior, A., 2019. Influence of cutting condition on machinability aspects of Inconel 718. Journal of Engineering Research, 7(2): 315-332.

Köpf, A., Lux, B. and Haubner, R., 2001. Substrate Effects during Nucleation. New Diamond and Frontier Carbon Technology, 11(1).

Krishnia, L. and Tyagi, P.K., 2018. Growth and characterization of polycrystalline diamond films on silicon using sugarcane bagasse as carbon precursor at atmospheric pressure by thermal chemical vapor deposition. Diamond and Related Materials, 87, pp.18-26.

Krishnia, L., Yadav, B.S., Palnitkar, U., Satyam, P.V., Gupta, B.K., Koratkar, N.A. and Tyagi, P.K., 2018. As-pyrolyzed sugarcane bagasse possessing exotic field emission properties. Applied Surface Science, 443, pp.184-190.

Kuo, C.T., Lin, C.R. and Lien, H.M., 1996. Origins of the residual stress in CVD diamond films. Thin Solid Films, 290, pp.254-259.

Lal, R. and Singh, R.C., 2018. Experimental comparative study of chrome steel pin with and without chrome plated cast iron disc in situ fully flooded interface lubrication. Surface Topography: Metrology and Properties, 6(3): 035001.

Mallika, K. and Komanduri, R., 1999. Diamond coatings on cemented tungsten carbide tools by low-pressure microwave CVD. Wear, 224(2), pp.245-266.

Nesladek, M., Vandierendonck, K., Quaeyhaegens, C., Kerkhofs, M. and Stals, L.M., 1995. Adhesion of diamond coatings on cemented carbides. Thin Solid Films, 270(1-2), pp.184-188.

Polini, R., D'antonio, P., Casto, S.L., Ruisi, V.F. and Traversa, E., 2000. Cutting performance and indentation behaviour of diamond films on Co-cemented tungsten carbide. Surface and Coatings Technology, 123(1), pp.78-83.

Rana, R., Murtaza, Q. and Walia, R.S., 2021. Optimization using genetic algorithm of tribological behaviour of WC tool material. Indian Journal of Engineering and Materials Sciences (IJEMS), 27(4): 889-896.

Schäfer, L., Höfer, M. and Kröger, R., 2006. The versatility of hot-filament activated chemical vapor deposition. Thin Solid Films, 515(3), pp.1017-1024.

Sein, H., Ahmed, W., Jackson, M., Ali, N. and Gracio, J., 2003. Stress distribution in diamond films grown on cemented WC–Co dental burs using modified hot-filament CVD. Surface and Coatings Technology, 163, pp.196-202.

Sun, F., Ma, Y., Shen, B., Zhang, Z. and Chen, M., 2009. Fabrication and application of nano–microcrystalline composite diamond films on the interior hole surfaces of Co cemented tungsten carbide substrates. Diamond and Related Materials, 18(2-3), pp.276-282.

Tang, W., Wang, Q., Wang, S. and Lu, F., 2001. Adherent diamond coatings on cemented carbide substrates with different cobalt contents. Diamond and related materials, 10(9-10), pp.1700-1704.

Ullen, N.B., Hasak, S.M.A. and Dirikolu, M.H., 2020. Factors influencing the machinability during turning sinter-hardened Cu-Ni-Mo based steel: Dependency on cutting speed, feed rate and cutting depth. Journal of Engineering Research, 8(4): 236-257.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2021.3123174, IEEE Journal of Emerging and Selected Topics in Power Electronics

Digitally Controlled Power Management Circuit with Dual-functioned Single-stage Power Converter for Vibration Energy Harvesting

Hyunjun Jung, Yamini Sharma, and Lei Zuo

***Abstract*— This paper proposes digitally controlled power management circuit (PMC) for vibration energy harvesting in vehicle suspension system. The PMC is tested with a mechanical motion rectifier based electromagnetic energy harvester that converts from bi-directional vibration to unidirectional rotation, with improved energy harvesting performance and efficiency. Topology of one-stage DC-DC converter is adapted, and it is digitally controlled while considering the uncertainty of vibration sources. In order to efficiently transfer the harvested power to the electrical load, the algorithm of digital controller is designed to have four operation modes, which are impedance matching mode, voltage regulation mode, generator protection mode, and low power mode, for increasing the PMC efficiency while reducing the power consumption of controller. Experimental results show that proposed circuit can successfully operate and switch between the four operation modes. In the evaluation of PMC performance with 2mm and 2Hz excitation, the efficiency of proposed one-stage power converter is 86%. The total power consumption of controller in active mode and low power mode are 109 mW and 2.52 mW respectively while harvesting about 3W vibration energy.**

***Index Terms*— Energy harvesting, power management circuit, vibration, vehicle suspension, impedance matching, battery charging**

I. INTRODUCTION

SMART technologies such as wireless sensing and control system in the intelligent vehicles have been spotlighted for safety and comfort driving, monitoring, control, and logistics management [1-4]. However, reliable and economic power supply remains a challenging issue to overcome to adapt the smart technologies in real applications [5-7]. Motor vehicles such as cars, trucks, and trains require multiple power sources to power sensors and control systems, which result in the complexity of vehicle design [8-9]. The active or semi-active suspensions have been researched in the past several decades, demonstrating improved vehicle dynamic performance [10-13]. However, they require additional power supply and consume more energy [14]. In case of railed vehicle such as freight train, the power source is located at locomotive, and it is difficult to extend the power cable to the individual railcars because of the up to 200 trailer lengths [15]. Battery or diesel engines can be used for each trailer but leads to high logistics cost for replacement or refueling.

Energy harvesting technologies can be a promising solution to supply power for sensing (and control) in intelligent road and rail vehicles [16-17]. Energy harvesting systems can harvest the wasted energy in the vehicles, which can supply sustainable power as well as improving the fuel economy [18-20]. The electromagnetic vibration energy harvesting from suspensions has been actively developed, because the amount of harvestable energy is large, and the size of the electromagnetic transducer is relatively small [21-22]. In our previous work [23], electromagnetic energy harvesting suspension of automotive was designed and verified with a field test, and 19 W was obtained at 48 km/h (30mph) of vehicle speed. In [15], we tested a mechanical motion rectifier (MMR) based electromagnetic energy harvester in freight railcar and concluded 9.2-14.5 W can be obtained at 90 km/h of train speed.

Power management circuit (PMC) is an important interface that transfers the power from electromagnetic vibration energy harvester to the electrical load [24-26]. The PMC shall have multiple functions: regulating the voltages produced from irregular vibration, maximizing the power output or energy conversion efficiency, protecting the generator from overloading, and properly charging battery. Some batteries such as Nickel-Cadmium (NiCd), Lithium Ion (Li-ion) can have increased life span with pulse charging. Charging the battery with railcar energy harvesting system in real application could provide natural pulse charging [27], which allows long battery lifetime.

Guo et al. [28] proposed buck-boost converter with an analog controller for railcar energy harvesting. In this case, the power converter can transfer the maximum power with impedance matching, but this circuit doesn't consider the battery protection that is critical for power supply functionality. Wang et al. [29] proposed a two-stage impedance matching and voltage regulation circuit. Power management circuit therein can detect the power generated by electromagnetic energy harvesters, and then determine the operation mode between active and sleep to minimize the power consumption of circuit. Voltage regulator for battery charging was adopted in [29], but the two-stage converter for impedance matching and voltage regulation reduced the total circuit efficiency. Connor et al. [27] proposed

Hyunjun Jung, Yamini Sharma, and Lei Zuo are with Center for Energy Harvesting Materials and Systems, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: jung87@vt.edu; yaminisharma244@vt.edu; *leizuo@vt.edu. *Corresponding).

2168-6777 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:58:16 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2021.3123174, IEEE Journal of Emerging and Selected Topics in Power Electronics

one stage DC-DC power converter with an analog controller for highly efficient power conversion. The power converter is able to switch the operation mode between impedance matching mode for maximum power transfer and voltage regulation mode for battery protection. However, power consumption of controller was over 200 mW, which can be significant when vibration is small, and there is no consideration for sleep mode to save power consumption of circuit. In addition, the power management circuit must also consider the generator protection to avoid exceeding generator current limit.

In this paper, a digitally controlled power management circuit with a multi-functional single-stage power converter is proposed and tested with freight railcar energy harvester for reliable and sustainable power supply. Topology of one stage DC-DC converter is adapted, and it is digitally controlled to match the impedance for maximum power transfer. Algorithm of a digital controller is also designed to reduce the power consumption of the controller with consideration of intermittent and uncertain vibration source. In addition, the digital controller also protects the battery from overcharging and the generator from overloading.

The rest of the paper is organized as follows: Section 2 presents the MMR electromagnetic energy harvester as well as the analysis of ambient source in railcar suspension for the design of proposed power management circuit. Section 3 proposes a digitally controlled power management circuit with multi-functional single-stage power converter. Section 4 presents the experimental results to demonstrate the efficiency and effectiveness of the circuit. Section 5 concludes the paper.

II. MMR ELECTROMAGNETIC ENERGY HARVESTING FROM FREIGHT RAILCAR SUSPENSION

The MMR electromagnetic energy harvester is designed to produce energy by translating the linear motion of the suspension vibration into rotational motion to drive a permanent magnetic generator. Fig.1 shows the installation of the MMR electromagnetic energy harvester on a freight K5 bogie, which is used for freight railcar made by CRRC Yangteze Ltd. Co. [15]. The K5 bogie mainly consists of two wheelsets, side frames, bolster, spring plank, center plate, roller bearing adapters, and coil springs. The MMR electromagnetic energy harvester was installed between the spring plank and the bolster in parallel with the coil springs to convert the energy from suspension displacement to the 3-phase AC generator.

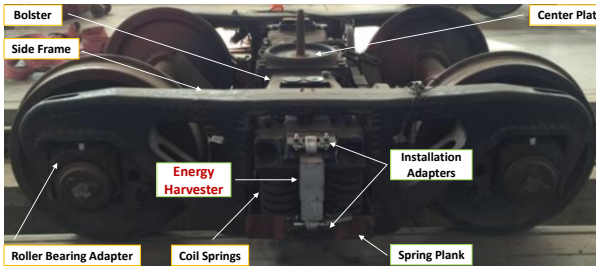


Fig. 1. The installation of the MMR electromagnetic energy harvester on a freight K5 bogie [Reproduced and drawn with permission from Ref. [15]].

Fig. 2 shows the MMR based electromagnetic energy harvester. It consists of rack and pinion gear, Mechanical Motion Rectifier (MMR) gearbox, planetary gearhead, and electrical generator. The bi-directional relative vibration between the wheel axle and the bogie of the freight railcar is converted into a bi-directional rotation of the shaft by the rack and pinion mechanism. The MMR is embedded with two one-way clutches that only allow one directional motion transmission. The MMR converts the bi-directional rotation of the pinion shaft to unidirectional rotation of the geared generator, which could improve the energy harvesting performance and efficiency. To transfer the maximum power from the vibration system, impedance matching between energy harvester and the electrical load is required. However, load condition can be changed based on the power demand, hence the energy harvester cannot maintain the maximum power transfer condition using the electrical load directly. To solve this problem, a DC-DC converter can be used for

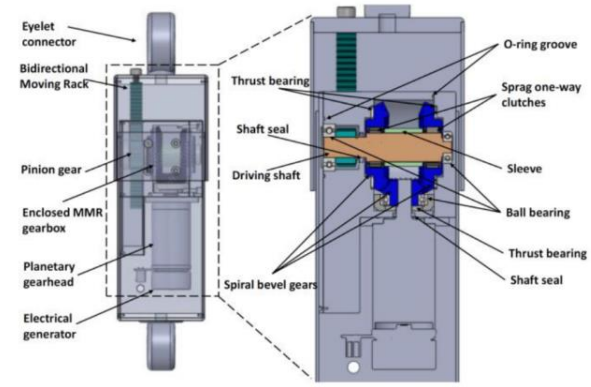


Fig. 2. MMR based electromagnetic energy harvester [Reproduced and drawn with permission from Ref. [31]].

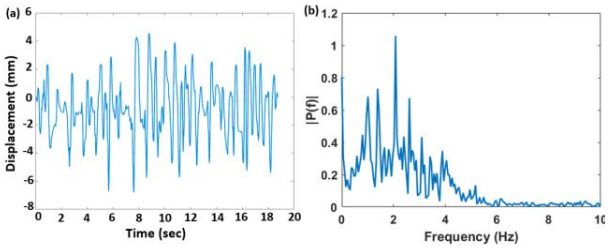


Fig. 3. Experimental (a) displacement of suspension and (b) its FFT at freight railcar [Reproduced and drawn with permission from Ref. [31]].

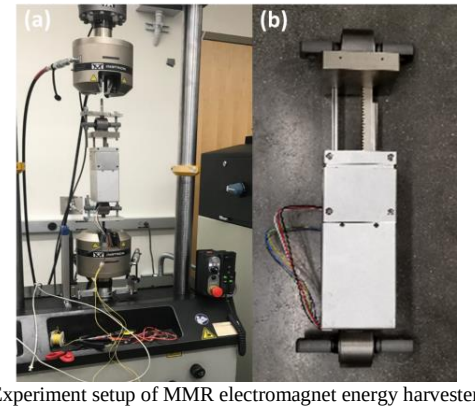


Fig. 4. Experiment setup of MMR electromagnet energy harvester on the Instron machine.

2168-6777 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:58:16 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2021.3123174, IEEE Journal of Emerging and Selected Topics in Power Electronics

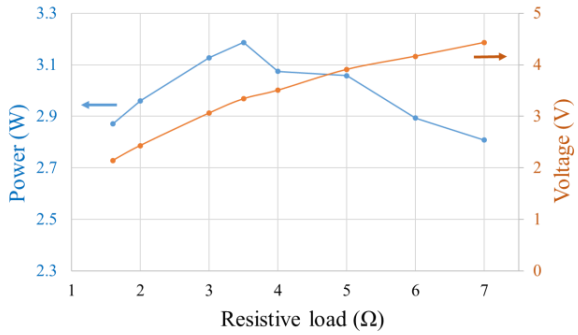


Fig. 5. RMS power and voltage with various resistive loads.

emulating impedance value to match with energy harvester for extracting the maximum power from the harvester. Before designing the DC-DC converter, impedance matching value of MMR electromagnetic energy harvester is investigated. In [30], the experimental suspension displacement of the freight railcar was recorded, as shown in Fig. 3(a). Based on the displacement, Fast Fourier Transform (FFT) is obtained as shown in Fig. 3(b). The highest amplitude occurred at 2 Hz. To obtain impedance matching value of the MMR energy harvester under highest amplitude condition, an Instron machine is used for mocking up the mechanical excitation using 2 Hz and 2 mm sinusoidal input to the MMR electromagnetic energy harvester, as shown in Fig. 4. The optimal resistance was found using a various resistive loads connected in parallel to the MMR electromagnetic energy harvester and the full bridge rectifier. As shown in Fig. 5, as resistive load increases, voltage at resistive load increases. In the given example, maximum power is transferred when the external load resistance is 3.5Ω. For more general cases, online perturbation and observation (P&O) or other algorithms [32, 33] can be used to find the best resistance.

III. Design of Digitally Controlled Power Management Circuit with Multi-Functional Single-Stage Power Converter

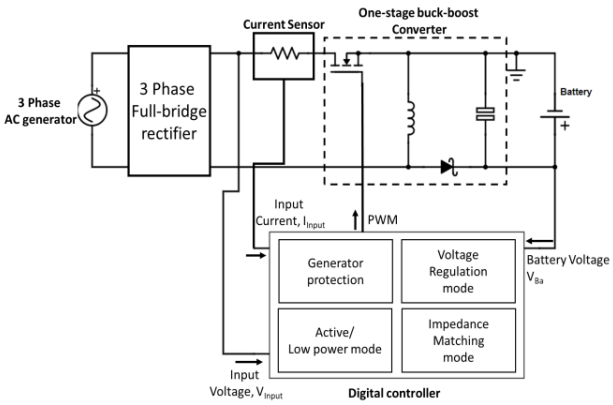


Fig. 6. Proposed power management circuit archtecture.

A. Circuit architecture

Fig. 6 show the proposed power management circuit (PMC) architecture. Proposed electromagnetic energy harvester can work in four different modes: the matched impedance mode, battery voltage regulation mode, low power mode, and generator protection mode. In the impedance matching mode, the one-stage DC-DC converter emulates this ideal resistance value to extract maximum power from the harvester. In the case of voltage regulation mode, the single-stage DC-DC converter protects the battery from over-charging by regulating its output voltage. To switch the mode between impedance matching and voltage regulation, the battery voltage is sensed. When the battery voltage is lower than 80% of maximum battery voltage, i.e. the battery is not fully charged yet, the one-stage DC-DC converter operates in impedance matching mode to extract maximum power. When the battery voltage is higher than 80% of maximum battery voltage, i.e. the battery is almost fully charged, the converter operates in voltage regulation mode to protect battery from over-charging. When converter input current is higher than the current limit of the generator in the energy harvester, the controller is set to generator protection mode to make a large equivalent resistance value of input DC-DC converter to limit the generator current to a safe level. When the energy harvester does not generate sufficient power, the controller is set to low power mode to shut down the entire circuit to reduce the power loss of circuit.

B. Power stage design

1) Three phase full-bridge rectifier

The output voltage of the proposed electromagnetic energy harvester is much larger than typical vibration energy harvesting systems. Therefore, low forward drop Schottky diodes are used for rectification. The Schottky diodes (FUS45-0045B from IXYS Corporation) have forward drop of 550mV at 15A.

2) One-stage DC-DC buck-boost converter

The voltage from proposed energy harvester can be varied, which can be higher or lower than battery voltage in real applications. Thus, a buck-boost type topology is utilized for proposed energy harvester. The equivalent input resistances of the buck-boost converter in continuous and discontinuous modes are shown in Table I [27]. From the table, the equivalent input resistance of the buck-boost converter in continuous conduction mode (CCM) is dependent on the load resistance R and duty cycle D. This makes it difficult to maintain a fixed resistance value since the load resistance value can change under different operating conditions. In discontinuous conduction mode (DCM), the equivalent resistance value depends only on the duty cycle D, if the converter is operated with fixed PWM period T_s. This way, the duty cycle of the

TABLE I
THE EQUIVALENT INPUT RESISTANCE OF BUCK-BOOST CONVERTER IN DIFFERENT CONDUCTION MODES

Conduction mode	Active mode (mW)
Continuous conduction mode	$R_{in,CCM} = \frac{V_{in}}{I_{in}} = \frac{R(1-D)^2}{D^2}$
Discontinuous conduction mode	$R_{in,DCM} = \frac{V_{in}}{I_{in}} = \frac{2L}{D_1^2 T_s}$

2168-6777 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:58:16 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2021.3123174, IEEE Journal of Emerging and Selected Topics in Power Electronics

buck-boost converter can be set to achieve a desired equivalent input resistance.

C. Control design

The TI MSP432 digital controller is used to decide the operating mode of the proposed one-stage DC-DC converter: impedance matching mode, battery voltage regulation, low power mode, and generator protection mode; as well as control the DC-DC converter to realize the required operation. TI MSP432 has an in-built 14-bit ADC which is used to sense the input voltage V_{input}, the input current I_{input}, and output voltage or the battery voltage V_{Ba} of the buck-boost converter for controlling the operation modes. Timer A in MSP432 is used for generating the PWM. Since the harvester's peak voltage can

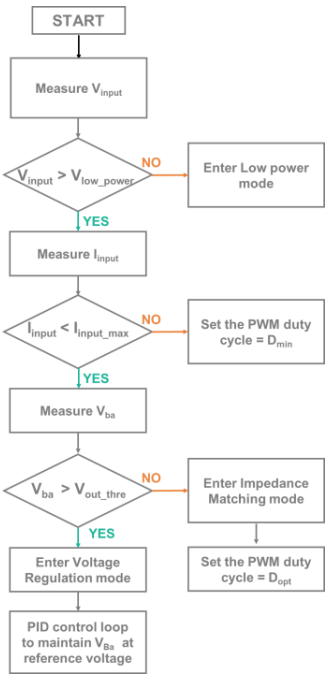


Fig. 7. The algorithm of proposed MSP432 digital controller.

be very large, a voltage divider of gain 0.1 is connected with input of one-stage DC-DC converter for protecting the ADC in MSP432. We take the sensing voltage divider as V_{input}. Fig. 7 shows the algorithm of proposed MSP432 controller. Initially,

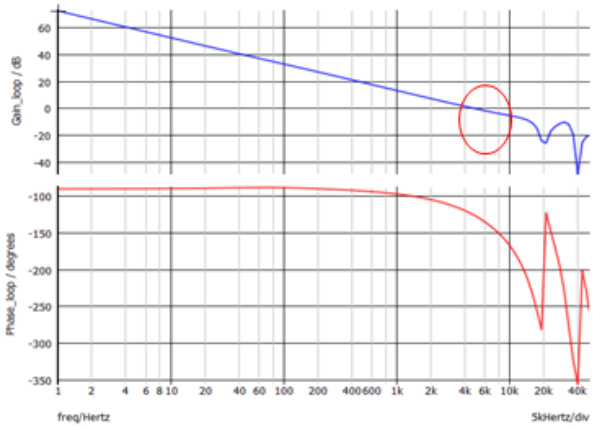


Fig. 8. The control-to-output transfer function (a) gain vs (b) phase vs frequency.

ADC reads the input voltage V_{input} to determine the mode between active and low power mode. When V_{input} is higher than V_{low_power} (generator output after the rectifier is over 1V), ADC reads the input current with active mode status. Otherwise, it transitions to low power mode to minimize the power consumption of proposed circuit. The ADC also reads the input current I_{input} from the current sensor for the protection of the energy harvester generator. When I_{input} is higher than the maximum constant current limit of the generator I_{input_max} (1.6A for the Maxon Motor EC-I 40, model# 496652), the controller sets a low PWM duty cycle D_{min} for high input impedance to reduce the current. Otherwise, ADC reads battery voltage V_{ba} to decide between impedance matching and battery voltage regulation mode. When the battery voltage is lower than 80% of maximum battery voltage V_{out_thre}, the controller sets the impedance matching mode with D_{opt} to extract maximum power. When the battery voltage is higher than 80% of maximum battery voltage, controller sets voltage regulation mode to protect the battery from over-charging.

1) Low power mode and generator protection mode

The low power mode is activated when the input voltage is insufficient to charge the battery, or the energy harvester does not generate power. In the low power mode, the PWM is shutoff and the clock frequency of the MCU is reduced to save power. When V_{input} is lower than 0.1V (voltage after the rectifier is less than 1V), it transitions to low power mode to minimize the power consumption of proposed circuit. Otherwise, proposed PMC is activated for charging the battery. After activation, input voltage can be zero in some instances because the input of DC-DC converter is connected with ground when the switch in buck-boost converter turns on. This event unintentionally enables low power mode. To prevent it, MCU monitors V_{input} for 5 ms after V_{input} becomes lower than 0.1V. If V_{input} becomes higher than 0.1V within 5 ms, the MCU maintains the active mode. Otherwise, it transitions to low power mode.

Generator protection mode is activated when I_{input} is higher than the maximum current limit of generator. Duty cycle of buck-boost converter is directly related to input current because duty cycle controls the input resistance of buck-boost converter. In order to instantly reduce the current, the MCU sets low PWM duty cycle D_{min}, such as 1%.

2) Impedance matching mode

In the impedance matching mode, the buck-boost converter emulates an ideal resistance value. This ideal value can be determined by feedback control, P&O, or other methods [28, 31, 32]. To demonstrate the working principle of the multi-mode, here we set the ideal resistance value based on experiment as shown in Fig. 5 based on a particular vibration excitation by connecting a resistor to the harvester and calculating the power for different resistance values. For the given example, the energy harvester generates maximum power when the resistance is 3.5Ω. In order to emulate the matched impedance value, the buck-boost converter is controlled by MSP432 controller in the discontinuous conduction mode (DCM)

$$R_{in} = \frac{2L}{D^2 T_s} \qquad (1)$$

R_{in} is emulated input impedance of DC-DC converter, L is inductance in the DC-DC converter, D is duty cycle of PWM, and T_s is switching period of PWM. In the experiment L =33μH,

2168-6777 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:58:16 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2021.3123174, IEEE Journal of Emerging and Selected Topics in Power Electronics

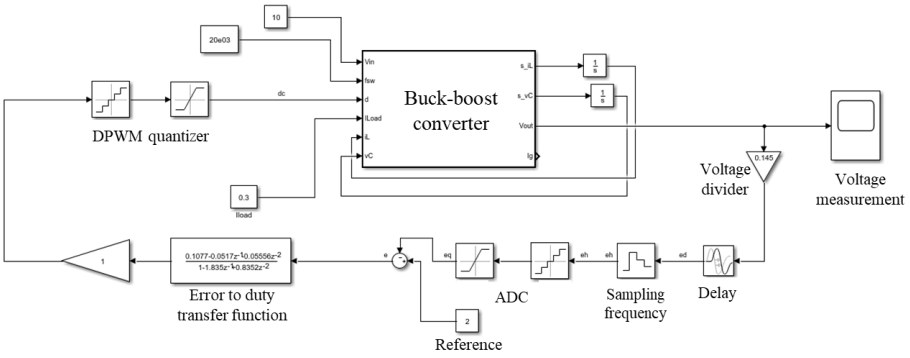


Fig. 9. Simulink block diagram of converter with feedback control system.

T_s= $\frac{1}{20{,}000}$ sec, and D = 61% is determined to set the 3.5Ω of R_{in}. That is, the MSP432 controller is configured to provide a 20kHz PWM with duty cycle set to 61%.

3) Voltage regulation mode

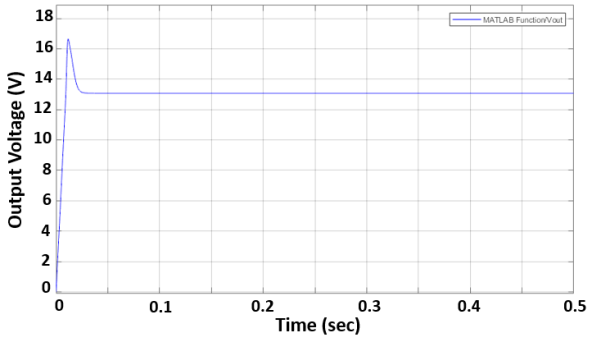


Fig. 10. The regulated output voltage step response as a results of Simulink simulation.

In the voltage regulation, a lead-lag feedback control is designed to maintain a constant output voltage of the one-stage DC-DC buck-boost converter. For the buck-boost converter with DCM mode, a model of the converter given with pole and one zero in [34] is adapted

$$s_{zero} = \frac{1}{R_{ESR}C_{out}} \quad (2)$$

$$s_{pole} = \frac{2}{R_{load}C_{out}} \quad (3)$$

where C_{out} is the output capacitor (=660μF), R_{ESR} is the equivalent series resistance (ESR) of the output resistance (= 45mΩ from the datasheet), and R_{load} is the load resistor (7.2Ω, considering the output current of 1.67A and output voltage of 13V), which are set based on the worst-case condition. Using these values, the zero is located at 5.3kHz and the pole at 67Hz. To reduce the steady error of output voltage and achieve a bandwidth of 5kHz, a 2nd order analog compensator is used with one integrator pole, one pole at 5.3kHz, and one zero at 67Hz. The gain is decided based on the required bandwidth and is set to 512. Control type-2 compensator is used for the control-to-output loop transfer function H(s), which is given by

$$H(s) = \frac{k(1+\frac{s}{\omega_{zero}})}{s(1+\frac{s}{\omega_{pole}})} = \frac{512(1+\frac{s}{2\cdot\pi\cdot 67})}{s(1+\frac{s}{2\cdot\pi\cdot 5300})} = \frac{512(1+0.0024s)}{s(1+0.00003s)} \quad (4)$$

Circuit simulation is carried out in the analog domain for analyzing the performance of feedback control. The control-to-output loop transfer function H(s) is shown in Fig. 8, which shows the gain and phase plots. Simulation results show that a bandwidth ~5kHz is obtained, and the phase margin is more than 50°, which result in the stable condition for regulating the voltage. Conversion from the analog transfer function into digital domain is carried out for digital control implementation by using Tustin bilinear transform given by [34].

$$z = e^{sT_s} \approx \left(\frac{1+s\left(\frac{T_s}{2}\right)}{1-s\left(\frac{T_s}{2}\right)}\right) \leftrightarrow s = \left(\frac{2}{T_s}\right)\left(\frac{z-1}{z+1}\right) \quad (5)$$

Substituting the value of s in Equation (5) to Equation (4), the compensator transfer function H(z) from voltage error to duty cycle in digital domain comes out to be

$$H(z) = \frac{duty}{error} = \frac{0.1077-0.0517z^{-1}-0.05556z^{-2}}{1-1.835z^{-1}+0.8352z^{-2}} \quad (6)$$

in which the T_s in Tustin bilinear transform is taken as one order smaller than the PMW switching period, $1/T_s$ =187kHz. To verify the above transfer function, a buck-boost converter and feedback control system in the digital domain is simulated in Simulink. The simulink block diagram of converter with feedback control system is shown in Fig. 9. DCM capacitor voltage in the buck-boost converter block is used for feedback control. As shown in Figure 9, a gain of 0.145 is used to represent the divider on the output voltage in analog domain. The delay block is used to model the controller operations delay (such as ADC time delay, calculation delay etc.), which is set to 4μs. The ADC block represents the ADC in the MSP432 controller, and the sampling frequency is set to 187 kHz. The quantization level is set to 0.0002 V (= $3.3V/2^{14}$) due to the 14-bit ADC in MSP432 controller. The reference voltage is compared with the output voltage to calculate error The aim of the digital controller H(z) is to minimize this error by changing the duty cycle of the converter. The digital duty cycle signal is converted into analog duty cycle using a DPWM quantizer, and then the analog duty cycle forces the buck-boost converter to produce a constant output voltage V_{Ba} on the capacitor under time varying input voltage or output load within the target specifications.

2168-6777 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:58:16 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2021.3123174, IEEE Journal of Emerging and Selected Topics in Power Electronics

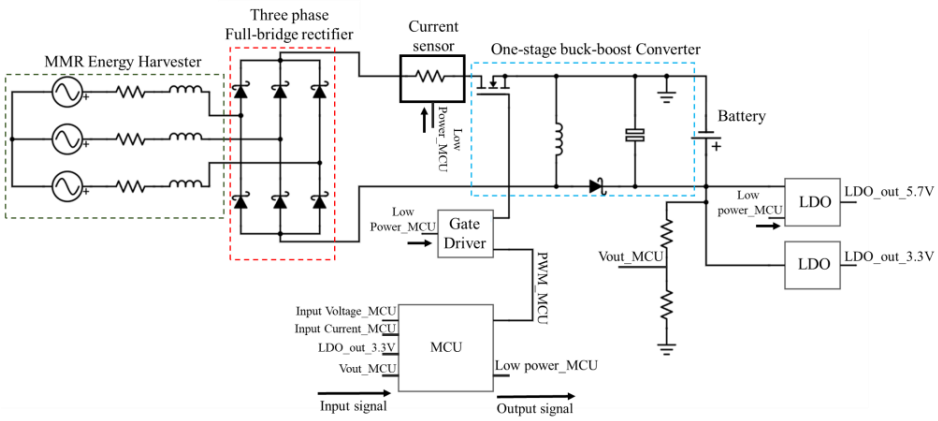


Fig. 11. The schematic of proposed PMC.

Simulink simulation is carried out with input voltage of 20V, switching frequency of 20kHz, and output load current of 1.67A, which are the same conditions as simulation in analog domain. Fig. 10 shows regulated output voltage as a result of Simulink simulation. At the initial point around 0.03 second, there is an overshot of 29% on the output voltage. After that, regulated constant voltage of 13 V is produced. The final step is to convert the digital domain transfer function into C language code for MSP432. The equation comes out to be

$$d[n] = 0.1077e[n] - 0.0517e[n-1] - 0.05556e[n-2] + 1.835d[n-1] - 0.8352d[n-2] \quad (7)$$

where d[n] is the current duty cycle, d[n-1] is the previous duty cycle and so on. e[n] is the error between the measured output voltage and a reference voltage. This digital domain transfer function controls the duty cycle of PWM to maintain an output voltage around the reference voltage.

D. Circuit implementation

Fig. 11 shows the schematic of proposed PMC. It consists of a three-phase bridge rectifier, a single-stage buck-boost DC-DC converter, a gate driver, two linear drop-out (LDO) voltage regulators and an input current sensor. The buck-boost converter consists of an inductor, N-channel MOSFET switch, Schottky diode, and input and output capacitors. The negative terminal of the battery is connected to the ground. The PWM signal (PWM_MCU) from MSP432 microcontroller unit (MCU) is 3.3V as shown in Fig 11, which is not enough to efficiently drive the MOSFET switch. Thus, the PWM is sent to the gate driver that is used to amplify the output signal of the MCU. One LDO takes input from the 12V battery and is used to produce an output of 5.7V (LDO_out_5.7V) for powering the gate driver and the current sensor, as shown in Fig. 11. Another LDO takes input from the battery to produce an output of 3.3V (LDO_out_3.3V) to power the MCU. Since the maximum input voltage of ADC in the MCU controller is 3.3V, a resistive divider network is connected to the output voltage node (Vout_MCU) and then fed to the controller ADC for protecting the MCU. The current sensor consists of an amplifier, shunt resistor, and passive low pass filter. The shunt resistor and amplifier are used to sense the input current, and the passive low pass filter is used at the amplifier output to measure the constant (average) input current for generator protection. Output signal of the current sensor (input current_MCU) and input voltage (input voltage_MCU) are sent to ADC inputs of the MCU to control the generator protection and low power modes. In addition, the MCU sends the output signal (low power_MCU) to gate driver and current sensor for controlling the activation or deactivation to save power.

IV. EXPERIMENT AND RESULTS



Fig. 12. (a) the proposed circuit realized in a printed circuit board (PCB), (b) test setup.

The proposed circuit realized in a printed circuit board (PCB) is shown in Fig. 12 (a). The PCB is connected with the MCU and a 12 V lead acid battery (Power Sonic, 12180) as shown in Fig. 12 (b). In order to verify the performance of the proposed circuit, the harvester is attached to Instron machine to simulate the railway bogie movement, as shown in Fig. 4. The output of the three-phase generator is connected to the power management circuit. Considering the typical excited frequency of railway suspension (2 Hz), the system is excited with input between 1.5Hz and 2.5Hz.

A. Impedance matching and Voltage regulation mode

Fig. 13 shows input resistance measurements of the buck-boost converter at different excitation frequencies, for showing the input impedance stability. In the test, the equivalent input resistance of the buck-boost converter is maintained at a fixed value of 3.5Ω at different frequencies. In addition, with 2 Hz and 2 mm sinusoidal input, the measured input power of one-stage DC-DC power converter is 3.05 W, which is very close to output power (3.18 W) at optimum resistive load (3.5 Ω), as shown in Fig. 5. Test results show that proposed circuit in impedance matching mode can successfully match the impedance for maximum power transfer in different input conditions. For the voltage regulation test, a DC power supply is connected to the proposed circuit as the source voltage. This

2168-6777 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:58:16 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2021.3123174, IEEE Journal of Emerging and Selected Topics in Power Electronics

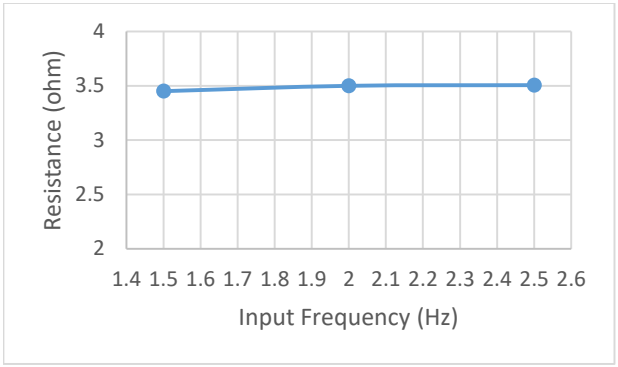


Fig. 13. Measurement results of input resistance of the buck-boost converter at different excitation frequencies.

test is carried out to verify operation of the circuit over the entire source voltage range. Fig. 14 (a) shows the output of the buck-boost converter at different source voltages. In the test, the proposed buck-boost converter maintains a targeted output voltage of 13V at different source voltages. Fig. 14 (b) shows the input current vs. source voltage for the cases with and without voltage regulation. As source voltage increases in the case without voltage regulation, input current increases due to the absence of the feedback loop. However, as source voltage increases in the case with voltage regulation, input current is limited to maintain the regulated output voltage for battery protection.

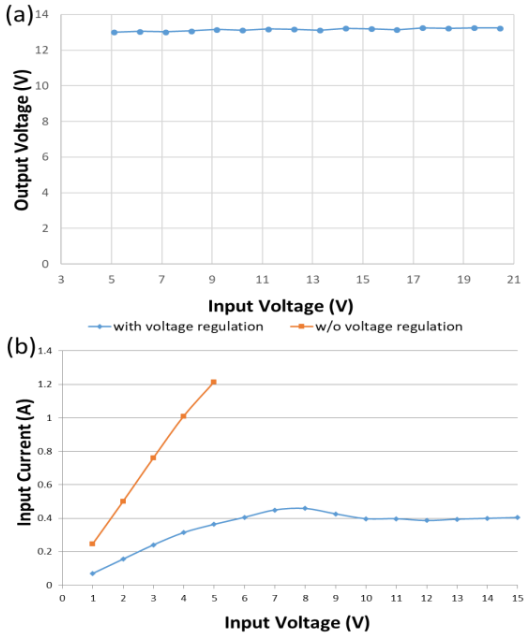


Fig. 14. (a) the output of the buck-boost converter at different input voltages (b) the input current vs source. voltage for two different cases: 1) with voltage regulation, 2) without voltage regulation.

B. Low power mode and generator protection mode

Fig. 15 shows the testing results of transient waveforms between active mode and low power mode. The top graph in purple is source voltage (after 1/10 voltage divider), the middle one in yellow is PWM generated by the MCU, and the bottom one in blue is the signal of the low-power MCU. The low power mode is activated when source voltage is less than 0.1V (after 1/10 voltage divider), as shown in Fig 15 (a). In this mode, the MCU shuts the PWM down, and the clock frequency is reduced. Furthermore, the low-power signal turns the gate driver off to save power. In Fig. 15 (b) shows the transient waveforms from low power mode to active mode. As the source voltage goes above 0.1V, the PWM and gate driver are turned on. The active mode operation resumes. To prevent false entry into low power mode, the controller reads source voltage about 5 ms after source voltage is less than 0.1V. If the voltage is higher than 0.1 V within 5 ms, active mode is maintained, otherwise, the MCU transitions to low power mode as shown in Fig 15 (a). Fig. 16 shows the I_{input}, V_{input} and PWM signal in the active mode (impedance matching mode), and generator protection mode. The top graph in purple is the output signal of current sensor I_{input}, the middle one in yellow is PWM generated by the MCU, and the bottom one in blue is the source voltage V_{input}. If source voltage is higher than 0.1V, the MCU enters the active mode to generate PWM as shown in Fig. 16(a). In this mode, as I_{input} is lower than the maximum current limit of the generator (1.6 A), MCU generates PWM with 61% of duty cycle. Once I_{input} is higher than the maximum 1.6 A, controller set 1% of PWM duty cycle, which results in the reduction of current as shown in Fig. 16(b). Testing results show that proposed PMC can successfully operate in low power mode to minimize the power consumption of the circuit and in generator protection mode for protecting the energy harvester.

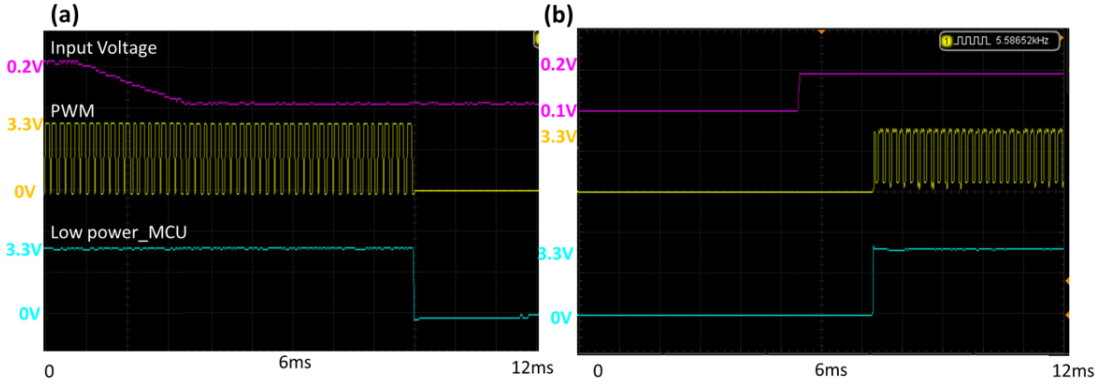


Fig. 15. Operation of low power mode: (a) from active to low power mode, and (b) from low power mode to active mode.

2168-6777 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:58:16 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2021.3123174, IEEE Journal of Emerging and Selected Topics in Power Electronics

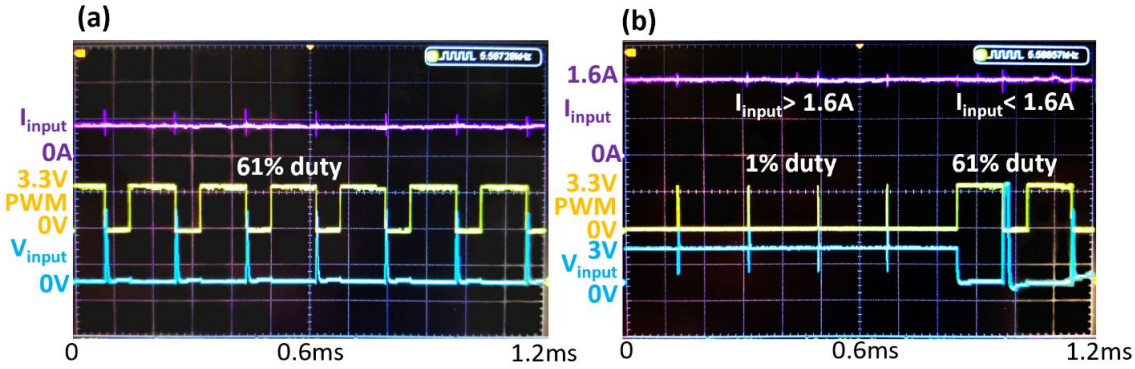


Fig. 16. Test results of generator protection mode.

Additionally, the algorithm of generator protection mode is intended to protect the generator, but the algorithm can be also used for limiting the maximum battery charging current to protect the battery in the impedance matching mode.

C. Circuit efficiency and power consumption

TABLE II
CONTROLLER POWER LOSS

Controller power loss	Active mode (mW)	Low power mode (mW)
Gate driver + current sensor	33.0	0.055
LDO for gate driver and current sensor	39.0	0.065
LDO for MCU	27.2	1.74
MCU	10.3	0.66
Total	109.5	2.52

The system is excited with 2mm, 2Hz excitation to measure the input and output power. The input power to the proposed one-stage power converter is 3.05W, and the output power comes out to be 2.74W. Thus, the efficiency of proposed one-stage power converter is 89% but considering the output power (3.18 W) at optimum resistive load, efficiency is 86%. The controller power consumption consists of the losses due to the gate driver, current sensor, LDOs (5.5V for gate driver and current sensor, 3.3V for MCU), and MCU. Table II shows the power consumption breakdown from different components. The average power dissipation of the controller is measured with a multi-meter (Fluke-87-V Series). The total controller power loss comes out to be ~109mW. The controller power consumption in active and low power mode are given in Table II. The total controller loss comes down to 2.52 mW as the PWM and gate driver get disabled and the clock frequency reduces. About 107 mW of power is saved in the low power mode.

The total loss of proposed circuit can be divided into two parts: buck-boost converter losses (conduction loss, switching loss, inductor loss, diode loss) and controller losses (gate driver loss, LDO, current sense resistor loss, current sensor loss and MCU loss). From the efficiency results, the total loss of the circuit at 2Hz excitation is 440 mW while outputting 2.74W. It is important to note that buck-boost converter losses (or any other DC-DC converter) will be present in any energy harvesting circuit that uses switch-mode power supply. Table III shows the comparison of controller power loss and efficiency between a previous work on similar application using analog control [27] and proposed DC-DC converter. The reported controller power loss was 224mW [27], which is about 2 times higher than controller for the proposed converter, and the efficiency of proposed one-stage power converter is 86%, which is 12% higher than the one-stage power converter in [27].

TABLE III
COMPARISON OF CONTROLLER POWER LOSS AND EFFICIENCY

		[27]	Proposed circuit
Controller power loss	Active mode	224 mW	109 mW
	Sleep mode	N/A	2.52 mW
Efficiency of DC-DC converter		74%	86%

V. CONCLUSION

In this paper, a digitally controlled power management circuit with multi-functional single-stage power converter is proposed for vibration energy harvesting in a vehicle suspension system. Topology of a one stage DC-DC converter is adapted, and it is digitally controlled and designed to maximize the power output while reducing power consumption of the controller with consideration of intermittent vibration sources. In addition, digital controller protects the generator and battery charging when it is needed. The proposed digital controller decides the operating mode of proposed one-stage DC-DC converter: impedance matching mode, voltage regulation, generator protection, low power mode. The proposed PMC is realized in a printed circuit board (PCB) and connected with the MMR electromagnetic energy harvester to evaluate the performance of proposed PMC.

Experimental results show that proposed circuit in impedance matching mode can successfully match the impedance for maximum power transfer in different input conditions. In the voltage regulation mode, proposed PMC maintains a fixed output voltage of 13V at different input voltages from 3V to 20V for battery protection. When the harvester does not generate sufficient output power, the

2168-6777 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:58:16 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2021.3123174, IEEE Journal of Emerging and Selected Topics in Power Electronics

controller detects this event, and the proposed PMC switches from active to low power mode to minimize power consumption of the circuit. In the evaluation of PMC performance, the efficiency of the proposed one-stage power converter is 86%, which is 12% higher than the one-stage power converter in [27], and the total controller power loss is less than 50% compared to previous work [27].

In this paper, we adapt the commercial passive three phase full-bridge rectifier for AC-DC conversion. It has 550 mV voltage drop, which will resulted in a power loss based on V (voltage drop) × I (input current). In our previous simulation work for maximizing the power extraction from train suspension energy harvesting system using conjugate impedance matching [31], it was shown that 17-45% of power output is increased compared to resistive impedance matching. The complicated impedance matching system could consume additional power, but it is worth designing the power management circuit in the future. In this paper, the open-loop control without feedback in the impedance matching mode is used, but impedance of energy harvester can be changed as temperature changes, hence the controller with temperature sensing can be an alternative method to improve the accuracy of the impedance matching. In addition, tracking the maximum power point using P&O algorithm considering dynamic vibration of railcar can be used for maximum power transfer [32]. In summary, future work should include 1) the active three-phase full-bridge rectifier to minimize the forward voltage drop for reducing rectifier power loss 2) the minimization of controller loss, and 3) feedback control of the impedance matching to maximize the power output at various vibration inputs, considering the various inductive impedance values.

ACKNOWLEDGMENT

The authors thank the funding support from NSF #1738689 and CRRC Yangtze Co., Ltd. We also thank Yu Pan, Teng Lin, Hongjip Kim and Weihan Lin for help with the mechanical hardware.

REFERENCE

[1] Bernal E, Spiryagin M, Cole C. Onboard condition monitoring sensors, systems and techniques for freight railway vehicles: a review. IEEE Sens J 2019; 19: 4–24.

[2] Li C, Luo S, Cole C, Spiryagin M. An overview: modern techniques for railway vehicle on-board health Veh Syst Dyn 2017; 55: 1045–70.

[3] J. Engelbrecht, M. Johannes Booysen, G. Rooyen, FJ Bruwer, Survey of smartphone-based sensing in vehicles for intelligent transportation system applications, IET Intelligent Transport Systems, 2015, 9, 10: 924 - 935.

[4] Y Pan, T Lin, F Qian, C Liu, J Yu, J Zuo, L Zuo. Modeling and field-test of a compact electromagnetic energy harvester for railroad transportation. Applied Energy. 2019; 247: 309-321.

[5] Hrovat D. Survey of advanced suspension developments and related optimal control applications. Automatica 1997; 33: 1781–817.

[6] Abdelkareem MA, Makrahy MM, Abd-El-Tawwab AM, EL-Razaz A, Ali MKA, Moheyeldein M. An analytical study of the performance indices of articulated truck semi-trailer during three different cases to improve the driver comfort. Proc Inst Mech Eng Part K J Multi-body Dyn 2018; 232: 84–102.

[7] Karnopp D. Active and semi-active vibration isolation. J Vib Acoust 1995; 117: 177–85.

[8] D Maurya, P Kumara, S Khaleghian, R Sriramdas, MG Kang, RA Kishore, V Kumard, H Song, J Park, S Taheri, and S Priya. Energy harvesting and strain sensing in smart tire for next generation autonomous vehicles. Applied Energy 2018; 232: 312-322.

[9] Z Yang, S Zhou, J Zu, and D Inman. High-Performance Piezoelectric Energy Harvesters and Their Applications. Joule 2018; 2: 642-697.

[10] X Ma, PK Wong, J Zhao. Practical multi-objective control for automotive semi-active suspension system with nonlinear hydraulic adjustable damper. Mechanical Systems and Signal Processing. 2019; 117: 667-688.

[11] Y Shoukry, MW El-Kharashi, and S Hammad. MPC-On-Chip: An Embedded GPC Coprocessor for Automotive Active Suspension Systems. 2010; 2 :31-34.

[12] A Stribrsky, K Hyniova, J Honcu, and A Kruczek. Energy recuperation in automotive active suspension systems with linear electric motor. Mediterranean Conference on Control & Automation, Athens, Greece, June 27-29, 2007.

[13] A Aly, and F A. Salem. Vehicle Suspension Systems Control: A Review. International Journal of Control, Automation and Systems. 2013; 2: 46-54.

[14] M Abdelkareem, L Xu, M Ali, A Elagouz, J Mi, S Guo, Y Liu, L Zuo, Vibration energy harvesting in automotive suspension system: A detailed review. Applied Energy. 2018; 229: 672-699.

[15] Y Pan, F Liu, R Jiang, Z Tub, L Zuo, Modeling and onboard test of an electromagnetic energy harvester for railway cars. Applied Energy. 2019; 250: 568-581.

[16] K. H. Lam, Solar energy harvesting for electric vehicles Energy Systems for Electric and Hybrid Vehicles, 2016, p. 129 – 153.

[17] A. Pirisi, M. Mussetta, F. Grimaccia, D. Caputo, G. Gruosso, R.E. Zich, An innovative device for traffic energy harvesting 6th IET International Conference on Power Electronics, Machines and Drives (PEMD 2012), 2012, p. 53 – 53.

[18] Li C, Tse PW. Fabrication and testing of an energy-harvesting hydraulic damper. Smart Mater Struct 2013; 22: 065024.

[19] Galluzzi R, Xu Y, Amati N, Tonoli A. Optimized design and characterization of motor-pump unit for energy-regenerative shock absorbers. Applied Energy 2018; 210: 16–27.

[20] X He, G Xiao, B Hu L Tan, H Tang, S He, Z He. The applications of energy regeneration and conversion technologies based on hydraulic transmission systems: A review. Energy Conversion and Management. 2020; 205: 112413.

[21] C. Nagode, M. Ahmadian and S. Taheri. Vibration-based energy harvesting systems for on-board applications, Joint Rail Conference, Pueblo, CO, USA, March 16-18, 2011.

[22] D. Dinulovic, M. Shousha, M. Al-Batol, T. Zafar, J. Bickel, H. Ngo, and M. Haug, "Dual-Rotor Electromagnetic Based Energy Harvesting System for Smart Home Applications," in IEEE Transactions on Magnetics, vol., no., pp. , 2020.

2168-6777 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:58:16 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2021.3123174, IEEE Journal of Emerging and Selected Topics in Power Electronics

[23] Z Li, L Zuo, G Luhrs, L Lin, and Y Qin. Electromagnetic Energy-Harvesting Shock Absorbers: Design, Modeling, and Road Tests. IEEE Transactions on Vehicular Technology. 2013; 62: 1065–1074.
[24] Mahmoud Shousha, Dragan Dinulovic, Martin Haug, Tomislav Petrovic, and Abdelmomen Mahgoub, A power management system for electromagnetic energy harvesters in battery/batteryless applications, IEEE Journal of Emerging and Selected Topics in Power Electronics, 2020; 8: 3644-3657.
[25] Andrew Jordan Dickson, Sarah Burton, Michael Shepertycky, Yan-Fei Liu, and Qingguo Li, Digitally Controlled Energy Harvesting Power Management System, IEEE Journal of Emerging and Selected Topics in Power Electronics, 2016; 4: 303-317.
[26] Danilo Porcarelli, Dora Spenza, Davide Brunelli, Alessandro Cammarano, Chiara Petrioli, Luca Benini, Adaptive Rectifier Driven by Power Intake Predictors for Wind Energy Harvesting Sensor Networks, IEEE Journal of Emerging and Selected Topics in Power Electronics, 2015; 3: 471-482.
[27] TO Connor. Design of a Power Management Circuit for Vibration Energy Harvesting from Freight Railcars. MS Thesis, Electrical and Computer Eng., Virginia Tech, May 2015.
[28] T Guo, R Kerley, DS Ha. Development of a Power Conditioning Circuit for Railcar Energy Harvesting. IEEE 56th International Midwest Symposium on Circuits and Systems (MWSCAS), Columbus, OH, USA, Aug. 4-7, 2013.
[29] J Wang, AJ Dancy, and DS Ha. Vibration Energy Harvesting Circuit with Impedance Matching and Wake-up for Freight Railcars. 44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, USA, Oct. 21-23, 2018.
[30] N Putra, B Ariantara. Electric motor thermal management system using L-shaped flat heat pipes. Applied Thermal Engineering. 2017; 126: 1156-1163.
[31] L Costanzo, M Vitelli, Y Pan, L Zuo. Maximizing the Power Extraction from Train Suspension Energy Harvesting System. ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Anaheim, California, USA, Aug. 18–21, 2019.
[32] N Kong, DS Ha. Low-Power Design of a Self-powered Piezoelectric Energy Harvesting System With Maximum Power Point Tracking. IEEE Transactions on Power Electronics. 2012; 27: 2298-2308.
[33] L Costanzo, T Lin, W Lin, AL Schiavo, M Vitelli, L Zuo, Power Electronic Interface with an Adaptive MPPT Technique for Train Suspension Energy Harvesters, IEEE Transactions on Industrial Electronics, 2020, DOI: 10.1109/TIE.2020.3009584
[34] R Erickson and D Maksimovic. Fundamentals of Power Electronics, 2020, Springer Nature (Chapter: Digital Control of Switched-Mode Power Converters).



Hyunjun Jung received his Bachelor degree in electrical engineering from Seoul National University of Science and Technology, Seoul, South Korea in 2012. He was an exchange PhD student at the department of electrical and computer engineering in Virginia Tech, and received his Ph.D in electrical engineering from Hanyang University, Seoul, South Korea, in 2017. He was a postdoctoral researcher at Energy harvesting & optimization lab, University of Maryland Baltimore County from 2017-2019. Since 2019, he was a postdoctoral researcher at the Center for Energy Harvesting Materials and Systems, Virginia Tech, and he is now with Pacific Northwest National Lab (PNNL). His research interest includes design of energy harvesting device, low power management circuit, and wireless sensor network.



Yamini Sharma was born in Delhi, India in 1992. She received her B. Tech in Electronics and Communication Engineering in 2016 from Delhi Technological University (formerly Delhi College of Engineering) in India. She worked as a System Engineer in Texas Instruments India, Bangalore from 2016-2018 in power delivery. She received her Master degree in Electrical Engineering at Center for Energy Harvesting Materials and Systems in Virginia Polytechnic Institute and State University, USA in 2020. She is currently working in Apple Inc, California as Hardware Development Engineer in wireless charging. Her research interests include high frequency power conversion and high density magnetic integration.



Lei Zuo received his BS from Tsinghua University in 1997 and MS and PhD degrees from MIT in 2002 and 2005. He is currently an endowed professor in mechanical, electrical, and ocean engineering at Virginia Tech and directs the NSF Industry-University Cooperative Research Center for Energy Harvesting Materials and Systems (CEHMS). He also holds a joint appointment as Chief Scientist at Pacific Northwest National Lab (PNNL). His research interest includes energy harvesting, marine renewables, mechatronics design, vibration control, smart structures, and advanced manufacturing. He has published over 340 papers and advised 66 PhD and Masters to completion their degrees.

2168-6777 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:58:16 UTC from IEEE Xplore. Restrictions apply.

Australasian Accounting, Business and Finance Journal

Volume 15
Issue 5 -*Special Issue - Sustainable Development Goals and Businesses as Active Change Agents.*

Article 3

2021

Does Gender Diversity on Board Promote Corporate Social Responsibility? An Empirical Analysis of Sustainable Development Goals

Amit Kumar Singh
University of Delhi, India

Hima Bindu Kota
Amity University, India

Varda Sardana
University of Delhi, India

Shubham Singhania
Delhi Technological University, India

Follow this and additional works at: https://ro.uow.edu.au/aabfj

Copyright ©2021 Australasian Accounting Business and Finance Journal and Authors.

Recommended Citation
Singh, Amit Kumar; Kota, Hima Bindu; Sardana, Varda; and Singhania, Shubham, Does Gender Diversity on Board Promote Corporate Social Responsibility? An Empirical Analysis of Sustainable Development Goals, *Australasian Accounting, Business and Finance Journal*, 15(5), 2021, 22-40. doi:10.14453/aabfj.v15i5.3

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Does Gender Diversity on Board Promote Corporate Social Responsibility? An Empirical Analysis of Sustainable Development Goals

Abstract

Recent regulatory changes in India require the firms to improve the appointment of female directors on corporate boards, and it is believed that such a regulation would prove to be a boon in terms of strategic decision making. The Board Capital Theory advocates that the appointment of women directors on board shall enhance various dimensions of the board capital breadth and help in better decision making. With growing consciousness for sustainable practices throughout the globe, it is pertinent to see whether the gender diverse boards can promote corporate social responsibility and create a business case for their upsurge, as it would give room for policy implications. This study investigates the impact of gender diverse boards on promoting corporate social responsibility, using multivariate regression with a sample of NIFTY 50 Index for the period 2014-2019. The study found insignificant positive relation among gender-diverse boards and sustainability. To check for the robustness of the study, we have used two diversity indices, Blau & Shannon index, to supplement our results.

Keywords

Gender Diversity, Sustainability, Corporate Governance, NIFTY 50 Companies.

This article is available in Australasian Accounting, Business and Finance Journal: https://ro.uow.edu.au/aabfj/vol15/iss5/3

Does Gender Diversity on Board Promote Corporate Social Responsibility? An Empirical Analysis of Sustainable Development Goal

Amit Kumar Singh[1], Hima Bindu Kota[2], Varda Sardana[3] and Shubham Singhania[4]

Abstract: Recent regulatory changes in India require the firms to improve the appointment of female directors on corporate boards, and it is believed that such a regulation would prove to be a boon in terms of strategic decision making. The Board Capital Theory advocates that the appointment of women directors on board shall enhance various dimensions of the board capital breadth and help in better decision making. With growing consciousness for sustainable practices throughout the globe, it is pertinent to see whether the gender diverse boards can promote corporate social responsibility and create a business case for their upsurge, as it would give room for policy implications. This study investigates the impact of gender diverse boards on promoting corporate social responsibility, using multivariate regression with a sample of NIFTY 50 Index for the period 2014-2019. The study found insignificant positive relation among gender-diverse boards and sustainability. To check for the robustness of the study, we have used two diversity indices, Blau & Shannon index, to supplement our results.

Keywords: Gender Diversity, Sustainability, Corporate Governance, NIFTY 50 Companies.

JEL Classification: G30, G38, J16, M14

[1] *University of Delhi, India.*
[2] *Amity University, India*
[3] *University of Delhi, India.*
[4] *Technological University, India.*

1. INTRODUCTION

Corporate Social Responsibility (CSR) reporting refers to the act of communicating the effect of various economic actions undertaken by the firms on the social and environmental domain. In simple words, we can infer that any impact caused by the firms towards the society and environment has to be informed to all the stakeholders (Jamali and Neville, 2011; Jamali, El Dirani and Harwood, 2015). The Iron Law of Responsibility states that the corporate entities shall make voluntary efforts to make valuable contribution towards the society, since they have been recklessly using society's resources. In this context, it becomes imperative for the firms to make contributions towards the society. To promote such measures by the firms, the Companies Act 2013 has made it compulsory for some companies to contribute 2% of the average net profits towards CSR related activities.

Since all major decisions regarding the management of the firms are undertaken by the board of directors, they are imperative for promoting CSR related issues (Muttakin, Khan and Mihret, 2018). Owing to various debacles taking places across the globe, the board composition has undergone a sea change, and the research studies on corporate governance has seen an upsurge. With growing consciousness throughout the world, the major changes in composition have been in terms of making the board diverse, with a specific emphasize on the inclusion of females- through mandatory or voluntary provisions. Amongst the various boards related components, gender diversity has received growing interest from researchers.

The Sustainable Development Goals of the United Nations, specifically SDG-5 has also emphasized the growing role of gender equality on various dimensions be it business, academic, political & social arena. Women have been acting as change agents as their presence in the process of making decisions tends to influence the board strategy (Nielsen and Huse, 2010b) and their presence have made things more independent & objective (Fondas, 2000). Some authors also tend to believe that they bring upon a different perspective, different outlook, various alternatives in terms of decision making and leads to problem solving(Rao and Tilt, 2016).In the light of the benefits it is believed that gender equality on board would prove to be highly beneficial not only as a matter of social upliftment of women but also in economic sense.

Countries like Norway, Sweden and Denmark have mandated the appointment of female directors on the board of companies. On similar lines, in India, the Companies (Amendment) Act, 2013 requires the board of certain category of companies to have at least one-woman director. According to Moreno-Gómez, Lafuente and Vaillant (2018), social perspective of feminist groups means that females tends to choose strategies that are different from that of males and hence, gender diversity plays a big role in decision making. Board gender diversity may present a lot of opportunities in terms of having a diverse opinion, and it might affect the CSR related disclosures and other essential issues, since it is unnatural to have a homogenous set of stakeholders for any business. Taking these studies as the basis, we can postulate that the working style of female directors is different from that of their male counterparts (Alonso-Almeida, Perramon and Bagur-Femenias, 2017). Therefore, having gender diverse boards would certainly help in better CSR related decisions.

Previous research studies in the field of gender diversity have looked it from the perspective of financial performance (Campbell and Mínguez-Vera, 2008; Abdelzaher and Abdelzaher, 2019; Herrera-Cano and Gonzalez-Perez, 2019). But there has not been much research contributions in terms of investigating whether gender diverse boards affect corporate social responsibility.

For the Indian scenario, Sanan (2016) in her study looked at the impact of presence of both males and females on boards on social performance to capture the socially responsible behaviour of the firms, by creating a CSR index. However, it did not give significant results. Motivated by the fact that not much studies have been done in this respect, it is pertinent for the interest of various stakeholders to know whether the presence of women directors on corporate boards have any significant business case or they are a mere form of tokenism.

Therefore, this study focusses on evaluating the impact of gender diversity on board in promoting CSR. Section 2 outlines the conceptual model developed for the study. Section 3 highlights the extensive literature review on gender diversity & CSR. Section 4 mentions about the research gap based on the extensive literature. Section 5 and 6 covers the objective and hypothesis. Section 7 looks after the research methodology. Section 8 and 9 elaborates on the results and conclusions, and lastly, section 10 summarises the policy implications, future scope, and limitations of the study.

2. CONCEPTUAL MODEL FOR THE STUDY

The systematic review of several studies based on theories such as agency, stakeholder, resource dependence, and legitimacy theory postulates how gender diversity on board can affect the decisions with respect to corporate social responsibility and promote sustainable development. In a large number of studies, the results have suggested that the presence of women directors on board have impacted the CSR positively (Bear, Rahman and Post, 2010). There are evidences which also report that the presence of gender diversity on board impacts several dimensions of CSR, such as the level of charity by the firms (Wang and Coffey, 1992) as well as the environment based CSR (Post, Rahman and Rubow, 2011). Thus, in this study, we have used CSR expenditure as a proxy variable to measure the impact caused by gender diversity on board in promoting corporate social responsibility.

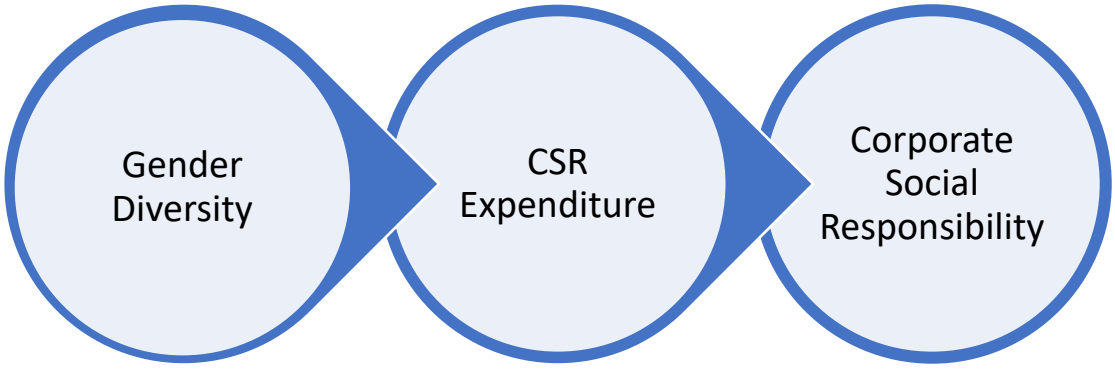


Figure.1 Conceptual Model on Gender Diversity & Corporate Social Responsibility

3.REVIEW OF LITERATURE

3.1 Role of Gender Diversity on Board

Diversity, in terms of gender, on corporate boards in a firm, can influence decision making in

various ways. As per Nielsen and Huse (2010a), board of directors usually work in groups or teams, and hence variation in a group can certainly lead to rise in knowledge, abilities as well as information. Increased innovation, novice ideas, greater market penetration capability, efficiency in problem solving & corporate leadership, efficient global relationships; all of these benefits can accrue to a firm on account of gender diversity in its leadership (Robinson and Dechant, 1997). The existence of women on board may help in improving the image of the firm, which would ultimately have a positive impact on customer behaviour (Smith, Smith and Verner, 2006). Women are found to have provided uncommon perspectives, and bring in unique experiences as well as working to the firm, while making decisions (Daily and Dalton, 2003). Female directors have an inverse relation with accounting manipulations (Abdullah, Ismail and Nachum, 2016) as well, and they are more inclined towards asking questions, so that the decisions are not finalised without adequate discussions (KONRAD, KRAMER and ERKUT, 2008). Stakeholder and Gender Socialisation Theory advocates that heterogeneity in the top management will certainly lead to the better decision-making quality (Chams and García-Blandón, 2019). Rahman, Ibrahim and Che Ahmad (2017) also found that heterogeneity at the top-level management plays a role in reducing the level of agency conflicts within the company.

On the other hand, a homogenous board, with directors having similarity on thought, shall certainly lead to congruence in perspective and conformity in decision making (Van Knippenberg, De Dreu and Homan, 2004). It has also been observed that heterogenous boards achieve better during uncertainty, whereas homogenous boards achieve better in stable conditions (Nielsen and Huse, 2010a). Any negative effects of board gender diversity, such conflicts, may get moderated overtime when directors become aware about each other (Harjoto, Laksmana and Lee, 2015).When it comes to sustainability, Nadeem, Zaman and Saleem (2017) suggest that women's representation on board would improve decision making in the context of sustainability as the diversity present shall together bring diverse range of expertise and knowledge. Female representation on board brings in different work styles on board and leads to better adaptation to the environment (Mínguez-Vera and López-Martínez, 2010).

3.2 Corporate Social Responsibility of Companies and their Disclosure

As per Liao *et al.* (2018), corporate sustainability disclosures are meant to consider the contribution of companies or entities in broadly 3 parameters: social, economic sustainability & environmental aspect. In India, Clause 55 of Listing Agreement requires firms to have Business Responsibility Reports (BRR) based on the Framework of National Voluntary Guidelines (NVG). However, there is limited literature and studies undertaken as to disclosure of CSR requirements by corporates.

While undertaking disclosures, legitimacy theory, that is congruency between the society's expectations and company's value systems plays an important role. The quality of such disclosures would also be subjective and context dependent (Beattie, McInnes and Fearnley, 2004).According to a study by Prasad, Mishra and Kalro (2017), the extent of disclosures have increased from 2011-12 to 2014-15, but they are not significant. Moreover, the quality of disclosures has also improved, but like quantity, this is also not significant.

3.3 Women and Sustainability Practices

Multiple studies have pointed towards the outlook of women in general, as well as women directors towards corporate sustainability and social responsibility. Males and females differ in their values especially when it comes to the corporate social performance (Post, Rahman and Rubow, 2011). It has also been advocated that firms with women representation on the

boards are found to violate less environmental concerns (Donaldson and Preston, 1995). According to Stakeholder Theory, women possess some communal qualities such as helpfulness, kindness, sympathy, interpersonal sensitivity, etc. which may facilitate their say, based on stakeholders claims and expectations (Eagly, Johannesen-Schmidt and Van Engen, 2003).

Women's communal qualities also help them in taking decisions, considering their social responsibility (Tourigny, Han and Baba, 2017). This is reiterated by the Social Role Theory as well, as per which women are shown to portray communal qualities (such as generosity, social orientation, concern) while men showcase agentic qualities (such ambitious, self-directed) (Doherty and Eagly, 1989; Eagly and Wood, 1991). Men are prone to pay emphasise on their tasks, whereas women are social butterflies and are oriented towards others (Major and Forcey, 1985; Eagly and Karau, 1991) While men are inclined towards following and promoting fairness and obligations, women are more likely to have long term relationships and caring for others' needs (Gilligan, 1977). Hence, presence of women directors on board, even if one, shall create a difference in the sustainability practises of the firm (Zaichkowsky, 2014).

The reason for this is that women directors are relatively less power-hungry, and hence they show strong traits like kindness & compassion. This enables them look up to the protection of the nature and other people (Adams and Funk, 2012). Women are considered to be more ethical, have better communication skills, and better participation which enables them to have better concern for society and environment, and hence better CSR reporting (Kesner, 1988).

Agency Theory postulates that information failure as well as agency conflicts on the board shall reduce due to the presence of women directors, thereby improving the quality of corporate social responsibility (Reguera-Alvarado, de Fuentes and Laffarga, 2017). The Resource Dependency Theory advocates that the presence of female directors shall turn out to be fruitful for the CSR disclosures, since they would, through their connections, maintain legitimacy between the society and stakeholders (Lückerath-Rovers, 2013). Additionally, the Gender Socialisation Theory believes that the leadership styles adopted by female directors is more ethical and socially inclined than that of male directors, and hence they are more stakeholder-oriented, which enhances the quality of CSR reporting (Landry, Bernardi and Bosco, 2016).

According to Zahid *et al.* (2020), diversity on board leads to a significant impact on the corporate sustainability disclosures (CSD), and CSD was found to be better in firms where the women were present in the top management. In another study Al Fadli *et al.*, (2019) found that the presence of female directors on board of Jordanian public listed companies enhanced CSRT reporting practises. Similarly, as per Pucheta-Martínez, Bel-Oms and Olcina-Sempere (2019), increase in the ratio of women directors (independent & institutional directors) on boards, up to a certain threshold limit would lead to better CSR disclosure. However, increasing the proportion beyond the threshold limit may lead to fall or decrease in the CSR disclosures. One exception to this is the study conducted by Sanan (2016), which did not find any significant impact of gender based diversity on firms financial and social performance.

4. RESEARCH GAP

The past literature makes it clear that gender diversity & financial performance is a well-researched area, both globally as well as in the Indian context. Much of the earlier literature has focussed on finding the nexus between corporate governance and financial performance (Rahman, Ibrahim and Ahmad, 2017). However, there is not much literature which taps the relation between gender diversity and sustainability or corporate social responsibility. There is

a dearth of research, specifically in the Indian context. Hence, as per our knowledge, this study is amongst the first ones that tries to ascertain the effect of gender diversity and presence of women on company boards, on the sustainability and socially responsible behavior of companies.

Due to differences in institutional and cultural environment of various developing & developed countries various attributes of corporate governance, in the form of board diversity and the like, may vary. Byron and Post (2016) have documented that institutional level factors will certainly affect the outcomes of the studies in different countries. In India, the economy has been rooted in terms of socio-cultural barriers for women, whereas in the West side, the women's life in terms of the working, lifestyle and education is not entirely based on the social and cultural norms. This makes it even more essential to undertake such a study in a developing country, like India.

5. OBJECTIVES

The major aim of the study is to determine the impact of gender diversity on large firms' board, on promoting corporate social responsibility.

6. HYPOTHESIS

H_{o1}: The proportion of women on board has no effect on the CSR expenditure of firm.
H_{a1}: The proportion of women on board affects CSR expenditure of firm.
H_{o2}: Leverage in a firm has no effect on the CSR expenditure of firm.
H_{a2}: Leverage in a firm affects the CSR expenditure of firm.
H_{o3}: The size of a firm has no effect on the CSR expenditure of firm.
H_{a3}: The size of a firm affects the CSR expenditure of firm.
H_{o4}: Return on assets of a firm has no effect on the CSR expenditure of firm.
H_{a4}: Return on assets of a firm affects the CSR expenditure of firm.
H_{o5}: Age of a firm has no effect on the CSR expenditure of firm.
H_{a5}: Age of a firm affects the CSR expenditure of firm.
H_{o6}: Board size of a firm has no effect on the CSR expenditure of firm.
H_{a6}: Board size of a firm affects the CSR expenditure of firm.

7. RESEARCH METHODOLOGY

7.1 Data Collection Sources

We have used five sources for the purpose of data collection. These are as follows:
1. Prime Database Group: The database was used to collect complete data on directors of the corporates and composition of boards for various Indian firms listed on National Stock Exchange (NSE).
2. Indianboards.com: This was used as a supplementary source to collect data regarding board of directors of firms in our sample.
3. CMIE PROWESS: Prowess was used to collect indicators and measures of financial performance and CSR performance of firms on NSE.
4. Capitaline Database: This database was tapped for sourcing annual reports of companies.
5. Corporate Governance & CSR Reports of the firms: These reports were used to collect data regarding CSR expenditure of firms in our sample.

7.2 Sample

We started with the NIFTY 50 companies for the analysis, which have been taken as the large firms. Out of these, banking companies were eliminated as these companies follow the Banking

Regulations Act, 1949, and hence have different regulations regarding CSR. After removing the companies with missing data, a total of 42 firms were finally included for the study.
A five-year sample period, from 2014 to 2019 was considered to measure the impact of gender diversity on corporate sustainability of firms, thereby leading to a total of 210 data points. The period beginning from 2014 has been considered because the mandatory compliance with CSR requirements became a part of the Companies Act in the year 2013.

7.3 Variables

Dependent Variable

Corporate Social Responsibility has become the buzzword in the current scenario, and efforts are being made in all respect by corporate houses to make valuable contribution to the society at large. Therefore, to measure the efforts undertaken by these firms, we have used CSR expenditure incurred by the firms as the dependent variable.

CSR Expenditure

Corporate Social Responsibility is measured as the expenditure incurred by a company on social as well as sustainability related issues. This variable shall act as a proxy for Corporate Social Responsibility initiatives taken by the firm. Though other studies have used the CSR ratings provided by some credible database sources (Francoeur *et al.*, 2019) or have created CSR reporting index (Ahmad, Rashid and Gow, 2017) using content analysis of the annual reports to measure the CSR based aspects, in case of India such ratings are not available as of now, hence CSR expenditure as a variable has been widely used in several studies (Verma and Vijaya Kumar, 2014; Prasad, Mishra and Bapat, 2019).

Independent Variables

Table 1: Summary of Independent Variables

Independent Variable	Measure	Literature
P-Women	Percentage of women directors on board with respect to the total directors.	(Campbell and Mínguez-Vera, 2008; Giraldez-Puig and Berenguer, 2018; Haldar *et al.*, 2018; Chancharat and Chancharat, 2019; Hashim, Ahmed and Huey, 2019)
Blau Index	A measure of gender diversity	(Campbell and Mínguez-Vera, 2008; Issa and Fang, 2019; Singh, Singhania and Sardana, 2019)
Shannon Index	A measure of gender diversity	(Campbell and Mínguez-Vera, 2008; Issa and Fang, 2019)
Leverage	Calculated as the ratio of total long-term debt to the total assets.	(Sanan, 2016; Karim, Manab and Ismail, 2019; Pucheta-Martínez and Gallego-Álvarez, 2019)
Firm Size	Measured in terms of the natural log of the total assets	(Karim, Manab and Ismail, 2019; Pucheta-Martínez, Bel-Oms and Olcina-Sempere, 2019)
Profitability (Return on Asset)	Percentage of profits earned against the total assets employed by a firm	(Pucheta-Martínez, Bel-Oms and Olcina-Sempere, 2019; Valls Martínez, Cruz Rambaud and Parra Oller, 2019)

Age	Measured as the number of years for which the firm has lived from the date of its incorporation.	(Francoeur *et al.*, 2019; Zahid *et al.*, 2020)
Board Size	Number of directors' present in the Board of the firm.	(Sanan, 2016; Charumathi and Rahman, 2019; Issa and Fang, 2019; Pucheta-Martínez, Bel-Oms and Olcina-Sempere, 2019)

Proportion of women on board (P-woman)

P-woman refers to the proportion of female directors on the corporate boards. This variable has been used as a proxy measure of gender diversity on the board of the firms. Literature suggests that existence of women directors on board leads to a positive impact on organisational innovation, diversity in perspectives and hence better decision making (Torchia, Calabrò and Huse, 2011). Board diversity also provides opportunities for strategic alliances and better relationship with other firms and corporate groups (Haldar, Shah and Nageswara Rao, 2015). Hence, various studies have used P-woman to measure for gender diversity on board and its impact on various firm related aspects (Campbell and Mínguez-Vera, 2008; Giraldez-Puig and Berenguer, 2018; Haldar *et al.*, 2018)

BLAU Index

The Blau index is a well know variable used to measure diversity. It is calculated as:

$$D = 1 - \sum_{i=1}^{N} p_i^2$$

Where p denotes the proportion of objects in the categories and
N denotes the number of categories.
The value of Blau Index varies from 0 to 0.5 (lowest to highest diversity). It has been employed in order to improve the the robustness of the study (Campbell and Mínguez-Vera, 2008; Issa and Fang, 2019; Singh, Singhania and Sardana, 2019).

SHANNON Index

The Shannon index aims to state the uncertainty in predicting the species identity of any person chosen randomly from a given set of data. It is calculated as:

$$H' = -\sum_{i=1}^{S} p_i \ln p_i$$

Where p denotes the proportion of objects in the categories and
S denotes the number of categories.
The value of Shannon Index varies from 0 to 0.69 (lowest to highest diversity). Like Blau Index, Shannon Index also enhances the robustness of the study (Campbell and Mínguez-Vera, 2008; Issa and Fang, 2019; Singh, Singhania and Sardana, 2019)

Control Variables

Leverage

Leverage of any organisation is calculated as the ratio of debt to assets. In the previous literature. it has been documented that leverage shall negatively affect CSR (Waddock and Graves, 1997; Campbell, 2007) and has been empirically used in several other studies (Sanan, 2016; Issa and Fang, 2019; Karim, Manab and Ismail, 2019; Pucheta-Martínez, Bel-Oms and Olcina-Sempere, 2019).

Firm Size

Firm size has been measured with respect to the total assets. Based on the extant literature, it has been observed that larger firms are inclined towards behaving in a socially responsible manner (McWilliams and Siegel, 2001; Barnea and Rubin, 2010) and therefore we expect that it is certainly going to affect the corporate social responsibility activities. This variable has been used in multiple past studies as well (Issa and Fang, 2019; Karim, Manab and Ismail, 2019; Pucheta-Martínez, Bel-Oms and Olcina-Sempere, 2019).

Profitability (Return on Assets)

Return on assets is calculated as the percentage of profits earned against the total assets employed in a firm. This variable influences the financial performance and, thus, is likely to affect the corporate social responsibility measures of the firms (Haniffa and Cooke, 2005). The CSR disclosure of profitable firms is also larger as has been used and advocated in various other studies (Pucheta-Martínez, Bel-Oms and Olcina-Sempere, 2019; Valls Martínez, Cruz Rambaud and Parra Oller, 2019).

Age

The age of the firm is measured by the number of years the firm has lived from the date of its incorporation. Also, the age of the firm defines the number of years for which the firm has been in existence and has been using societal resources. Based on the Iron Law of Responsibility, age of the firm should have a positive impact on the CSR (Jo and Harjoto, 2011) and has been used in several studies as well to measure social performance & sustainability issues (Francoeur *et al.*, 2019; Zahid *et al.*, 2020).

Board Size

Board size refers to the number of directors' present on the board of the firm. Board size is a crucial element in the decision making of the firms. It has been suggested that larger boards tends to disclose more CSR related information (Esa and Ghazali, 2012) and therefore it is likely that the board size shall impact the decisions taken with respect to the corporate social responsibility measures undertaken by the firms, and has been used extensively in literature (Sanan, 2016; Issa and Fang, 2019; Karim, Manab and Ismail, 2019; Pucheta-Martínez, Bel-Oms and Olcina-Sempere, 2019).

7.5 Research Technique and Model Specification

The technique employed for the study is the Multivariate Regression Analysis. Ordinary least squares (OLS) method of regression is used for the analysis.

We have employed a balanced panel data as this method allows us to remove any unobservable heterogeneity in the sample. In case where unobservable heterogeneity is present, and it is correlated with our independent variables, we conduct an estimation by fixed effects model. Otherwise, we undertake a random effects model. To check for such kind of correlation, we used the Hausman Test (Hausman and Taylor, 1981), whose results have been displayed in the upcoming sections.

Three models were run using the OLS regression, the specifications of which are as follows:

$$\textit{Model 1- } CSR_{it} = \beta_0 + \Sigma\beta_1 P\text{-}WOMAN_{it} + \Sigma\beta_2 LEVER_{it} + \Sigma\beta_3 F\text{-}SIZE_{it} + \Sigma\beta_4 ROA_{it} + \Sigma\beta_5 AGE_{it} + \Sigma\beta_6 B\text{-}SIZE_{it} + \varepsilon_{it}$$

$$\textit{Model 2- } CSR_{it} = \beta_0 + \Sigma\beta_1 BLAU_{it} + \Sigma\beta_2 LEVER_{it} + \Sigma\beta_3 F\text{-}SIZE_{it} + \Sigma\beta_4 ROA_{it} + \Sigma\beta_5 AGE_{it} + \Sigma\beta_6 B\text{-}SIZE_{it} + \varepsilon_{it}$$

Model 3- $CSR_{it} = \beta_0 + \Sigma\beta_1 SHANNON_{it} + \Sigma\beta_2 LEVER_{it} + \Sigma\beta_3 F\text{-}SIZE_{it} + \Sigma\beta_4 ROA_{it} + \Sigma\beta_5 AGE_{it} + \Sigma\beta_6 B\text{-}SIZE_{it} + \varepsilon_{it}$

where,
CSR refers to CSR expenditure, our proxy for Sustainability Measure.
P- WOMAN represents the percentage of women directors on board.
LEVER refers to Leverage of firm.
F-SIZE refers to the size of the firm.
ROA refers to the Return on Assets.
AGE is the age of the firm.
B-SIZE refers to the size of the firm's board.
BLAU & SHANON are the indices which are our proxy for gender diversity.
i refers to the various firms included in the sample of the time.
t is the time measured in years from 2014 to 2019.

8. RESULTS AND ANALYSIS

Descriptive Statistics

Table 2 clearly depicts that the mean average of the two indices are 0.26 and 0.19, respectively. These low numbers point to the fact that the diversity in the firms' boards is quite less to have any impact on Corporate Social Responsibility. Not only this, but the average proportion of females on board is 17.4% only amongst the 210 firm year observations, which is again low. This points to the fact that there has been no true enhancement in gender diversity in these firms. The average assets are Rs.1590.86 crores and the return on assets is 5.388% which speaks of sound financial health of the firms on an average. The lower leverage ratio, which stand at an average of approximately 0.157, are good as it depicts that less assets are financed by using the debt.

Table 2: Descriptive Statistics

Measure	BLAU	SHANNON	P-WOMAN	TOTAL ASSETS	ROA	LEVER	AGE	CSR	B-Size
MEAN	0.261	0.195266	17.403	1590.856	5.388	0.157	49.94	98.80	12.22
MAXIMUM	0.468	0.661563	37.5	9318.82	22.4	1.947	148	904	25
MINIMUM	0.142	0.271189	7.6923	152.93	-20.22	-0.751	9	0	4
STD. DEV	0.083	0.100128	7.004	1791.165	6.161	0.524	27.86	138.67	3.22

Source: Author's own calculations

Hausman Test

We performed a Hausman test to test for Endogeneity in our models.

Table 3: Hausman Test

Test Summary	Chi-Sq statistic	Chi-Sq-d.f	p-value
Cross section random	21.858	8	0.000

Source: Author's own calculations

The test clearly suggests the use of Fixed Effects Model to run our regression tests.

Table 4: Model Fitness Determination

Models	**R**	**R2**	**Adjusted R2**
1	.706	.499	.484
2	.707	.500	.485
3	.707	.499	.484

Source: Author's own calculations

Table 4 shows the fitness of each of the model. As is clear from the table, all the three models are a fair fit with an adjusted R-squared of 48.4% approximately.

Table 5: Summary Results of Regression Models

Model 1			Model 2			Model 3		
Variable	t-statistic	p-value	Variable	t-statistic	p-value	Variable	t-statistic	p-value
P-woman	0.022	0.982	Blau	-0.394	0.694	Shannon	-0.332	0.740
Leverage	-4.015	0.000*	Leverage	-4.045	0.000*	Leverage	-4.002	0.000*
Firm Size	13.139	0.000*	Firm Size	13.161	0.000*	Firm Size	13.145	0.000*
ROA	3.516	0.001*	ROA	3.549	0.000*	ROA	3.553	0.000*
Age	2.470	0.014**	Age	2.406	0.017**	Age	2.430	0.016**
B-Size	0.001	0.999	B-Size	-0.46	0.963	B-Size	-0.040	0.968

Source: Author's own calculations
***Significant at 1%**
****** Significant at 5%**

Table 5 summarizes the results of our regression models. Significance at 1% and 5% has been tested. Under Model 1, the percentage of women on board has a positive but insignificant impact on the CSR expenditure. This means that gender diversity or presence of women on board of companies does not have any impact on the sustainable or socially responsible behavior of firms as measured by CSR expenditure. The same is reiterated by our Model 2 and Model 3, where the Blau Index and Shannon Index, respectively, also have positive but insignificant impact on CSR expenditure.

As far as the control variable are concerned, majority of them show the expected signs and magnitude of impact under all the models. Leverage, that is the proportion of debt in the firm's total capital has a negative and significant impact on CSR.

Similarly, with the increase in firm size and return on assets, a firm has greater funds at its disposal for CSR initiatives. Hence, they have a positive and significant impact on CSR activities and expenses. As for age of firm, the longer the firm has been in business, the positive is the impact on corporate social expenditure. Lastly, the size of board, that is, the number of directors on board does not seem to have any significant impact on CSR expenditure in our

study.

9. CONCLUSION

The results of our study suggest that gender diversity has a positive impact on CSR expenditure of the firm, although the impact is not significant. However, this also indicates that appointment of women on firms' board has not led to any negative effect on the social responsibility as well. The insignificant relationship points to the fact that such is a mere token exercise. Clearly, the ratio of women is less to influence CSR decisions. Also, gender discrimination and stereotyping challenge is one of the major reasons why women have not been able to make full contribution towards the corporate strategy & leadership positions in their firms (Galbreath, 2011; Issa and Fang, 2019). According to De Cabo, Gimeno and Nieto (2012), there still exists prejudices and judgements towards the abilities of the women directors with respect to the handling of management positions. Moreover, if we diversify the board owing to the external pressures, such kind of diversity is value destroying and firm demeaning (Weisbach, 1998).

According to some of the previous literature (KONRAD, KRAMER and ERKUT, 2008; Torchia *et al.*, 2010) it is said that the critical mass of women directors plays a significant role. These could have been some of the reasons due to which we could not see any significant impact of diversity on CSR activities. Thus, we may conclude that appointment of women on board is done by the companies to just fulfil the requirements of the Companies Act, 2013. In simple words, we can say that the stereotypes and prejudices with respect to the female directors, as well as the low level of diversity, have led to diversity acting as a bane than a boon for the CSR activities. In terms of the control variables, it is found that firm size has positive and significant effect on the CSR initiatives. It is more likely for larger firms to be exposed to the check and balances, and therefore they tend to be more focussed towards the stakeholders, leading to promoting better CSR activities and reputation of the firm (Valls Martínez, Cruz Rambaud and Parra Oller, 2019). Also since large sized firms are better equipped financially, they are able to spend more on CSR related activities (Boulouta, 2013; Sial *et al.*, 2018; Francoeur *et al.*, 2019). Firms use society's resources to a great extent and the more profitable the firm is, the more is the expectations from the firms with respect to the transparency and ethical activities (Báez *et al.*, 2018). Thus, it has been found that firms with better profitability tend to have positive and significant impact on CSR (Dienes and Velte, 2016; Issa and Fang, 2019). Leverage of the firm has also turned out to be positively affecting the CSR based activities since it enhances the firm's capacity to undertake more such initiatives because of the financial stability and debts being covered by the assets of the firms. This is in conformation with prior studies (Issa and Fang, 2019; Pucheta-Martínez, Bel-Oms and Olcina-Sempere, 2019). Age of the firm has also shown significant and positive relationship with the CSR activities, since an older firm is more aware about its reputation and hence takes more careful view of the CSR related activities, as has been documented in the past literature as well (Francoeur *et al.*, 2019).

Board size also has shown a positive but insignificant result. This may be because even after having larger boards, majority of them were male members, and their inclination towards the CSR activities would have been relatively less. Positive relation between board size and CSR has been found in several studies (Issa and Fang, 2019; Pucheta-Martínez, Bel-Oms and Olcina-Sempere, 2019).

10. POLICY IMPLICATIONS, FUTURE SCOPE AND LIMITATIONS

The study, apart from making valuable contributions to the literature, also provides some

recommendations for the policymakers and firms. Firstly, policy makers and firms must consider the regulations in full spirit and not in mere letter form. Just following the law while maintaining the stereotypes and prejudices with respect to the abilities of women directors, is not going to add to the benefits that could have been derived at large because of diversity. Secondly, stricter laws with respect to appointment of independent women directors on boards can bring about a change in how gender diversity is perceived and implemented within firms. Mere appointment of female directors is acting as a hoax since the female directors are related to the promoters and do not bring any advantage on the table due to lack of transparency. Thirdly, the critical mass theory shall be considered while framing policies regarding the minimum female directors that should mandatorily be appointed on the board of certain categories of firms, so that a fair representation can allow women directors to voice their opinions in the boardroom.

As a limitation of this study, the sample includes 42 firms, which makes it less representative of the whole population. A larger sample may give better results. Moreover, in future studies, a larger time horizon and other proxy variables that are more representative of the CSR related elements such as disclosure index, can also be constructed for better and adequate results.

REFERENCES

Abdelzaher, A. and Abdelzaher, D. (2019). Women on Boards And Firm Performance In Egypt: Post The Arab Spring. *The Journal of Developing Areas*, 53(1), 225-241. doi: 10.1353/jda.2019.0013.

Abdullah, S. N., Ismail, K. N. I. K. and Nachum, L. (2016). Does having women on boards create value? the impact of societal perceptions and corporate governance in emerging markets. *Strategic Management Journal*, 37(3), 466-476. doi: 10.1002/smj.2352.

Adams, R. B. and Funk, P. (2012). Beyond the glass ceiling: Does gender matter?. *Management Science*, 58(2), 219–235. doi: 10.1287/mnsc.1110.1452.

Ahmad, N. B. J., Rashid, A., & Gow, J. (2017). Board independence and corporate social responsibility (CSR) reporting in Malaysia. Australasian Accounting, Business and Finance Journal, 11(2), 61-85.doi:10.14453/aabfj.v11i2.5

Al Fadli, A., Sands, J., Jones, G., Beattie, C., & Pensiero, D. (2019). Board gender diversity and CSR reporting: Evidence from Jordan. Australasian Accounting Business and Finance Journal, 13(3), 29-52.doi:10.14453/aabfj.v13i3.3

Alonso-Almeida, M. D. M., Perramon, J., & Bagur-Femenias, L. (2017). Leadership styles and corporate social responsibility management: Analysis from a gender perspective. Business Ethics: *A European Review*, 26(2), 147-161.doi: 10.1111/beer.12139.

Báez, A. B., Báez-García, A. J., Flores-Muñoz, F., & Gutiérrez-Barroso, J. (2018). Gender diversity, corporate governance and firm behavior: The challenge of emotional management. *European Research on Management and Business Economics*, 24(3), 121-129.doi: 10.1016/j.iedeen.2018.07.001.

Barnea, A., & Rubin, A. (2010). Corporate social responsibility as a conflict between shareholders. *Journal of Business Ethics*, 97(1), 71-86.doi: 10.1007/s10551-010-

0496-z.

Bear, S., Rahman, N., & Post, C. (2010). The impact of board diversity and gender composition on corporate social responsibility and firm reputation. *Journal of Business Ethics*, 97(2), 207-221. doi: 10.1007/s10551-010-0505-2.

Beattie, V., McInnes, B., & Fearnley, S. (2004, September). A methodology for analysing and evaluating narratives in annual reports: a comprehensive descriptive profile and metrics for disclosure quality attributes. *Accounting Forum*, 28(3), 205-236). doi: 10.1016/j.accfor.2004.07.001.

Boulouta, I. (2013). Hidden connections: The link between board gender diversity and corporate social performance. *Journal of Business Ethics*, 113(2), 185-197.doi: 10.1007/s10551-012-1293-7.

Byron, K., & Post, C. (2016). Women on boards of directors and corporate social performance: A meta-analysis. *Corporate Governance: An International Review*, 24(4), 428-442.doi: 10.1111/corg.12165.

De Cabo, R. M., Gimeno, R., & Nieto, M. J. (2012). Gender diversity on European banks' boards of directors. *Journal of Business Ethics*, 109(2), 145-162.doi: 10.1007/s10551-011-1112-6.

Campbell, J. L. (2007). Why would corporations behave in socially responsible ways? An institutional theory of corporate social responsibility. *Academy of Management Review*, 32(3), 946-967.doi: 10.5465/AMR.2007.25275684.

Campbell, K., & Mínguez-Vera, A. (2008). Gender diversity in the boardroom and firm financial performance. *Journal of Business Ethics*, 83(3), 435-451.doi: 10.1007/s10551-007-9630-y.

Chams, N., & García-Blandón, J. (2019). Sustainable or not sustainable? The role of the board of directors. *Journal of Cleaner Production*, 226, 1067-1081.doi: 10.1016/j.jclepro.2019.04.118.

Chancharat, S., & Chancharat, N. (2019). Board structure, ownership structure, and performance of Thai listed companies. *Australasian Accounting, Business and Finance Journal*, 13(3), 53-70.doi: 10.14453/aabfj.v13i3.4.

Charumathi, B., & Rahman, H. (2019). Do women on boards influence climate change disclosures to CDP?–evidence from large Indian companies. *Australasian Accounting, Business and Finance Journal*, 13(2), 5-31. doi: 10.14453/aabfj.v13i2.2.

Daily, C.M. and Dalton, D.R. (2003). Women in the boardroom: a business imperative. Journal of Business Strategy, 24(5). https://doi.org/10.1108/jbs.2003.28824eaf.002.

Dienes, D., & Velte, P. (2016). The impact of supervisory board composition on CSR reporting. Evidence from the German two-tier system. Sustainability, 8(1), 63., doi: 10.3390/su8010063.

Doherty, E. G. and Eagly, A. H. (1989). Sex Differences in Social Behavior: A Social-Role Interpretation. *Contemporary Sociology*.18(3), 343. doi: 10.2307/2073813.

Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review*, 20(1), 65-91.doi: 10.5465/amr.1995.9503271992.

Eagly, A. H., Johannesen-Schmidt, M. C., & Van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: a meta-analysis comparing women and men. *Psychological Bulletin*, 129(4), 569.doi: 10.1037/0033-2909.129.4.569.

Eagly, A. H., & Karau, S. J. (1991). Gender and the emergence of leaders: A meta-analysis. J*ournal of Personality and Social Psychology*, 60(5), 685.doi: 10.1037/0022-3514.60.5.685.

Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin*, 17(3), 306-315.doi: 10.1177/0146167291173011.

Esa, E., & Ghazali, N. A. M. (2012). Corporate social responsibility and corporate governance in Malaysian government-linked companies. *Corporate Governance*,13(3),292-305. doi: 10.1108/14720701211234564.

Fondas, N. (2000). Women on boards of directors: gender bias or power threat?. *Women on Corporate Boards of Directors* (pp. 171-177). doi: 10.1007/978-90-481-3401-4_12.

Francoeur, C., Labelle, R., Balti, S., & Bouzaidi, S. E. (2019). To what extent do gender diverse boards enhance corporate social performance?. *Journal of Business Ethics*, 155(2), 343-357. doi: 10.1007/s10551-017-3529-z.

Galbreath, J. (2011). Are there gender-related influences on corporate sustainability? A study of women on boards of directors. *Journal of Management & Organization*, 17(1), 17-38.doi: 10.1017/s1833367200001693.

Gilligan, C. (1977). In a different voice: Women's conceptions of self and of morality. *Harvard Educational Review*, 47(4), 481-517. doi: 10.17763/haer.47.4.g6167429416hg5l0.

Giraldez-Puig, P., & Berenguer, E. (2018). Family female executives and firm financial performance. *Sustainability*, 10(11), 4163. doi: 10.3390/su10114163.

Haldar, A. *et al.* (2018) 'Corporate Performance: Does Board Independence Matter? – Indian evidence', *International Journal of Organizational Analysis*. 26(1),185-200. doi: 10.1108/IJOA-12-2017-1296.

Haldar, A., Shah, R., & Nageswara Rao, S.V.D.(2015). Gender diversity in large listed Indian companies. *Corporate Ownership & Control*, 12(3-5), 573-580. https://doi.org/10.22495/cocv12i3c5p8

Haniffa, R. M., & Cooke, T. E. (2005). The impact of culture and governance on corporate social reporting. *Journal of Accounting and Public Policy*, 24(5), 391-430. doi: 10.1016/j.jaccpubpol.2005.06.001.

Harjoto, M., Laksmana, I., & Lee, R. (2015). Board diversity and corporate social responsibility. *Journal of Business Ethics*, 132(4), 641-660. doi: 10.1007/s10551-014-2343-0.

Hashim, F., Ahmed, E. R., & Huey, Y. M. (2019). Board Diversity and Earning Quality: Examining the Role of Internal Audit as a Moderator. *Australasian Accounting, Business and Finance Journal*, 13(4), 73-91. doi: 10.14453/aabfj.v13i4.6.

Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. Econometrica: *Journal of the Econometric Society*, 1377-1398. doi: 10.2307/1911406.

Herrera-Cano, C. and Gonzalez-Perez, M.A. (2019). Representation of Women on Corporate Boards of Directors and Firm Financial Performance. Diversity within Diversity Management (Advanced Series in Management, Vol. 22), Emerald Publishing Limited, Bingley, 37-60. https://doi.org/10.1108/S1877-636120190000022003

Issa, A. and Fang, H. X. (2019). The impact of board gender diversity on corporate social responsibility in the Arab Gulf states. *Gender in Management*, 34(7), pp. 577–605. doi: 10.1108/GM-07-2018-0087.

Jamali, D., & Neville, B. (2011). Convergence versus divergence of CSR in developing countries: An embedded multi-layered institutional lens. *Journal of Business Ethics*, 102(4), 599-621.doi: 10.1007/s10551-011-0830-0.

Jamali, D. R., El Dirani, A. M., & Harwood, I. A. (2015). Exploring human resource management roles in corporate social responsibility: The CSR-HRM co-creation model. *Business Ethics: A European Review*, 24(2), 125-143. doi: 10.1111/beer.12085.

Jo, H., & Harjoto, M. A. (2011). Corporate governance and firm value: The impact of corporate social responsibility. *Journal of Business Ethics*, 103(3), 351-383.doi: 10.1007/s10551-011-0869-y.

Karim, S., Manab, N. A., & Ismail, R. B. (2019). The dynamic impact of board composition on CSR practices and their mutual effect on organizational returns. *Journal of Asia Business Studies*.doi: 10.1108/JABS-07-2019-0214.

Kesner, I. F. (1988). Directors' characteristics and committee membership: An investigation of type, occupation, tenure, and gender. *Academy of Management Journal*, 31(1), 66-84.doi: 10.5465/256498.

Konrad, A. M., Kramer, V., & Erkut, S. (2008). The impact of three or more women on corporate boards. *Organizational Dynamics*, 37(2), 145-164. doi: 10.1016/j.orgdyn.2008.02.005.

Landry, E. E., Bernardi, R. A., & Bosco, S. M. (2016). Recognition for sustained corporate social responsibility: Female directors make a difference. *Corporate Social Responsibility and Environmental Management*, 23(1), 27-36.doi: 10.1002/csr.1358.

Liao, P. C., Shih, Y. N., Wu, C. L., Zhang, X. L., & Wang, Y. (2018). Does corporate social

performance pay back quickly? A longitudinal content analysis on international contractors. *Journal of Cleaner Production*, 170, 1328-1337.doi: 10.1016/j.jclepro.2017.09.230.

Lückerath-Rovers, M. (2013). Women on boards and firm performance. *Journal of Management & Governance*, 17(2), 491-509.. doi: 10.1007/s10997-011-9186-1.

Major, B., & Forcey, B. (1985). Social comparisons and pay evaluations: Preferences for same-sex and same-job wage comparisons. *Journal of Experimental Social Psychology*, 21(4), 393-405.doi: 10.1016/0022-1031(85)90038-1.

McWilliams, A., & Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective. *Academy of Management Review*, 26(1), 117-127.doi: 10.2307/259398.

Mínguez-Vera, A. and López-Martínez, R., 2010. Female directors and SMES: An empirical analysis. *Journal of Global Strategic Management*, *8*(2), pp.34-46. doi: 10.20460/JGSM.2010415823

Moreno-Gómez, J., Lafuente, E., & Vaillant, Y. (2018). Gender diversity in the board, women's leadership and business performance. *Gender in Management: An International Journal*, 33(2), 104-122. doi: 10.1108/GM-05-2017-0058.

Muttakin, M. B., Khan, A., & Mihret, D. G. (2018). The effect of board capital and CEO power on corporate social responsibility disclosures. *Journal of Business Ethics*, 150(1), 41-56.doi: 10.1007/s10551-016-3105-y.

Nadeem, M., Zaman, R. and Saleem, I. (2017) 'Boardroom gender diversity and corporate sustainability practices: Evidence from Australian Securities Exchange listed firms', *Journal of Cleaner Production*. Elsevier B.V., 149, pp. 874–885. doi: 10.1016/j.jclepro.2017.02.141.

Nielsen, S. and Huse, M. (2010a) 'The contribution of women on boards of directors: Going beyond the surface', *Corporate Governance: An International Review*. 18(2),136-148. doi: 10.1111/j.1467-8683.2010.00784.x.

Nielsen, S. and Huse, M. (2010b) 'Women directors' contribution to board decision-making and strategic involvement: The role of equality perception', *European Management Review*. 7(1),16-29. doi: 10.1057/emr.2009.27.

Post, C., Rahman, N., & Rubow, E. (2011). Green governance: Boards of directors' composition and environmental corporate social responsibility. *Business & Society*, 50(1), 189-223.doi: 10.1177/0007650310394642.

Prasad, M., Mishra, T. and Bapat, V. (2019) 'Corporate social responsibility and environmental sustainability: Evidence from India using energy intensity as an indicator of environmental sustainability', *IIMB Management Review*. Elsevier Ltd, 31(4), 374–384. doi: 10.1016/j.iimb.2019.07.014.

Prasad, M., Mishra, T., & Kalro, A. D. (2017). Environmental disclosure by Indian companies: an empirical study. *Environment, Development and Sustainability*, 19(5), 1999-2022.doi: 10.1007/s10668-016-9840-5.

Pucheta-Martínez, M. C., Bel-Oms, I., & Olcina-Sempere, G. (2019). Commitment of independent and institutional women directors to corporate social responsibility reporting. *Business Ethics: A European Review*, 28(3), 290-304.doi: 10.1111/beer.12218.

Pucheta-Martínez, M. C., & Gallego-Álvarez, I. (2020). Corporate environmental disclosure practices in different national contexts: The influence of cultural dimensions. *Organization & Environment*, 33(4), 597-623.doi: 10.1177/1086026619860263.

Rahman, H.-U.-, Ibrahim, M. Y. and Ahmad, A. C. (2017) 'Corporate governance reforms and shareholders' confidence in emerging markets', *World Journal of Science, Technology and Sustainable Development*. 14(1),60-74. doi: 10.1108/wjstsd-09-2016-0054.

Rahman, H. U., Ibrahim, M. Y., & Ahmad, A. C. (2017). Physical characteristics of the chief executive officer and firm accounting and market-based performance. *Asian Journal of Accounting and Governance*, 8, 27-37.doi: 10.17576/ajag-2017-08-03.

Rao, K., & Tilt, C. (2016). Board composition and corporate social responsibility: The role of diversity, gender, strategy and decision making. *Journal of Business Ethics*, 138(2), 327-347.doi: 10.1007/s10551-015-2613-5.

Reguera-Alvarado, N., De Fuentes, P., & Laffarga, J. (2017). Does board gender diversity influence financial performance? Evidence from Spain. *Journal of Business Ethics*, 141(2), 337-350.doi: 10.1007/s10551-015-2735-9.

Robinson, G., & Dechant, K. (1997). Building a business case for diversity. *Academy of Management Perspectives*, 11(3), 21-31. doi: 10.5465/ame.1997.9709231661.

Sanan, N. K. (2016). Board gender diversity, financial and social performance of Indian firms. *Vision*, 20(4), 361-367.doi: 10.1177/0972262916673006.

Sial, M. S., Zheng, C., Cherian, J., Gulzar, M. A., Thu, P. A., Khan, T., & Khuong, N. V. (2018). Does corporate social responsibility mediate the relation between boardroom gender diversity and firm performance of Chinese listed companies?. *Sustainability*, 10(10), 3591.doi: 10.3390/su10103591.

Singh, A. K., Singhania, S. and Sardana, V. (2019) 'Do Women on Boards affect Firm's Financial Performance? Evidence from Indian IPO Firms', *Australasian Accounting, Business and Finance Journal*, 13(2), pp. 53–68. doi: 10.14453/aabfj.v13i2.4.

Smith, N., Smith, V. and Verner, M. (2006) 'Do women in top management affect firm performance? A panel study of 2,500 Danish firms', *International Journal of Productivity and Performance Management*. 55(7), 569-593. doi: 10.1108/17410400610702160.

Torchia, M. *et al.* (2010) 'Critical mass theory and women directors' contribution to board strategic tasks', *Corporate Board: Role, Duties and Composition*.6(3),42-51. doi: 10.22495/cbv6i3art4.

Torchia, M., Calabrò, A., & Huse, M. (2011). Women directors on corporate boards: From

tokenism to critical mass. *Journal of Business Ethics*, 102(2), 299-317.doi: 10.1007/s10551-011-0815-z.

Tourigny, L., Han, J. and Baba, V.V. (2017). Does gender matter? A study of trust and its outcomes in the manufacturing sector in mainland China. *Gender in Management*, 32(8), 554-577. https://doi.org/10.1108/GM-05-2016-0106

Valls Martínez, M. del C., Cruz Rambaud, S. and Parra Oller, I. M. (2019) 'Gender policies on board of directors and sustainable development', *Corporate Social Responsibility and Environmental Management*, 26(6), pp. 1539–1553. doi: 10.1002/csr.1825.

Van Knippenberg, D., De Dreu, C. K., & Homan, A. C. (2004). Work group diversity and group performance: an integrative model and research agenda. *Journal of Applied Psychology*, 89(6), 1008.doi: 10.1037/0021-9010.89.6.1008.

Verma, A., & Kumar, C. V. (2014). An analysis of CSR expenditure by Indian companies. *Indian Journal of Corporate Governance*, 7(2), 82-94.doi: 10.1177/0974686220140201.

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance–financial performance link. *Strategic Management Journal*, 18(4), 303-319. doi: 10.1002/(SICI)1097-0266(199704)18:4<303::AID-SMJ869>3.0.CO;2-G.

Wang, J., & Coffey, B. S. (1992). Board composition and corporate philanthropy. *Journal of Business Ethics*, 11(10), 771-778. doi: 10.1007/BF00872309.

Hermalin, B. E., & Weisbach, M. S. (1998). Endogenously Chosen Boards of Directors and Their Monitoring of the CEO. *The American Economic Review*, 88(1), 96-118. doi: 10.1017/CBO9781107415324.004.

Zahid, M., Rahman, H, Ali, W., Khan, M., Alharthi, M., Qureshi, M., & Jan, A. (2020) 'Boardroom gender diversity: Implications for corporate sustainability disclosures in Malaysia', *Journal of Cleaner Production*. 244. doi: 10.1016/j.jclepro.2019.118683.

Zaichkowsky, J. L. (2014). Women in the board room: One can make a difference. *International Journal of Business Governance and Ethics*, 9(1), 91-113. doi: 10.1504/IJBGE.2014.062774.

Effect of asymmetries of stiffness and mass on parametric instabilities of hollow rotor system

DOI: 10.36909/jer.ICCEMME.15745

Sandeep Chauhan*, Chandan Kumar**, Ashish Gupta***, Vikas Rastogi****

*Assistant Professor, Department of Mechanical Engineering, G L Bajaj Institute of Technology and Management, Greater Noida-201306, India

*Research Scholar, Dr. A.P.J. Abdul Kalam Technical University, Uttar Pradesh, Lucknow, India

**Associate Professor, Department of Mechanical Engineering, BIT Sindari, 828123, India

***Associate Professor, Department of Mechanical Engineering, G L Bajaj Institute of Technology and Management, Greater Noida-201306, India

**** Professor, Design Engineering Centre, Department of Mechanical Engineering, Delhi Technological University, Delhi-110042, India.

*Email: ashish.j001@gmail.com; Corresponding Author.

ABSTRACT

The present work aims to understand the influence of the asymmetries on the dynamics of the hollow rotor system. The study comprises the analytical study through the extended Lagrangian Hamiltonian approach and validation conducted with the simulation study through the bond graphs. This study involves an investigation of the dynamics of the hollow rotor system through two case studies. First case study considers a hollow rotor system with small asymmetry in stiffness while another study considers a hollow rotor system with asymmetry in mass. The lumped analysis involves the development of a mathematical model by considering the symmetry breaking of a hollow rotor caused by mass or stiffness. The study shows good agreement in simulation and analytical results.

Key words: hollow rotor, asymmetric rotor, extended Lagrangian Hamiltonian mechanics, umbra time.

INTRODUCTION

Most of the rotary machines used for engineering applications are asymmetric due to misalignment or unbalance. Sometimes, symmetric systems become asymmetric during service due to initiation and growth of fatigue crack, slit, notches, wear and tear, etc. The presence of a crack causes variations in stiffness, mass, and damping. These variations make the system asymmetric and alter the dynamic behavior of the rotor. Analysis of such a system is necessary for the accurate prediction of the presence of these asymmetries, which are dangerous for the structural health of the rotor system. The cross-sectional geometry also affects the dynamics of the rotor due to alteration in the stiffness of the rotor system. Mostly solid cylindrical shaped rotor used due to ease in construction and analysis. However, the attention of researchers shifts towards a hollow rotor-shaft system owing to a lesser weight of shafts for transmitting the same power and consequently lesser requirement of energy for operation. Asymmetries greatly affected the performance of the hollow rotor system due to lesser stiffness. Thus, it becomes necessary to analyze the effects of asymmetries on the dynamics of the hollow rotor.

Dimarogonas (1996) presented a paper to elaborate on several aspects of local flexibility associated with the asymmetries in a rotating system. While Kumar et al. (2009) elaborated various methodologies employed by researchers to understand the effect of asymmetries due to crack on the dynamics of a rotor system. Generally, the Lagrangian-based approach is preferred for analyzing dynamics of the rotating system due to its inherent characteristics of describing conservation laws associated with symmetry properties. Some studies have been initiated to enlarge the applicability of the Lagrange equation to a rotating system subjected to holonomic constraints and dissipative forces. Vujanovic and Djukic (1978, 1989, 1973, 1975) attempted to show the connection between symmetries and associated conservation law. Vujanovic (1978, 1989) proposed a new approach to obtain

constants of motion of a rotary system with non-potential forces with non-holonomic constraints.

As any rotating component with asymmetries consists of non-potential and dissipative forces and in such a situation, Lagrangian-based approaches failed to explain satisfactorily dynamics of an asymmetric rotary system involving non-holonomic constraints, frictional forces, and time fluctuating parameters. Analysis of these systems requires knowledge of invariants of motion of the system due to the existence of a relation between symmetries and invariant of motions (1977, 1997) of such systems. Noether's theorem (1971) complemented the knowledge and contributed significantly to determine invariance properties. Extended Noether theorem together with Hamiltonian mechanics was found useful in understanding the dynamic behavior of such an asymmetric rotor system. For improving understanding in this direction, Mukherjee (1994) introduced the umbra Lagrangian-Hamiltonian mechanics approach by utilizing virtual time as umbra time. The approach involved the development of an extended form of the equation of motions as umbra-Lagrange's equation. Some other studies by Mukherjee et al. (2001, 2006, and 2011) further illustrate concepts of umbra Lagrangian. The approach was applied to analyze thermo-mechanical systems (Rastogi et al., 2011), electro-mechanical system (Mukherjee et al., 2007) and a one-dimensional continuous system (Mukherjee et al., 2009). Kumar et al. (2012) extend the approach to study the dynamics of a solid rotor system with asymmetry in stiffness. Another study by Jain et al. (2018) utilized the approach to understand the dynamic behavior of multi-rotor-system where symmetries are broken in terms of the rotor stiffness.

Different researcher has conducted some studies to analyze the dynamics of the hollow rotor system. Among them, a major contribution comes from Chen et al (2010) who utilized the concept of local equivalent Young’s modulus for understanding the dynamic behavior of a

motor built-in high speed hollow rotating shaft through experimental and numerical study. Some other studies by Chen et al. (2011, 2012, and 2013) enhanced understanding of the influence of the friction on the natural frequency of the hot fit rotor-hollow shaft assembly. Chen and Yang (2010) also performed an experimental study of a hollow shaft with a hot-fitted rotor to determine the optimal value of interference for dynamical stiffness of a shaft fitted with a hot-fit rotor. In another study, Asdaque and Behera (2014) analyzed the dynamic behavior of hollow tapered shaft-rotor to investigate the effects of length and speed on the dynamics of the hollow tapered shaft-rotor system. Further, Guo et al. (2014) theoretically studied an NES-rotor system by developing motion equations for NES rotor as well as TMD rotor dynamical systems by applying Lagrange's equations and determined amplitude of vibration at resonance condition. One study by Cavalini et al. (2017) established the suitability of the composite hollow shaft as a rotor while other study by Y. Yongfeng et al. studied dynamic behavior of hollow-cracked shaft.

The present work is initiated to develop an extension of Lagrangian Hamiltonian mechanics for hollow rotor systems with asymmetries in stiffness and mass by utilizing the approach introduced by Mukherjee et al. (2009, and 2011). The methodology adopted for analysis is presented in sufficient detail for two case studies in subsequent sections.

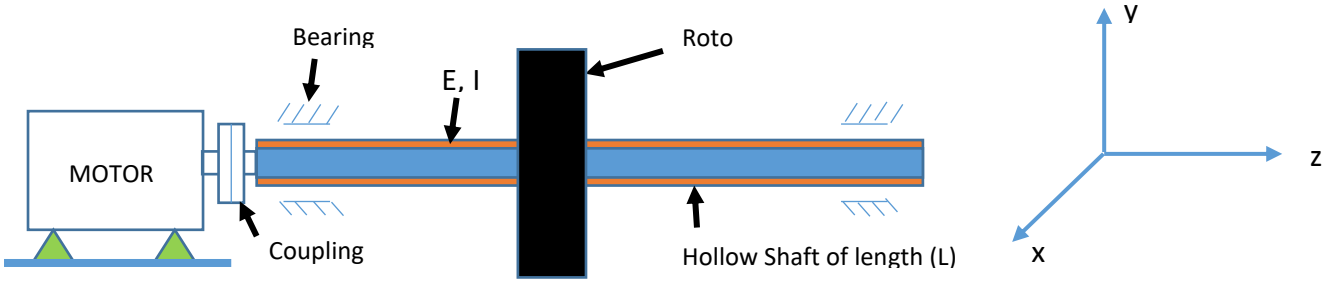


Figure 1 Rotor with DC motor having asymmetries in stiffness

CASE STUDY I: INSTABILITIES DUE TO ASYMMETRY IN STIFFNESS

This case study considers a rotor with internal and external damping and transverse stiffness due to flexure of the shaft driven by a DC motor as shown in Figure 1. The rotor model admits small asymmetries in stiffness. The work represents the rotor by two discrete pairs of equivalent

springs and dashpots in the two orthogonal transverse directions. Such dampers include external damping and a part of the internal damping. The study considers the additional circulatory forces due to the internal damping by using modulated gyrator elements in the bond graph model. The analysis neglects shaft twist due to high torsional stiffness assumption.

The umbra-Lagrangian is derived from bond graph model of rotor system by adopting the method developed by Mukherjee (1997, 2001, 2011) and is expressed as

$$L^* = \frac{1}{2}m\dot{x}^2(\eta) + \frac{1}{2}m\dot{y}^2(\eta) - \frac{1}{2}K\dot{x}^2(\eta) - \frac{1}{2}(K-\Delta K)\dot{y}^2(\eta) - \frac{1}{2}J\dot{\theta}^2(\eta) - \frac{1}{2}R_i\dot{\theta}(t)\{x(\eta)y(t) - x(t)y(\eta)\} - (R_a + R_i)\big(x(\eta)\dot{x}(t) + y(\eta)\dot{y}(t)\big) - \{R_i\big(x(t)\dot{y}(t) - y(t)\dot{x}(t)\big) - R_b\left(\dot{\theta}(t)\right) - \Omega\}\theta(\eta) \quad (1)$$

Here the symbol represented the standard quantity. Now by applying the extended Noether's equation, Noether's rate equation for asymmetric rotor is written as

$$\frac{\mathrm{d}}{\mathrm{dt}}\left[\mathrm{m}\big(\dot{\mathrm{x}}(\mathrm{t})\mathrm{y}(\mathrm{t}) - \dot{\mathrm{y}}(\mathrm{t})\mathrm{x}(\mathrm{t})\big) - \Delta\mathrm{K}\mathrm{y}(\mathrm{t})\mathrm{x}(\mathrm{t})\right] - \dot{\theta}(\mathrm{t})\mathrm{R}_\mathrm{i}\big(\mathrm{x}^2(\mathrm{t}) + \mathrm{y}^2(\mathrm{t})\big) + (\mathrm{R}_\mathrm{a} + \mathrm{R}_\mathrm{i})\big(\mathrm{x}(\mathrm{t})\dot{\mathrm{y}}(\mathrm{t}) - \dot{\mathrm{x}}(\mathrm{t})\mathrm{y}(\mathrm{t})\big) = 0 \quad (2)$$

On assuming an orbit of $x(t) = A\cos\omega t \text{ and } y(t) = A\sin\omega t$, we get

$$\dot{\theta}(t) = \omega(1 + \frac{R_a}{R_i}) \quad (3)$$

Thus, Noether's rate equation becomes

$$\frac{d}{dt}\left[m\big(\dot{x}(t)y(t) - \dot{y}(t)x(t)\big) - \Delta K y(t)x(t)\right] = 0 \quad (4)$$

The above equation consists of symmetric and un-symmetric part. On equating un-symmetric part to zero, one may obtain invariants of motions as

$$\frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{m}(\dot{\mathrm{x}}(\mathrm{t})\mathrm{y}(\mathrm{t}) - \dot{\mathrm{y}}(\mathrm{t})\mathrm{x}(\mathrm{t})] = 0 \quad (5)$$

Now, the umbra-Hamiltonian of the system is represented as

$$\mathrm{H}^* = \frac{\mathrm{p}_\mathrm{x}^2(\eta)}{2\mathrm{m}_{\mathrm{xx}}} + \frac{\mathrm{p}_\mathrm{y}^2(\eta)}{2\mathrm{m}_{\mathrm{yy}}} + \frac{1}{2}\mathrm{K}\mathrm{x}^2(\eta) + \frac{1}{2}(\mathrm{K}-\Delta\mathrm{K})\mathrm{y}^2(\eta) + \frac{1}{2\mathrm{J}}\mathrm{p}_\theta^2(\eta) - \mathrm{R}_\mathrm{i}\dot{\theta}(\mathrm{t})\{\mathrm{x}(\eta)\mathrm{y}(\mathrm{t}) - \mathrm{x}(\mathrm{t})\mathrm{y}(\eta)\} + (\mathrm{R}_\mathrm{a} + \mathrm{R}_\mathrm{i})\big(\dot{\mathrm{x}}(\mathrm{t})\mathrm{x}(\eta) + \dot{\mathrm{y}}(\mathrm{t})\mathrm{y}(\eta)\big) + \{\mathrm{R}_\mathrm{i}\big(\mathrm{x}(\mathrm{t})\dot{\mathrm{y}}(\mathrm{t}) - \mathrm{y}(\mathrm{t})\dot{\mathrm{x}}(\mathrm{t})\big) + \mathrm{R}_\mathrm{b}\left(\dot{\theta}(\mathrm{t})\right) - \Omega\}\theta(\eta) \quad (6)$$

On applying the second theorem of umbra-Hamiltonian, $\lim_{\eta\to t}\left[\frac{dH_e^*}{d\eta}\right] = 0$, we got

$$R_i\dot{\theta}(t)[\dot{x}(\eta)y(t) - \dot{y}(t)x(t)] + (R_a + R_i)\left(\dot{x}^2(t) + \dot{y}^2(t)\right) + \left\{R_i\left(x(t)\dot{y}(t) - y(t)\dot{x}(t)\right) + R_b\left(\dot{\theta}(t) - \Omega\right)\right\}\dot{\theta}(t) = 0 \quad (7)$$

Assuming an orbit $x(t) = A\cos\omega t$, and $y(t) = A\sin\omega t$, we get

$$-R_i\dot{\theta}(t)A^2\omega + (R_a + R_i)A^2\omega^2 + \left\{R_iA^2\omega + R_b\left(\dot{\theta}(t) - \Omega\right)\right\}\dot{\theta}(t) = 0 \quad (8)$$

A B

On putting part A of equation (8) equal to zero yields

$$\dot{\theta}(t) = \omega\left(1 + \frac{R_a}{R_i}\right) \quad (9)$$

and on putting B part of equation (8) equal to zero gives the amplitude of rotor

$$A = \sqrt{\frac{R_b}{R_i\omega}\left[\Omega - \omega\left(1 + \frac{R_a}{R_i}\right)\right]} \quad (10)$$

When we apply $\lim_{\eta \to t}\left[\frac{dH_i^*}{d\eta}\right] = 0$, we got the following expression

$$\frac{1}{2}\left\{\left(\frac{p_x^2(\eta)}{m_{xx}} + \frac{p_y^2(\eta)}{m_{yy}}\right) + Kx^2(\eta) + (K - \Delta K)y^2(\eta) + \frac{1}{J}p_\theta^2(\eta)\right\} = 0 \quad (11)$$

Now putting $p_x(t) = m_{xx}\dot{x}(t), p_y(t) = m_{yy}\dot{y}(t)$ and

$p_\theta(t) = J\dot{\theta}(t)$ in Eq. 11, following equation is obtained

$$(m_{xx}\ddot{x} + Kx)\dot{x} + \left(m_{yy}\ddot{y} + Ky - \Delta Ky\right)\dot{y} = 0 \quad (12)$$

There are two types of frequencies available in Eq. (12). If symmetry part is considered, one may have

$$\omega_{xx} = \sqrt{\frac{K}{m}} \quad (13)$$

If un-symmetric part is considered than we have

$$\omega_{yy} = \sqrt{\frac{K - \Delta K}{m}} \quad (14)$$

Since these two vectors are perpendicular to each other, we achieve the final natural frequency as

$$\omega^2 = \frac{\omega_{xx}^2 + \omega_{yy}^2}{2} \quad \text{or} \quad \omega = \sqrt{\frac{K - \frac{\Delta K}{2}}{m}} \quad (15)$$

CASE STUDY-II INSTABILITIES DUE TO ASYMMETRY IN MASS

This case study aims to understand the behavior of the rotor system where the symmetries are broken in terms of variation in mass. The study considered the same rotor system used in the previous case along with small asymmetries in mass in the Y direction instead of asymmetry in stiffness. All assumptions made in the previous case study also applicable in the present case.

The umbra-Lagrangian function for the system is expressed as

$$L^* = \frac{1}{2}m\dot{x}^2(\eta) + \frac{1}{2}(m+\Delta m)\dot{y}^2(\eta) - \frac{1}{2}K\dot{x}^2(\eta) - \frac{1}{2}K\dot{y}^2(\eta) - \frac{1}{2}J\dot{\theta}^2(\eta) - \frac{1}{2}R_i\dot{\theta}(t)\{x(\eta)y(t) - x(t)y(\eta)\} - (R_a + R_i)\big(x(\eta)\dot{x}(t) + y(\eta)\dot{y}(t)\big) - \big\{R_i\big(x(t)\dot{y}(t) - y(t)\dot{x}(t)\big) - R_b\left(\dot{\theta}(t)\right) - \Omega\big\}\theta(\eta) \quad (16)$$

By following the procedure as explained in previous case study, the relationship to determine threshold frequency and the amplitude of rotor system are derived as given below:

$$\dot{\theta}(t) = \omega\left(1 + \frac{R_a}{R_i}\right) \quad (17)$$

$$A = \sqrt{\frac{R_b}{R_i\omega}\left[\Omega - \omega\left(1 + \frac{R_a}{R_i}\right)\right]} \quad (18)$$

Due to asymmetry of mass along y axis, system will have two types of frequencies. One frequency is obtained by considering symmetry part and given below:

$$\omega_x = \sqrt{\frac{K}{m}} \quad (19)$$

While other frequency is obtained by considering asymmetric part as given below:

$$\omega_y = \sqrt{\frac{K}{m\pm\Delta m}} \quad (20)$$

The final natural frequency can be expressed as

$$\omega^2 = \frac{\omega_{xx}^2 + \omega_{yy}^2}{2} \quad \text{or} \quad \omega = \sqrt{\frac{2K}{m\pm\Delta m}} \quad (21)$$

RESULTS AND DISCUSSION

This section presents the results obtained from the analysis of hollow rotor with asymmetries in mass and stiffness by utilizing an extended Lagrangian Hamiltonian mechanics approach. Simulation analysis of the hollow rotor with different asymmetries are performed by adopting bond graph technique for validating the analytical approach. Next sections include the detailed discussion on analytical and simulation results.

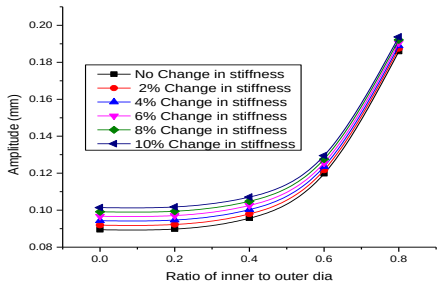
Analytical Results

Analytical results are obtained by adopting the approach discussed in previous section for both cases and discussed separately in subsequent sections. First section presents discussion of results pertaining to hollow rotor with consideration of asymmetry in stiffness while other section discussed the results pertaining to a hollow rotor system with asymmetry in mass.

Asymmetry in stiffness

By utilizing the equations developed previously, natural frequency and amplitude of vibration induced in the hollow shaft rotor system are obtained and presented graphically in figure (2-4). The study considers the hollowness ratio as the ratio of the internal diameter to the external diameter. The study calculates these results by assuming excitation frequency as 22 rad/s. The careful analysis of these results reveals an increasing trend in the amplitude of the vibration along with a decreasing trend in the natural frequency of the vibration for a hollow shaft rotor with different hollowness ratios. The results also show that there is no appreciable variation in the performance of the hollow rotor with hollowness ratio up to 0.4. This shows that hollow shaft with hollowness ratio value more than 0.4 is not suitable to employ for the rotor shaft. Further results are obtained by varying excitation frequencies as 26, 30, 34, and 38 rad/s, respectively and no effect observed on natural frequency as expected while increasing trend observed in amplitude of the vibration with increasing value of the excitation frequency as shown in figure 10 (b) for a hollow shaft rotors with the hollowness ratio as 0.2. Similar trends.

are observed for hollow shaft rotor with different hollowness ratios. These results are not included here to avoid repeatability.



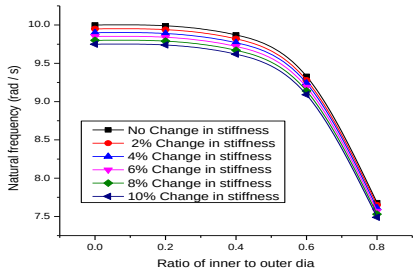
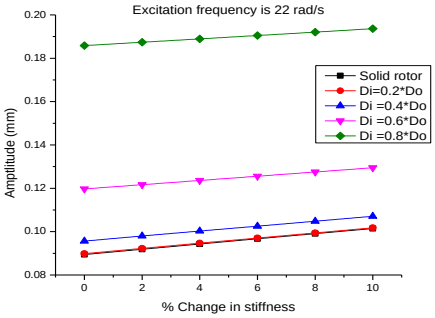


Figure 2 Effect of hollowness on amplitude **Figure 3** Effect of hollowness on natural frequency



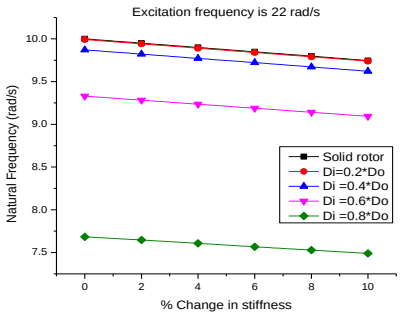


Figure 4 Amplitude Vs % Change in stiffness **Figure 5** Natural frequency Vs % Change in stiffness

Asymmetry in mass

By utilizing, the equations developed in previous section, the amplitudes and the natural frequencies of the vibration for a hollow rotor system having varying percentages of mass asymmetries along the y-axis obtained and presented in Figure 6 to Figure 9. The results depict the decreasing trend in the natural frequency of the system with increasing percentage of asymmetries of mass along the y-axis for hollow rotor with different hollowness ratios while reverse trend observed in the amplitudes of the vibrations with an increasing percentage of mass asymmetry. The results also show that there is little variation in the performance of a hollow shaft in comparison with a solid shaft if the hollowness ratio of the hollow shaft rotor is kept

below a certain limit (i.e. 0.4). The results also obtained by varying the excitation frequency and observed no affect in the excitation frequency while amplitudes increases with an increase in excitation frequency as shown by the figure 10 (a).

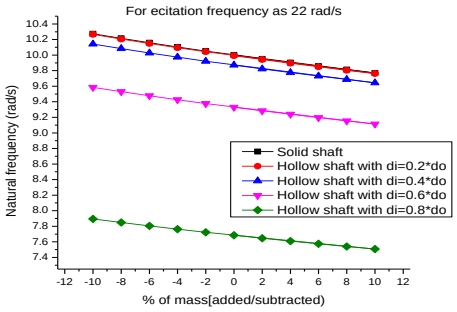


Figure 6 Frequency Vs % Change in mass

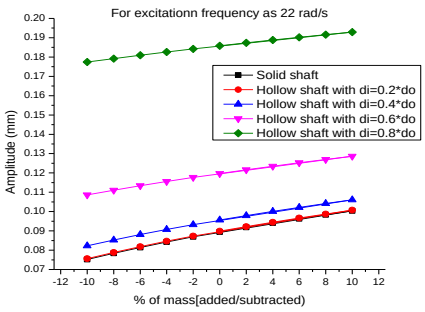


Figure 7 Amplitude Vs % Change in mass

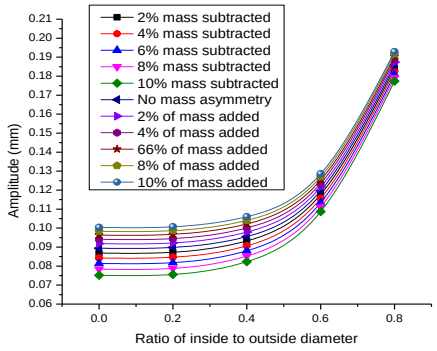


Figure 8 Effect of hollowness on amplitude

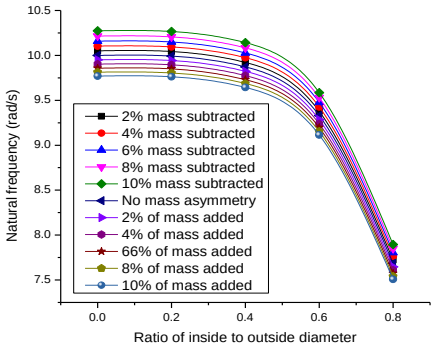
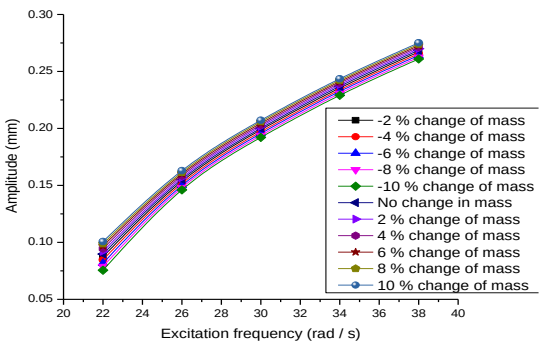
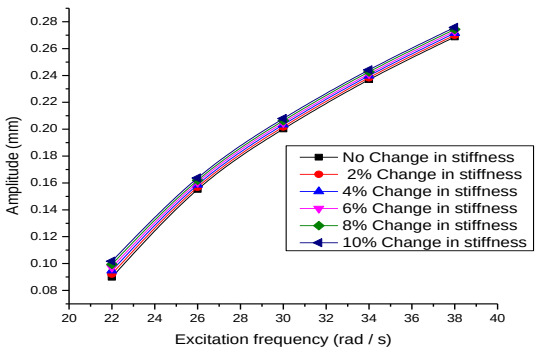


Figure 9 Effect of hollowness on natural frequency



(a) Hollow rotor with asymmetry in mass



(b) Hollow rotor with asymmetry in stiffness

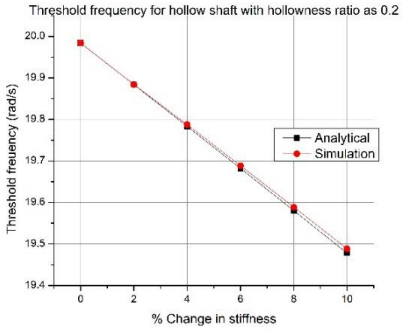
Figure 10 Effect of Excitation frequency on amplitude of hollo rotor when $D_i = 0.2*D_o$

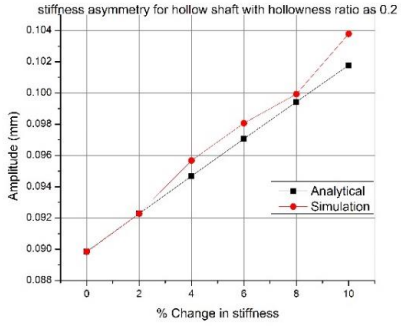
Simulation Results

The main purpose of this study is to gain an insight into dynamical behavior. Moreover, it provides validation of the theoretical results obtained in the previous section. The bond graph model of an asymmetric rotor with internal and external damping was developed and simulated through the software SYMBOLS-Shakti. The study performs a series of experimentation to show the effectiveness of the analysis described above. The work includes the simulation study performed for both the cases discussed in the previous section and presented in succeeding sub-sections. For simulation analysis of the hollow rotor with small asymmetry in stiffness and mass, the present work considers different input parameters as shown in Table-1. Amplitudes and threshold frequencies of the vibratory behavior of the hollow rotor with asymmetries are determined through the bond graph technique and presented graphically along with the corresponding analytical results through figures 14 to 17. Figures 14 and 15 presents the results for a hollow rotor with asymmetry in stiffness while figures 16 and 17 represent the corresponding results for a hollow rotor with asymmetry in mass. The threshold frequency of the hollow rotor system is found to be twice of the natural frequency of the hollow rotor system. The simulation results are in harmony with the analytical results and thus proves the utility of the extended Lagrangian Hamiltonian approach for predicting the dynamic behavior of hollow rotor with small asymmetries.

Table 1. Simulation parameters for hollow rotor with asymmetry in stiffness

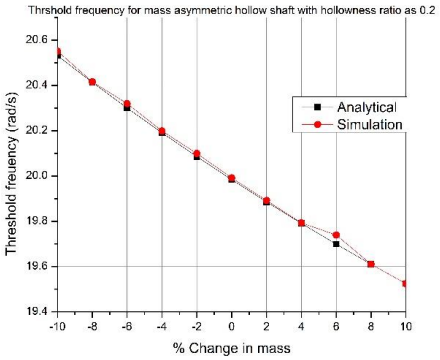
Parameter	K (N/m)	m (Kg)	J (Kg-m^2)	R_i (N-s/m)	R_a (N-s/m)	R_b (N-s/m)	Ω (rad/s)
Value	1000	10	1	5	5	0.2	22

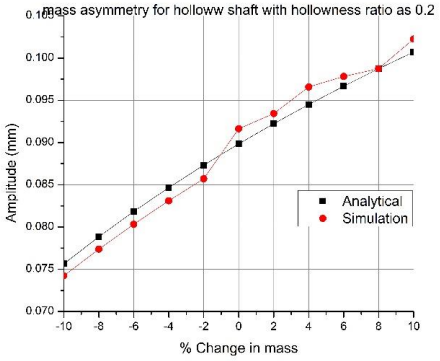




(a) Effect on threshold frequency **(b)** Effect on amplitude

Figure 11 Effect of stiffness asymmetry on hollow rotor system





(a) Effect on threshold frequency **(b)** Effect on amplitude

Figure 12 Effect of stiffness asymmetry on hollow rotor system

CONCLUSION

This work presents an analytical framework for understanding the changes in the dynamic behavior of the hollow rotor system caused by the asymmetries in terms of stiffness and mass of the shaft rotor. Following conclusion has been drawn from the study:

- The study develops expressions for determining the natural frequency and the amplitude of vibration of a hollow shaft rotor through Extended Lagrangian Hamiltonian formalism.
- The study obtains the umbra-Lagrangian function from its bond graph model, which remains invariant and leads to significant aspects of limiting dynamics when used with umbra Hamiltonian. The results exhibits marginal difference in the

behavior of a hollow rotor with a hollowness ratio less than 0.4 in comparison to a solid rotor. This fact justify the suitability of a hollow rotor for industrial purpose.

- The results of simulation study are in harmony with the analytical results and thus validate the limiting behavior obtained through present study.
- Harmony of simulation results with analytical results justify the use of an extended Lagrangian Hamiltonian approach for predicting the dynamic behavior of hollow rotor with small asymmetries.
- Further present study can be extended to incorporate more asymmetries in the hollow rotor like as transverse crack, circular crack and multi rotor system.

REFERENCES

Dimarogonas, 1996. Vibration of cracked structure: a state of the art review, Eng. Fract. Mech. 55 831-857.

C. Kumar, V. Rastogi, 2009. A brief review on dynamics of a cracked rotor, Int. J. Rot. Mach. 1-6, https://doi.org/10.1155/2009/758108. Article ID 758108.

B. Vujanovic, 1978. Conservation laws of dynamical systems via d'Alembert's principle, Int. J. Nonlinear Mech. 13 185-197.

B. Vujanovic, S.E. Jones, 1989. Variational Methods in Nonconservative Phenomena, Academic press, Boston

D. S. Djukic, 1973. A procedure for finding first integrals of mechanical systems with gauge-variant Lagrangians, Int. J. Nonlinear Mech. 8 479-488**.**

D. S. Djukic, B. Vujanovic, 1975. Noether's theory in classical nonconservative mechanics, Acta Mech. 23 17-27.

D. C. Karnopp, 1977. Lagrange's equations for complex bond graph systems, J. Dyn. Sys. Meas. Control 99 300-306.

A. Mukherjee, A.K. Samantaray, 1997. Umbra-Lagrange's equations through bond graphs, in: Proc. of ICBGM - International Conference on Bond Graph Modeling and Simulation, Phoenix, Arizona, USA, pp. 168-174.

E. Noether, 1971 Invariant variation problems, Transp. Theory Stat. Phys. 1 186-207.

A. Mukherjee, 1994. Junction structures of bond graph theory from analytical mechanics viewpoint, in: Proc. of CISS-1st Conference of International Simulation Societies. Zurich, Switzerland, pp. 661-666.

A. Mukherjee, V. Rastogi, 2006. A. Das Gupta, A methodology for finding invariants of motion for asymmetric systems with gauge-transformed Umbra lagrangian generated by bond graphs, Simulation 82 207-226.

A. Mukherjee, V. Rastogi, A. Dasgupta, 2011. Revisiting Umbra-Lagrangian-Hamiltonian mechanics: its variational foundation and extension of Noether's theorem and poincare-Cartan Integral, Int. J. Nonlinear Mech. 46 745-757.

V. Rastogi, A. Mukherjee, 2011. Conservation laws for gauge-variant Umbra-Lagrangian in Classical mechanics using Bond graphs, Simulation 87 301-312.

A. Mukherjee, V. Rastogi, A. Dasgupta, 2009. Extension of Lagrangian-Hamiltonian mechanics for continuous systems e investigation of dynamics of a one dimensional internally damped rotor driven through a dissipative coupling, Nonlinear Dyn. 58 107-127.

C. Kumar and Vikas Rastogi, 2012. Effects of asymmetric stiffness on parametric instabilities of rotor, Journal of computation and applied research in mechanical engineering, Vol 1 No 2, 119-128.

Anuj Kumar Jain , Vikas Rastogi , Atul Kumar Agrawal, 2018. A novel approach to study effects of asymmetric stiffness on parametric instabilities of multi-rotor-system, Journal of Sound and Vibration 413 159-172.

S. Y. Chen, C. Kung, T. T. Liao, & Y. H. Chen, 2010. Dynamic effects of the interference

fit of motor rotor on the stiffness of a high-speed rotating shaft, Transactions of the Canadian Society for Mechanical Engineering, 34(2), 243–261.

S. Y. Chen, C. Kung, Jung-Chun Hsu, 2011. Dynamic Analysis Of A Rotary Hollow Shaft With Hot-Fit Part Using Contact Elements With Friction, Transactions of the Canadian Society for Mechanical Engineering, Vol. 35, No. 3.

Shin-Yong Chen, 2013. The Dynamic Analyses and Verifications of A Hollow shaft With Hot-Fit Component Using 3d Finite Contact Element, Transactions of the Canadian Society for Mechanical Engineering, Vol. 37, No. 1.

S. Y. Chen, 2012 . An equivalent direct modeling of a rotary shaft with hot-fit components using contact element modal analysis results, Computers and Mathematics with Applications, 64(5), 1093–1099.

Shin-Yong Chen, Shyi-Kae Yang, 2010. The Optimal study of Finite Element Modelling for a High Speed Hollow Shaft with a Hot-fitted Motor Rotor, Internal symposium on Computer, Communication and Automation.

P M G Bashir Asdaque, R K Behera, 2014. Vibration Analysis of Hollow Profiled Shafts", International Journal of Current Engineering and Technology, Special Issue-3, 124–128.

C Guo, MA AL-Shudeifat, AF Vakakis, LA Bergman, DM McFarland, J Yan, 2014. "A. Vibration reduction in unbalanced hollow rotor systems with nonlinear energy sinks", Nonlinear Dynamics 79 (1), 527-538

Aldemir Ap Cavalini Jr, Thiago A M Guimarães, Bruno R M G da Silva, Valder Steffen Jr 2017. "Analysis of the Dynamic Behavior of a Rotating Composite Hollow Shaft", Latin American Journal of Solids and Structures 14, 1-16.

Yongfeng Yang, Qinyu Wu,, Yanlin Wang,, Weiyang Qin, Kuan Lu, 2019 "Dynamic characteristics of cracked uncertain hollow-shaft", Mechanical Systems and Signal Processing 124, 36–48

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/355675506

Effect of critical success factors (CSFs) on the organizational performance

Article · September 2021

DOI: 10.36037/IJREI.2021.5602

CITATIONS
0

READS
51

2 authors:



Radhey Shyam Mishra
Delhi Technological University
57 PUBLICATIONS **199** CITATIONS

SEE PROFILE



Rakesh Kumar
Delhi Technological University
10 PUBLICATIONS **9** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Population Statistics View project

Project Performance Measurement of TQM in Indian FMCG Industry View project

All content following this page was uploaded by Rakesh Kumar on 27 October 2021.

The user has requested enhancement of the downloaded file.

International Journal of Research in Engineering and Innovation Vol-5, Issue-6 (2021), 341-348

International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934



ORIGNAL ARTICLE

Effect of critical success factors (CSFs) on the organizational performance

R. S. Mishra, Rakesh Kumar

Department of Mechanical Engineering, Delhi Technological University Delhi, India

Article Information

Received: 20 June 2021
Revised: 03 August 2021
Accepted: 10 Sep 2021
Available online: 15 Sep 2021

Keywords:

TQM;
Critical success factors;
Organizational performance;
Structural Equation Modeling.

Abstract

Mediation analysis aims to uncover causal pathways along which effects are transmitted from potential causes. In present case, our interest in assessing the effective performances to which ultimately organizational performance can be affected. The former concerns the "direct effect" and the "indirect effect" or the "effect mediated.
In this paper, the developed model can provide some insights about the under lying process to identify causes for certain effects. This work is initiated in response to questions and suggestions from organizations how CSFs indicator to ensure that it more efficiently measures the performance effectiveness. While the modified indicator technical specification is being finalized, we encourage organizations to use quality improvement tools to understand more about the use of unconventional spaces or flow in their company.
©2021 ijrei.com. All rights reserved

1. Introduction

Critical Success Factors (CSFs) are predictors of performance effectiveness of organizational system or subsystems. There has been widespread interest in association between system's goal and its performance in context of organization. Organizations use to apply various strategies, stratagems and tactics like Toyota Production System (TPS), Lean manufacturing, Supply Chain Management (SCM), Just-In-Time (JIT), Six-sigma etc. to excel performance, TQM is one of them, organizations use to apply holistically. The performance of the organization mainly the aggregation of its financial, non-financial and operational performance. In a combination of these, the organization gets outcome such as effectiveness, efficiency, development and participant's satisfaction. After using all supports and efforts when the organization produces a product or service that is called the organizational performance. Performance measurement can facilitate the alignment of the goals of all individuals, teams, departments and processes with the strategic aims of the organization and incorporate the voice of the stakeholders in all planning and management activities, Oakland (2003). TQM is basically a strategy (towards continuous change), as well as an operationalized process, and can be also described as a holistic approach which seeks, through the improvement of quality, productivity and competitiveness (Pfau, 1989), to integrate all organizational functions and organizational objectives in a focus on meeting customer needs (Kumar et al., 2008). Performance of a company reflects to what degree the company accomplishes the corporate strategy and goals stated Öztayşi and Kutlu (2011). The General Accounting Office (GAO) study was one of the first studies trying to establish a link between TQM practices and the performance of companies, see GAO (1991). In his study, Malcolm Baldrige recipients and companies that had received a site-visit (i.e., companies that in a sense were close to receiving an award)

Corresponding author: R.S. Mishra
Email Address: hod.mechanical.rsm@dtu.ac.in
https://doi.org/10.36037/IJREI.2021.5602

were evaluated. The main conclusion from the GAO study was that the companies investigated had improved their operating results. There are several studies which have been carried out to determine the linkages or predictive correlations between TQM elements or factors and performance measures, Barros, S. et al. (2014). Demirbag, M. et al. (2006) conducted a study to determine the critical factors of TQM and measure their effect on organizational performance of SMEs operating in the Turkish textile industry. Critical success factor (CSF) theory was originally applied in other industries and areas, including general project management, manufacturing systems, and reengineering (Holland & Light, 2003). Yusuf. et al. (2007) emphasizes that as per TQM philosophy the customer requirements and business goals are inseparable. G. Muruganantham et al. (2018) mentioned that TQM provides a set of guidelines, which help to improve the performance of organization. Although in the way of measuring the MBNQA performance Wilson & Collier (2000) states that manufacturing system influence their performance variables through its mediating variables (leadership, information and analysis, strategic planning, human resource management, process management, business results and customer focus and satisfaction), but quality model, as he considered MBNQA model, directly influences company performance. To reinforce the benefits of TQM it is also advisable by Santos et al. (2007) to facilitate comparison across studies by avoiding differing conceptualizations and TQM-related measures. Above and all endorsed by Steven E. Brigham (1993) of reports survey by A. T. Kearney, TQM: A Business Process Perspective that TQM is an integrated management strategy that uses a collection of strategies to achieve corporate goals. TQM is most effective when it is a central, planned component of an organization's forward drive, one that necessitates top-level leadership, is based on a strong commitment to customers, and stresses significant improvements in "core" processes.

The development of operational strategies in alignment to firm's competitive strategy can serve to improve and tailor the product offering for customers as well as improve the internal efficiency and effectiveness of manufacturing plants (Robson et al., 2013). Therefore, an operational strategy is a subset of a firm's competitive strategy, Sahoo, (2020). Matching the dots, it can anticipate that the TQM drivers and enablers (CSFs) will positively affect the overall performance of the organization through positively affecting TQM performance shown in the figure. The TQM CSFs will enable the organization to perform better for its targeted goal.

However, sufficient research has not been carried out in this context. Therefore, we tried tends to the answer of following research questions through this study.

RQ1: Will Human Resource Management factors affect the performance of TQM?

RQ2: Will Top management commitment factors affect the performance of TQM?

RQ3: Will Process management factors affect the performance of TQM?

RQ4: Will Customer focus/ Customer Centricityfactors affect the performance of TQM?

RQ5: Will Supplier partnership/ Supplier's management factors affect the performance of TQM?

RQ6: Will Training and education factors affect the performance of TQM?

RQ7: Will Quality Information/Information Quality factors affect the performance of TQM?

RQ8: Will Strategic quality planning factors affect the performance of TQM?

RQ9: Will Culture and communication factors affect the performance of TQM?

RQ10: Will Benchmarking factor affect the performance of TQM?

RQ11: Will Social and environmental responsibility factors affect the performance of TQM?

RQ12: Will Innovation factor affect the performance of TQM?

The study involves formulation of hypotheses related to CSFs of TQM and its performance. Hypotheses are tested using the information and responses gathered from the experts of FMCGs industry. The study intends to investigate the relation between TQM CSFs and its performance in the Indian FMCGs industry. The main construct outlines the effects of TQM CSFs in FMCGs industry. The latent variables of all constructs have reflective type of observed variables. The intent is to understand the association of effect of TQM CSFs with sub factors on performance of TQM. The measurements models of the constructs developed were tested for fitness of data for further modelling.

The remaining part of the study is organized as follow. Section 2 explores the literature related to TQM CSFs and TQM performance and explains hypothesis development. Section 3 discusses research methodology comprising of hypothesis testing along with the data collection method. Section 4 deals with results and discussion, including the measurement model and SEM model of the study.

2. Literature Review

The critically examined factors which were responsible for success to achieve the intended goal is critical success factors of that system. Marais et al. (2017) states that CSFs are those aspects that must be well managed in order to achieve success. CSFs are combinations of activities and processes which are designed to support the achievement of the goals (Brotherton & Shaw, 1996, p. 114). Furthermore, CSFs are actionable, controllable by management to a variable extent, and potentially measurable (Brotherton & Shaw, 1996, p. 114). Walsh et al. (2002) states that a link existed between the source of the TQM initiative and driving force behind the TQM initiative in many organizations. Many organizations' TQM efforts originated in the quality department and also driven by the need for improved quality. Dixon et al. (1990) introduce two concepts, first, the link between strategies, actions and measures; and second, the acceptance of changing

performance measures. Performance measurement provides the feedback required to control and improve actions, which are themselves taken as a result of decision taken on strategies the organization is to follow, stated Sinclair and Zairi (2000). Odiorne (1987) states that the things for which we can devise indicators can be managed and the things for which we have no indicator can be out of control before realizing it. Performance measures derived from organization strategy with the purpose to implement the strategy, evaluate business performance, provide feedback and ensure communication, help in creating learning environment and continuously improving the organization. Zairi (1994) identifies that performance measurement has been the systematic assignment of number of activities. He further suggested that the function of measurement is to develop a method for generating a class of information that will be useful in a wide variety of problems and situations. Wilson & Collier (2000) states that manufacturing system influence their performance variables through its mediating variables (leadership, information and analysis, strategic planning, human resource management, process management, business results and customer focus and satisfaction) , but quality model, as he considered MBNQA model, directly influences company performance.

The performance solely depends on critical success factors is challengeable, for this some justification required, which need hypothesis development. The postulates were developed by the researchers for the estimation in this context which were under consideration for the study. The developed hypothesis needs further testing for whether that fits or unfits for the considered study, then acceptance or rejection of that hypothesis is decided. The testing of hypothesis is fundamental in statistics, and it could be considered as a "method" of making statistical decisions using experimental data.

The hypotheses are developed to test whether the TQM CSFs are positively related with financial, non-financial and operational performance of TQM. The responses were gathered through the questionnaire developed, based on the theoretical background of hypothesis. The hypotheses developed are as follows.

H_1: Human Resource Management factors (a) employee involvement (b) empowerment (c) recognition and reward (d) teamwork are positively related with Performance of TQM

H_2: Top management commitment factors (a) Top management support(b) Executive commitment (c) Leadership are positively related with Performance of TQM

H_3: Process management factors (a) Tools and techniques (b) Continuous improvement (c) Process design are positively related with Performance of TQM

H_4: Customer focus/ Customer Centricity factors (a) Customer and market focus (b) Customer satisfaction (c) Customer relationship are positively related with Performance of TQM

H_5: Supplier partnership/ Supplier's management factors (a) Cooperation with suppliers (b) Supplier quality management (c) Supplier relationship are positively related with Performance of TQM

H_6: Training and education factors (a) Learning (b) Knowledge and (c) Education & training are positively related with performance of TQM

H_7: Quality Information/Information Quality factors (a) Quality data and reporting (b)Internal quality information usage are positively related with Performance of TQM

H_8: Strategic quality planning factors (a) Quality policy (b) Quality planning (c) Vision &Plan statement are positively related with Performance of TQM

H_9: Culture and communication factors (a) Trust (b) Cultural change are positively related with Performance of TQM

H_{10}: Benchmarking factor (a) Competitors is positively related with Performance of TQM

H_{11}: Social and environmental responsibility factors (a) Wider community (b) Quality citizenship are positively related with performance of TQM

H_{12}: Innovation factor (a) Product innovation is positively related with Performance of TQM

3. Material & Methods

The present study attempts to (a) model a complex structure of TQM system and (b) Link the TQM CSFs with its performance for validly support the hypotheses of interest.

Questionnaire is designed for research purpose to conduct a survey and collect data for studying the effect of TQM CSFs on TQM performance by using AMOS 22. Most of the questions are adapted from peer reviewed works of Nguyen et al. (2016), Sadikoglu and Zehir (2010), Lakhal et al. (2006), Kaynak (2003), Saraph et al. (1989). The questionnaire sent to 500 quality experts, managers of quality departments of India based FMCGs industries to collect data from companies using TQM. A total of 395 respondents came back; of 395 responses, 30 responses are removed because of incomplete answers. As a consequence, the sample size of this research is 365, with the rate of response is 73%. Primary data (quantitative) was collected through a questionnaire comprising of structured questions and secondary data was collected from existing sources such as books, articles, journals, reports, and websites. Secondary data was found quick, easily accessible and inexpensive way of collecting data to better define the problem. Reliability analysis for the questionnaire as a whole, it is concluded that the questionnaire of the whole *Cronbach's* α value is 0.792, close to 0.8, so that the questionnaire has good reliability. In general, if the $\alpha > 0.9$, the questionnaire reliability is very good, if $0.8<\alpha< 0.9$, the questionnaire reliability is good, it is generally believed the questionnaire reliability is greater than 0.5 is reasonable. At the same time, the *Cronbach's* magnitude of each influencing factor is greater than 0.6, indicating that the reliability of each influencing factor is also better and credible.

The selection of variables considered its Magnitude, Articulation, Generality, Interestingness and credibility (MAGIC criteria suggested by Robert P. Abelson (1995)). The variables are transformed only for substantive reasons, not

statistical ones. In the view of Robert P. Abelson (1995), magnitude of any experimental effect(s) observed. Differences which are highly statistically significant are more persuasive than ones which are less significant or not at all significant. Articulation refers to the level of detail at which scientific results are presented. Results expressed in quantitative terms typically have more impact than those expressed only qualitatively, and, of these, more specific claims have greater impact than less specific. The generality of effects observed, by which appears to mean the extent to which observed effects are replicable and generalizable beyond the particular experimental context in which they were observed. Next is the interestingness of claims, which is a combination of their significance for current theories and their degree of surprisingness. The more general and the more interesting is a claim, the greater, typically, is its impact on an academic community.

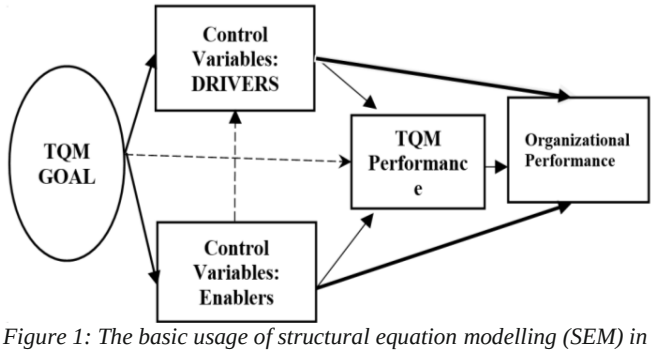


Figure 1: The basic usage of structural equation modelling (SEM) in path analysis with mediation.

The aim of this research is to develop links between different sets of variables. Through questionnaire data is collected on TQM performance in an Indian FMCGs industry. Responses are expressed in five-point Likert scale. We gathered data on TQM drivers and enablers, combined called critical success factors: (Human Resource Management (HRM); Top management commitment (TMC); Process management (PM); Customer focus and satisfaction (CFS); Supplier partnership (SP); Training and learning (TL); Information/analysis/data (INF); Strategic quality planning (SQP); Culture and communication (CC); Benchmarking (BHM); Social and environmental responsibility (SER); Innovation (INV)). To conduct multiple regression analysis on each independent variable with all of the TQM CSFs, we applied SEM for the connection between e.g. HRM factors (employee involvement, employee empowerment, recognition & reward and teamwork) on TQM performance. So, after determining that TQM performance is affected by HRM factors, we use data from the questionnaire survey to interpret the reasons for such connection. Through SEM, we dabbling into canonical correlation among the variables (dependent/independent). Kenneth and Judea (2013) regarded SEM as an inference engine that takes in two inputs, qualitative causal assumptions and empirical data, and derives two logical consequences of these inputs: quantitative causal conclusions and statistical measures of fit for the testable implications of the assumptions. Structural Equation Modelling (SEM) is used to test the relationships between TQM CSFs (Dependent and Independent variables) and TQM performance (unobserved) variables to test the hypotheses and confirm relationships between them. Amos, a structural equation modelling (SEM) software is used to accomplish this part of work. SEM can quickly create models to test hypotheses and confirm relationships among observed and latent variables--moving beyond regression to gain additional insight. This method is preferred by the researcher because it estimates the multiple and interrelated dependence in a single analysis. Every post-hoc analysis in CFA is guided not only with a statistical argument but also with conceptual appropriateness. After all, the CFA is meant to test a hypothesized model that is based on established theory. There's much to be discussed and described about what factor analysis is and what structural equation modelling is, and what the relation between them. Factor analysis is a method (or, more accurate, class of methods) of reducing the observed data into more compact "constructs". This is the "official" or "mainstream" definition. There are various ways (methods) of reducing data, and factor analysis becomes a little complicated in this regard.

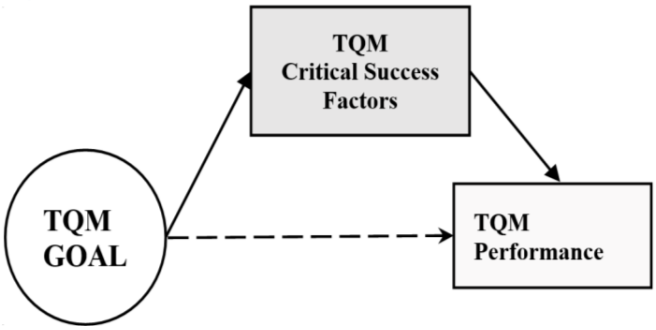


Figure 2: Proposed model for extroversion

Using SEM to understand a domain, if not even the causal relationships between different variables, many specific approaches have been developed but most of them involve much more stringent rules on the data generation, variability and observability of all relevant information than is usually the case anyway. In fact - the field of observational methods (i.e. SEM and other robust estimators) for estimating causal impact really do not care much about the interpretability of a model - because the causal interpretation is derived from the model predictions, not its structure. SEM is potentially one such methodology, as it allows for modelling complexities in behaviour (e.g., model loops, cross-lagged effects, autocorrelation structures, etc.), given adequate sample sizes. One of the strengths of SEM is its flexibility, which permits examination of complex associations, use of various types of data (e.g., categorical, dimensional, censored, count variables),

and comparisons across alternative models. However, these features of SEM also make it difficult to develop generalized guidelines regarding sample size requirements (MacCallum et al. 1999).

3.1 Confirmatory Factor Analysis (CFA)

CFA can be accomplished with SEM (some say that CFA is a form of SEM, I happen to conceive it the other way around; we're probably both partly wrong). Anyway, once you have your conceptual model (a simple sketch on a piece of paper with arrows pointing between variables, showing how and who influences/predicts who), you may proceed to the 'testing'. And now, you have to deal with the type of SEM that's suited for testing (in this case, you'd need to use covariance-based SEM — for instance, use a software package like IBM(TM)'s AMOS).

Another way to look at this question is to start from the types of modeling. Basically, we could use PLS modeling or covariance-based modeling. The first is more suited for exploratory analyses of relations between latent variables, whereas the second is more adequate for measuring the adequacy of the models (how well the model fits the observed data). Once you grasp the common points and the differences between the two methods, you may proceed to make analogies with factor analysis (exploratory versus confirmatory).

3.2 Exploratory Factor Analysis (EFA)

EFA (exploratory factor analysis) can be used to identify (hypothesize) latent constructs (which underlie a group of (co)related measured variables) and based on this 'mathematical argument' the case can be made that several latent constructs are inter-related in a specific way (such as a model specifies). Of course, would not proceed testing/building models just because EFA suggests some underlying constructs. Rather, some dogmatic (theoretical) reasons should first guide your EFA.

NB: We would also have to consider the type of EFA (what method).

The symbols in this diagram are the same as defined earlier. The new representations are the functions which provide a general way to represent the connections between the variables within the parentheses to those on the left-hand side of each node. Composite reliability that achieved 0.70 or above means the scale has good reliability. In general, composite reliability is greater than 0.6 and average variance extracted (AVE) is greater than 0.5, indicating that the reliability of this model is good. Composite reliability (sometimes called construct reliability) is a measure of internal consistency in scale items, much like Cronbach's alpha.

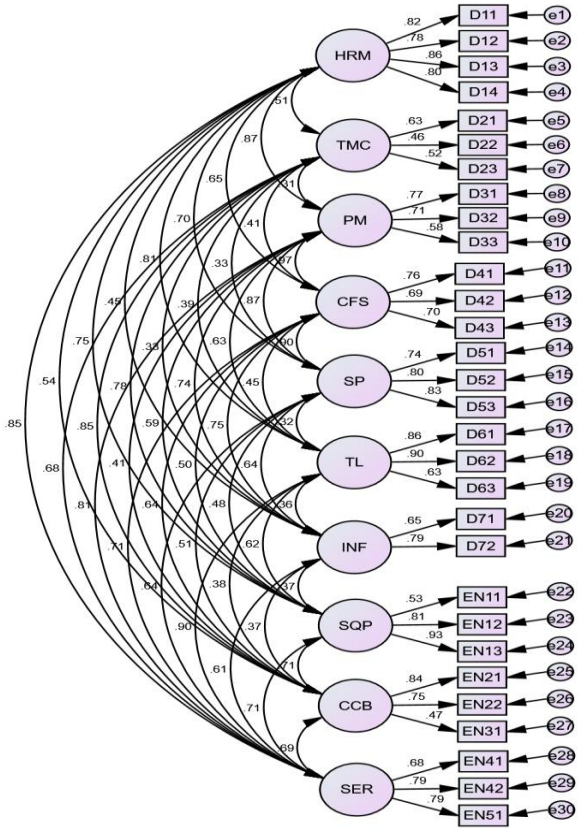


Figure 3: CFA Model

4. Results and discussion

According to the study, we hypothesized twelve paths including seven TQM drivers and five enablers (TQM critical success factors) and thirty sub-factors. Using the SEM investigated that impact of drivers, enablers and firm performance. Results exhibits all the paths are significant ($p < 0.05$). A SEM model divulges the critical success factors of TQM is directly and positively affects the TQM performance which further affect operational, financial and non-financial performances of TQM. Sideridis et al. (2014) advocated that SEM is potentially one such methodology, as it allows for modeling complexities in behavior (e.g., model loops, cross-lagged effects, autocorrelation structures, etc.), given adequate sample sizes. SEM models without measurement models are called path models.

Prior to fitting our SEM, table 3 consist the TQM CSFs as drivers and enablers. The first diagonal element of TQM drivers (D_1) represents the variance of the TQM CSFs which are (arbitrarily) ordered first, the second diagonal element represents those ordered second, and so on. Further, the first off-diagonal element of TQM CSFs (i.e., $D21$) represent the covariance of TQM CSFs for the factors which are ordered second with those which are ordered first, and so on.

Table 1: CFA results of Identified Constructs

Construct	Items	Factor Loadings	Composite Reliability (CR)	AVE
HRM (D_1)	D_{11}	0.82	0.887	0.663
	D_{12}	0.78		
	D_{13}	0.86		
	D_{14}	0.80		
TMC (D_2)	D_{21}	0.72	0.785	0.549
	D_{22}	0.79		
	D_{23}	0.71		
PM (D_3)	D_{31}	0.77	0.732	
	D_{32}	0.71	0.507	
	D_{33}	0.65		
CFS (D_4)	D_{41}	0.76	0.760	
	D_{42}	0.69	0.515	
	D_{43}	0.70		
SP (D5)	D51	0.74	0.833	0.626
	D52	0.80		
	D53	0.83		
TL (D6)	D61	0.86	0.871	0.694
	D62	0.90		
	D63	0.73		
INF(D7)	D71	0.70	0.715	0.557
	D72	0.79		
SQP (E1)	E11	0.69	0.855	0.666
	E12	0.81		
	E13	0.93		
CC (E2)	E21	0.84	0.821	0.605
	E22	0.75		
BHM (E3)	E31	0.74	---	
SER (E4)	E41	0.88	0.861	0.674
	E42	0.79		
INV (E5)	E51	0.79		

Table 2: Model fit

Goodness of Fit Indices	Results	Recommended Standard Value
CMIN/DF-degree of freedom	2.657	< 3
NFI (normed fit index)	0.90	≥ 0.90
NNFI (non-normed fit index)	0.92	≥ 0.90
CFI (comparative fit index)	0.92	≥ 0.90
GFI (goodness fit index)	0.91	≥ 0.90
AGFI(Adjusted goodness of fit index)	0.85	≥ 0.80
RMSEA (root mean square error of approximation)	0.06	<0.10

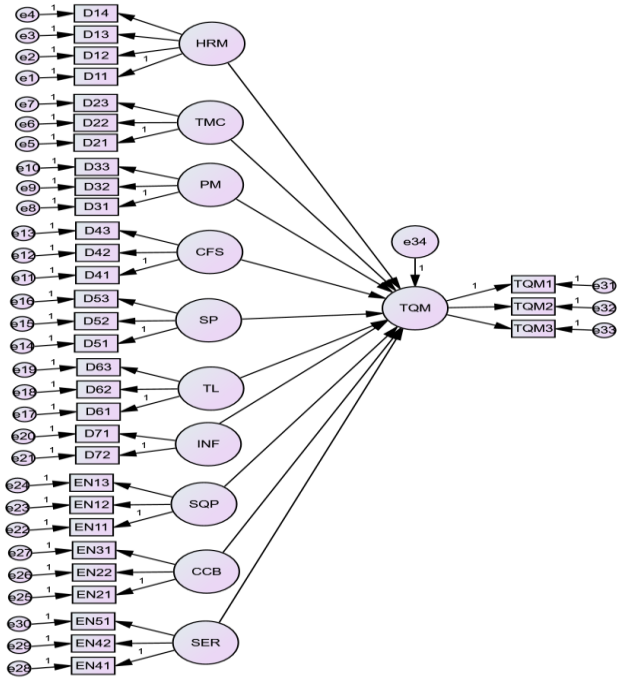


Figure 2: Path Diagram of Structural equation with twelve (two variables E_2, E_3 and E_4, E_5 are combined) explanatory variables

4.1 Assessment of model and hypotheses testing procedures

When the goodness of the model has been confirmed, the next is to test the hypothesized relationships among thevariables (TQM CSFs). Through the running of PLS Algorithm using Smart PLS, the hypothesized model is tested. Therefore, the path coefficients were generated as illustrated in the Figure 2. The p values in hypothesis testing are used to classify the data into two groups being 'significant' or 'insignificant' depending upon whether it 'rejects' or 'fails to reject' the null hypothesis. A level of significance (α level) is set between 0 and 1 as an arbitrary cut off value to determine statistical significance. Analysis of the linkage between the TQM critical success factors and the effectiveness of TQM provides an insight into the prevailing TQM system conditions that could improve/prohibit TQM effectiveness.

Table: 3 Bivariate correlations between variables. Shown below are the correlations between Human Resource Management (HRM); Top management commitment (TMC); Process management (PM); Customer focus and satisfaction (CFS); Supplier partnership (SP); Training and learning (TL); Information/analysis/data (INF); Strategic quality planning (SQP); Culture and communication (CC); Benchmarking (BHM); Social and environmental responsibility (SER); Innovation (INV)

TQM CSFs		D_1	D_2	D_3	D_4	D_5	D_6	D_7	E_1	$E_{2\ \&\ 3}$	$E_{4\ \&\ 5}$
TQM DRIVERS	D_1	--									
	D_2	0.51	-								
	D_3	0.87	0.31	-							
	D_4	0.65	0.41	0.97	-						
	D_5	0.70	0.33	0.87	0.90	-					
	D_6	0.81	0.39	0.63	0.45	0.32	-				
	D_7	0.45	0.33	0.74	0.75	0.64	0.36	-			
TQM ENABLERS	E_1	0.75	0.78	0.59	0.50	0.48	0.62	0.37	-		
	$E_{2\&}E_3$	0.54	0.85	0.41	0.64	0.51	0.38	0.37	0.71	-	
	$E_{4\ \&}$ E_5	0.85	0.68	0.81	0.71	0.64	0.90	0.61	0.71	0.69	-

Table 3: Hypotheses results and estimate

				Estimate	p value	Result
H_1	TQM	<---	HRM	.427	***	Fail to reject
H_2	TQM	<---	TMC	.740	***	Fail to reject
H_3	TQM	<---	PM	.571	***	Fail to reject
H_4	TQM	<---	CFS	.314	***	Fail to reject
H_5	TQM	<---	SP	.657	***	Fail to reject
H_6	TQM	<---	TL	.532	***	Fail to reject
H_8	TQM	<---	SQP	.560	***	Fail to reject
H_9,H_{10}	TQM	<---	CC&BHM	.454	***	Fail to reject
H_{11},H_{12}	TQM	<---	SER& INV	.642	***	Fail to reject
H_7	TQM	<---	INF	.475	***	Fail to reject

* Significant at 5% level of significance

Note: Human Resource Management (HRM); Top management commitment (TMC); Process management (PM); Customer focus and satisfaction (CFS); Supplier partnership (SP); Training and learning (TL); Information/analysis/data (INF); Strategic quality planning (SQP); Culture and communication (CC); Benchmarking (BHM); Social and environmental responsibility (SER); Innovation (INV). All pathways represent the influence of a factor independent from other influences in the model.

5. Conclusions

Structural equation modeling is a technique of modeling data based on the concept of "latent" variables/factors/constructs. It also has several purposes. It may be used to test predictions (and influences) or to measure (validate) models/theories. It also can be used in an "exploratory" fashion or for "confirmatory" purposes. Also, the modeling can be "variance-based" or "covariance-based". When testing predictions (and/or pathways of influence) one can think of a multiple regression model, which may have mediators and moderators. However, instead of using observed (measured) variables, using factors/constructs, which, in turn, are determined by (i.e. they are formative constructs) or are determining (i.e., they are reflexive constructs) the observed variables. Of course the SEM is not without limitations. The most widely used method of estimation is maximum likelihood (ML). Under the general assumptions of sufficiently large sample size, proper model specification, and residuals that are independent and normally distributed, ML provides asymptotically unbiased, consistent, and efficient parameter estimates and standard errors (Bollen, 1989).

References

[1] Barros, S., Sampaio, P., and Saraiva, P. (2014). The Relationship between Quality Approaches and their Impact on Portuguese Companies´ Quality Performance", International Conference on Industrial Engineering and Operations Management, Bali, Indonesia.

[2] Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

[3] Brotherton, B., & Shaw, J. (1996). Towards an identification and classification of critical success factors in UK hotels Plc. International Journal Hospitality Management, 15(2), 113-135.
[4] Demirbag M, Tatoglu E, Tekinkus M, Zaim S. (2006). An analysis of the relationship between TQM implementation and organizational performance: Evidence from Turkish SMEs. Journal of Manufacturing Technology Management, 17(6), Pp. 829–847.
[5] Dixon, J.R., Nanni, A.J., Vollmann, T.E. (1990). The new performance challenge: Measuring operations for world-class competition, Homewood, IL, Dow Jones-Irwin.
[6] Gao (1991). Management practices: US companies improve performance through quality efforts. Report No. GAO/NSIAD-91-190, US General Accounting Office (GAO), Washington, DC
[7] Holland, C. P., & Light, B. (2003). A framework for understanding success and failure in enterprise resource planning system implementation. In G. Shanks, P. B. Seddon, & L. P. Willcocks (Eds.), Second-wave enterprise resource planning systems (pp. 180-195). United Kingdom: Cambridge University Press.
[8] John S. Oakland, (2003). TQM: Text with Cases, Third Ed. Butterworth-Heinemann an imprint of Elsevier, Linacre House, Jordan Hill, Oxford OX2 8DP, 200 Wheeler Road, Burlington MA 01803.
[9] Kenneth A. Bollen and Judea Pearl, (2013), Eight Myths About Causality and Structural Equation Models- A TECHNICAL REPORT, In S.L. Morgan (Ed.), Handbook of Causal Analysis for Social Research.
[10] Kaynak, H. (2003). The relationship between total quality management practices and their effects on firm performance. Journal of operations management, 21(4), 405-435.
[11] Lakhal, L., Pasin, F., &Limam, M. (2006). Quality management practices and their impact on performance. International Journal of Quality & Reliability Management, 23(6), 625-646.
[12] Marais, M., du Plessis, E., &Saayman, M. (2017). A review on critical success factors in tourism. Journal of Hospitality and Tourism Management, 31, 1–12.
[13] MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. Psychological Methods, 4, 84-99.
[14] Nguyen, A. D., Pham, C. H., & Pham, L. (2016). Total quality management and financial performance of construction companies in Ha Noi. International Journal of Financial Research, 7(3), 41-53.
[15] Öztayşi B., Kutlu A.C. (2011). Determining the Importance of Performance Measurement Criteria Based on Total Quality Management Using Fuzzy Analytical Network Process. In: Wang Y., Li T. (eds) Practical Applications of Intelligent Systems. Advances in Intelligent and Soft Computing, vol 124. Springer, Berlin, Heidelberg.
[16] Odiorne, G. S. (1987). Measuring the unmeasurable: Setting standards for management performance. Business Horizons, 30(4), Pp 69-75, ISSN 0007-6813.
[17] Pfau, L.D. (1989). Total quality management gives companies a way to enhance position in global marketplace. Industrial Engineering, Vol. 4 No. 1, pp. 17-21.
[18] Robson, K., MacIntyre, J. and Trimble, R. (2013). Measuring the status and alignment of maintenance and manufacturing strategies - the development of a new model and diagnostic tool. Journal of Quality in Maintenance Engineering, Vol. 19 No. 4, pp. 381-397.
[19] Robert P. Abelson,(1995). Statistics as Principled Argument, Psychology Press, ISBN: 0805805281
[20] Sahoo, S. (2020). Aligning operational practices to competitive strategies to enhance the performance of Indian manufacturing firms. Benchmarking: An International Journal, ahead-of-print(ahead-of-print).
[21] Santos-Vijande, María Leticia, Alvarez-Gonzalez, Luis I. (2007). TQM and firms performance: An EFQM excellence model research based survey, International Journal of Business Science & Applied Management (IJBSAM), Vol.2(2),21-41; ISSN: 1753-0296
[22] Steven E. Brigham. (1993). TQM: Lessons we can Learn from Industry. Change: The Magazine of Higher Learning, 25:3, 42-48,
[23] Saraph, J. V., Benson, P. G., & Schroeder, R. G. (1989). An instrument for measuring the critical factors of quality management. Decision sciences, 20(4), 810-829.
[24] Sideridis, G., Simos, P., Papanicolaou, A., & Fletcher, J. (2014). Using Structural Equation Modeling to Assess Functional Connectivity in the Brain. Educational and Psychological Measurement, 74(5), 733–758.
[25] Sadikoglu, E., &Zehir, C. (2010). Investigating the effects of innovation and employee performance on the relationship between total quality management practices and firm performance: An empirical study of Turkish firms. International journal of production economics, 127(1), 13-26.
[26] Sinclair, D., & Zairi, M. (2000). Performance measurement: a critical analysis of the literature with respect to total quality management. International Journal of Management Reviews, 2(2), 145–168.
[27] V. Kumar, D. De Grosbois, F. Choisne and U. Kumar. (2008). Performance measurement by TQM adopters. The TQM Journal, Vol. 20(3), pp. 209 – 222.
[28] Wilson, D. D., & Collier, D. A. (2000). An Empirical Investigation of the Malcolm Baldrige National Quality Award Causal Model. Decision Sciences, 31(2), 361–383.
[29] Walsh, A., Hughes, H., & Maddox, D. P. (2002). Total quality management continuous improvement: is the philosophy a reality? Journal of European Industrial Training, 26(6), 299–307.
[30] Yusuf, Y., Gunasekaran, A., & Dan, G. (2007). Implementation of TQM in China and Organisation Performance: An Empirical Investigation. Total Quality Management & Business Excellence, 18(5), 509–530.
[31] Zairi, M. (1994). Measuring performance for Business Results. London, Chapman & Hall

Cite this article as: R. S. Mishra, Rakesh Kumar, Effect of critical success factors (CSFs) on the organizational performance, International journal of research in engineering and innovation (IJREI), vol 5, issue 6 (2021), 341-348.
https://doi.org/10.36037/IJREI.2021.5602

View publication stats

Trans Indian Inst Met
https://doi.org/10.1007/s12666-021-02440-1

ORIGINAL ARTICLE

Effect of Process Parameter on Tensile Strength and Hardness of SS304 Processed by Microwave Radiation

Uma Gautam[1] · **Vipin**[1]

Received: 13 August 2021 / Accepted: 27 September 2021
© The Indian Institute of Metals - IIM 2021

Abstract An investigation into the tensile strength and micro hardness of stainless steel 304 plates united by the heating radiation of microwaves has been studied in this paper. Three different sizes, 5 µm, 10 µm and 15 µm of nickel powder were used as an interfacing material. A domestic IFB microwave with 900 W output having a fixed frequency of 2.45 GHz was used during the experiment. The different sizes of nickel powder as slurry, variation in slurry by weight, and welding time were chosen as principal variables. Tensile strength and micro hardness were selected to study the mechanical strength of the joint. The centre composite design (CCD) was used as the design of experiments to optimize the results. The results showed that, as the welding time increases, tensile strength and hardness also increase, while tensile strength and micro hardness decrease as the particle size of nickel powder increases. There is a positive effect of slurry weight on the mechanical properties of the joint. From the result, the effect of welding time put up the most, followed by the particle size of powder and weight of slurry material. The optimum values as input variables of welding time, particle size, and slurry weight were found to be 13 min, 5 µm, and 3 g.

Keywords Microwave hybrid heating · Centre composite design · Tensile strength · Micro hardness · Temperature measurement

1 Introduction

In the EM spectrum, microwaves having a frequency range from 300 MHz to 300 GHz are functional for material processing applications. Microwaves are made up of forms of energy and fragments of the electromagnetic (EM) wave spectrum, which progresses in space and generated by the communication of magnetic and electric fields. From longest to shortest frequency, the spectrum includes number of rays having wavelengths between 1 and 1 mm [1]. The most used frequencies for domestic and industrial heating purposes are 915 MHz & 2.45 GHz with 900 W output rated power and that leads to eminent penetration depth for the materials. Hence, these are suitable for most laboratory reaction conditions [2]. Heating of a substance using EM radiation is found in a material's ability to effectively soak up energy. Microwave-processed materials have the prospect of speedy energy with efficiency comparable with other heating methodologies. In the microwave procedure, heat is transferable to all particles rather than to the surface of the material by transferring energy via electromagnetic waves, which enlarge the heat dissipation. In microwave processing, heat is independent of surface area, but turns to volume; thus, a contrary heating outline can be observed. Ahmad and Siores studied the inverted profile of heating which is generated by microwave heating. They characterized the ceramics of zirconia–alumina–silica at elevated temperatures and analysed the behaviour of temperature and time at different power levels [3]. The competence of turning electromagnetic emissions into heat is nearly100%. However, absorption of microwaves by metallic materials is a challenge because electromagnetic waves are reflected by metals at room temperature. Further, associating the metallic plates in distinct shapes inside domestic microwaves is described

✉ Uma Gautam
uma.mech.gautam@gmail.com

[1] Department of Mechanical Engineering, Delhi Technological University, New Delhi, India

Published online: 08 November 2021

already and a patent has also been filed on cladding & coating of metallic work pieces, which is a novel method to upgrade the physical effects of materials, a productive processing approach for those materials that are challenging, having time, space & energy savings, and new materials with better microstructures [4–6]. Shashi Prakash Dwivedi and Satpal Sharma observed the outcome of input values on the tensile strength of MS using microwave radiation. They selected three different rated powers of microwave ovens having a frequency of 2.45 GHz to control the output data. The Box–Behnken method of RSM was selected for experimental design. The outcomes exhibited that, as rated power increased, tensile strength decreased while tensile strength rose with increasing welding time & temperature [8]. Uma Gautam & Vipin, overviewed the materials that can be processed by microwave radiation. They also reviewed the various process parameters of microwave welding and discussed their effect on the weld joint [9]. Uma Gautam et al. studied the microwave welding of mild steel to examine mechanical properties at the weld zone. They used a mild steel specimen (20 × 20 × 2 mm) and welded it by a domestic microwave having a frequency of 2.45 GHz and an output power level of 900 W. They performed tests related to metallic mechanical properties and concluded that the joining occurred between 640-840 s. The Austenitic structure could be seen through microstructure testing with negligible porosity defects. The hardness on the welded area was 58–59 HRB and on HAZ was 51–52 HRB [10, 11]. Kanwarjeet Singh & Satpal Sharma investigated cobalt alloy (EWAC 1006 EE) to evolve clad on a P20 tool steel substrate using an LG domestic microwave having a fixed frequency of 2.45 GHz with a 900 W rated output by the hybrid heating method of microwave with suscepter. The experiments were carried out at 900 W power and concluded that there was a clad of 1 mm thickness evolved in 15 min. The resulted clad was characterized by SEM, XRD, and Vicker's microhardness, which revealed a uniform and crack-free clad due to volumetric microwave heating. The appearance of hard carbides like W2C and Cr23C6, which helped to enhance the hardness of the coating, could be observed by XRD. The value of Vicker's microhardness was found to be 761 + −1.5Hv [12–14].

2 Experimental Details

2.1 Material Selection

Stainless steel 304 has wide application in the fabrication industries due to its exemplary impervious corrosion, adequate formable property and adds value in strength as cold worked. It has a composition of 18% chromium and 8% nickel and is the most widely used material in the austenitic group [7]. Experiments were carried out using commercial grade 304 Stainless Steel plates having dimensions of (40 × 20 × 3) mm. A square butt configuration was chosen for joint preparation. The chemical composition and mechanical properties of 304 stainless steel are shown in Tables 1 and 2.

2.2 Selection of Inputs and Experimental Procedure

Microwave IFB multimode oven with a frequency of 2.45 GHz and a power output of 900 W were selected as fixed parameters. Three process parameters were chosen as input variables: welding time, powder size, and slurry weight. The first parameter was selected as welding time and trials were conducted to find out the optimum time at which metal edges fused to form a weld. The second parameter was nickel powder of three different sizes and to prepare the slurry we mixed it with epoxy resin. A different combination of nickel powder with epoxy was tried in different proportions. The best results were shown at an 85:15 and 90:10 ratio of nickel powder with resin. The third parameter was slurry weight in grams. Three values which gave the optimum result were selected for slurry weight by conducting a number of trials. Initially, trial runs were conducted to set ranges of input variables. First, welding time varied from 6 to 13 min in steps of 30 s. No joining could be observed during the variation in process

Table 2 Mechanical properties of SS 304

Property	Values
Ultimate tensile strength	706 MPa
Vickers microhardness	208Hv

Table 1 Chemical composition of stainless steel 304

Chromium (cr) %	Nickel (Ni) %	Manganese (Mn) %	Silicon (Si) %	Carbon (C) %	Phosphorous (P) %	Iron (Fe) %
18.29	8.05	1.14	0.25	0.058	0.041	71.13

time of upto 9 min. Little joining took place from 9 to 10 min. Welding time from 11 to 13 min produced a sound weld joint. After 13 min of process time, a distorted weld area was observed due to overheating of the specimen. The other two input variables were also selected by adopting the same process.

Firstly, a cavity slot of the same size & dimension as the metal pieces was made on the upper face of the insulating ceramic brick. The facing surfaces of metal pieces were cleaned with different grades of emery paper. After that, surfaces were cleaned ultrasonically for better results. After cleaning, facing surfaces were dipped into an acetone solution to avoid contamination. After that, prepared weighed slurry could be applied to cleaned faces and specimens placed into the slot of the brick. Insulating brick prevents direct contact of microwaves with specimens as metal reflects microwaves at room temperature. A 1 mm graphite sheet was placed over the specimen to prevent mixing of the slurry with charcoal powder. Fine stone charcoal powder was spread over the graphite sheet and specimen at all corners. The whole setup was placed inside the oven. As we switched on the microwave for a predetermined time, the charcoal powder absorbed the microwave and burning started within 40 s. The same heat transferred to the filler material through the graphite sheet. As the filler material started melting, heat transferred to facing surfaces and melting of surfaces started due to the diffusion process which took place between the slurry and

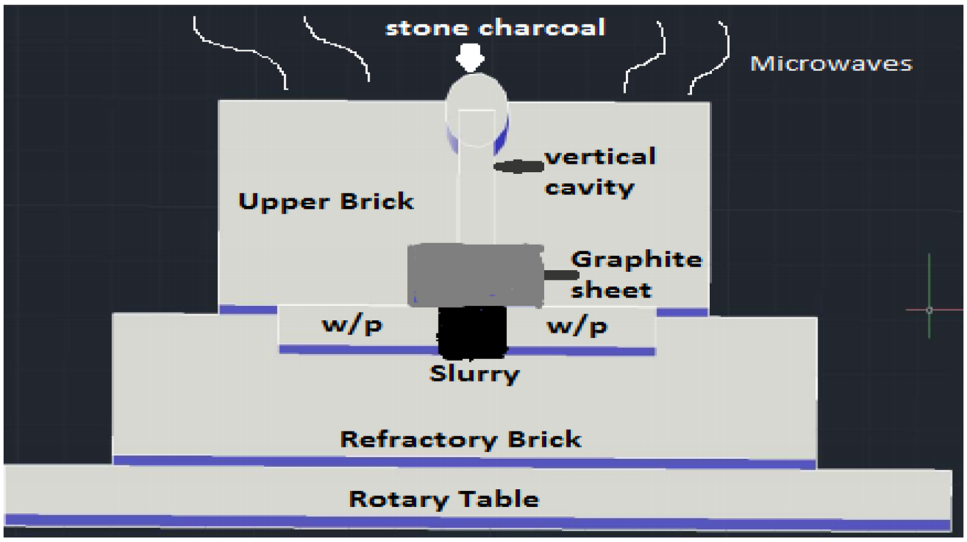


Fig. 1 Microwave Welding setup

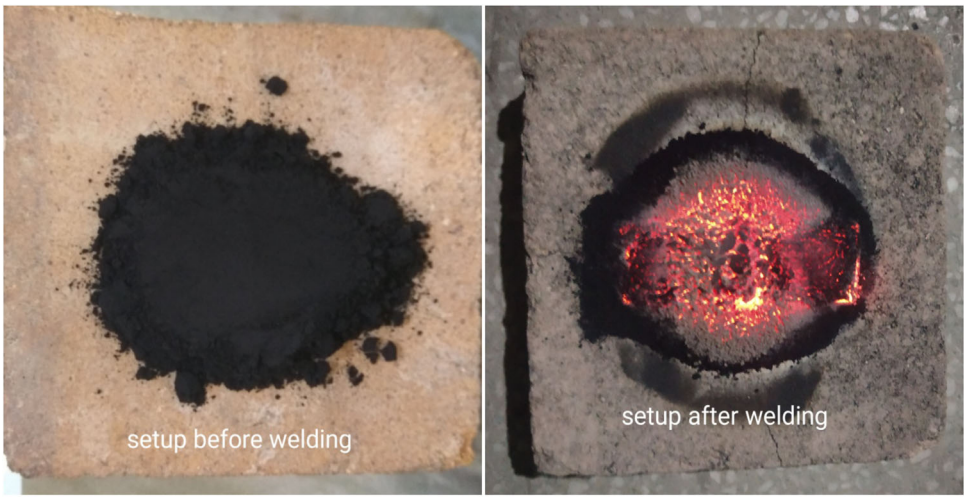


Fig. 2 Setup before welding & after welding

metal surfaces, which completed the joint. After completion of the joint, the whole setup was taken out of the oven and the metal pieces were exposed to environmental cooling. A joint area was filed prior to testing for better results. The same process was adopted for all the values of input variables as per design expert (Figs. 1, 2, and 3).

3 Result and Discussion

3.1 Analysis for Tensile Strength

In the present study, the effect of welding process time, powder size and slurry weight on tensile strength and hardness of the specimen have been studied. Mathematical modelling has been developed by giving input process variables for tensile strength using centred composite design with face centred points of response surface methodology using design expert software for design matrix (Table 3). Analysis of variance indicates the value of input variables and their interaction at 95% of the certified level. The Design summary shows the quadratic model and anova accepted the model with significant values. The ratio of variants among the factors to the variants within the factors is indicated by F-value. The probability factor which should be below 0.05 is indicated by *P*-value [15]. The significant F-value of the model is 1390.79. Values below 0.0500 of Prob > F specify model terms that are acceptable. The present study shows that all the model terms (A, B, C, AB, AC, BC, A^2, B^2, and C^2) are significant and all model terms are within range (less than 0.1000). The F-data of Lack of Fit is 0.39. The coefficient of determination (R*R) is 0.9992 for the response, which is used to check the fit model. The above value confirms that there is a 99.92% chance that experimental data is compatible with the predicted data of the model. This model gives R*R values less than 1 and, for the exactness of the model; this value always lies between 0 and 1. For the model to be more significant, the value of the adjusted determination coefficient should be high, and this model confirms that value (adj. R*R = 0.9985). The value of pred. R*R is 0.9968, which clarifies that the model can explain 99.6% of variation while forecasting new results. In this model, the value of the coefficient of variation is low (0.34), which indicates that there is less divergence between the experimental and predicted values. In this investigation, the value of adequate precision is 127.017 and values above 4 indicate desirability. From the above investigation, the final developed model for tensile strength, which includes significant factors and interactions, is

$$\begin{aligned}&\text{Model Equation for tensile strength}\\&= +1205.90000 - 168.90000 * \text{ Welding time}\\&+ 1.90000 * \text{ Slurry weight}\\&- 1.00000 * \text{ Powder size}\\&- 1.25000 * \text{Welding time} * \text{ Slurry weight}\\&+ 0.45000 * \text{ Welding time} * \text{ Powder size}\\&+ 0.25000 * \text{ Slurry weight} * \text{ Powder size} + 9.00000\\&* \text{ Welding time}^2\\&+ 3.50000 * \text{ Slurry weight} + 0.12000 * \text{Powder size}^2\end{aligned}$$

The normal plot of residual analysis can be used for acceptability of the model. It shows the deviation of values from the normal straight line. For the betterment of the model, values should lie around the straight line. Furthermore, to look into the adequacy of the model, a graph is plotted between the predicted and residual values. From the graph, it can be predicted that all values are in order and lie around the middle line and it also concluded

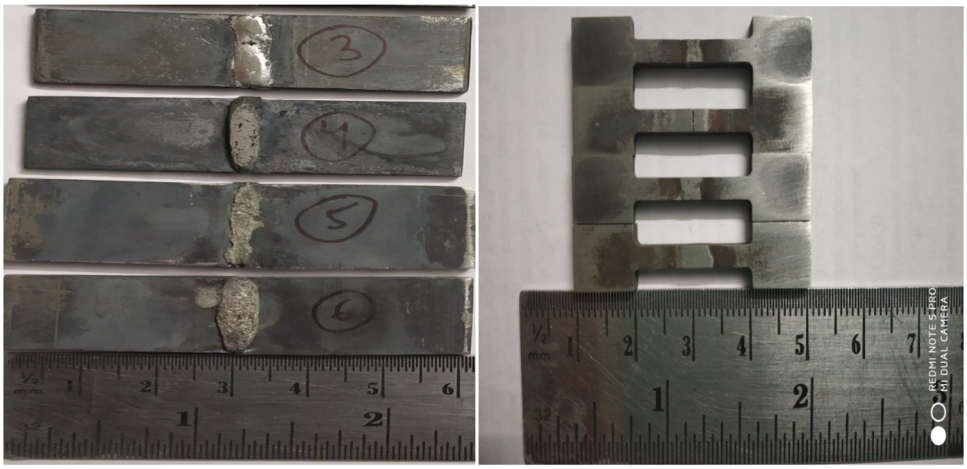


Fig. 3 Welded pieces and pieces for Tensile Test

Table 3 Tensile strength using CCD

Std	Run	Block	Welding time	Slurry weight	Powder size	Tensile strength	Hardness
20	1	Block 1	0	0	0	424	372
11	2	Block 1	0	− 1	0	425	370
4	3	Block 1	1	1	− 1	501	454
10	4	Block 1	1	0	0	483	438
13	5	Block 1	0	0	− 1	441	380
5	6	Block 1	− 1	− 1	1	368	308
12	7	Block 1	0	1	0	431	376
14	8	Block 1	0	0	1	414	352
19	9	Block 1	0	0	0	425	371
7	10	Block 1	− 1	1	1	381	312
16	11	Block 1	0	0	0	424	370
18	12	Block 1	0	0	0	424	369
1	13	Block 1	− 1	− 1	− 1	403	328
17	14	Block 1	0	0	0	424	371
8	15	Block 1	1	1	1	480	426
15	16	Block 1	0	0	0	420	366
9	17	Block 1	− 1	0	0	384	320
2	18	Block 1	1	− 1	− 1	498	450
3	19	Block 1	− 1	1	− 1	409	329
6	20	Block 1	1	− 1	1	474	414

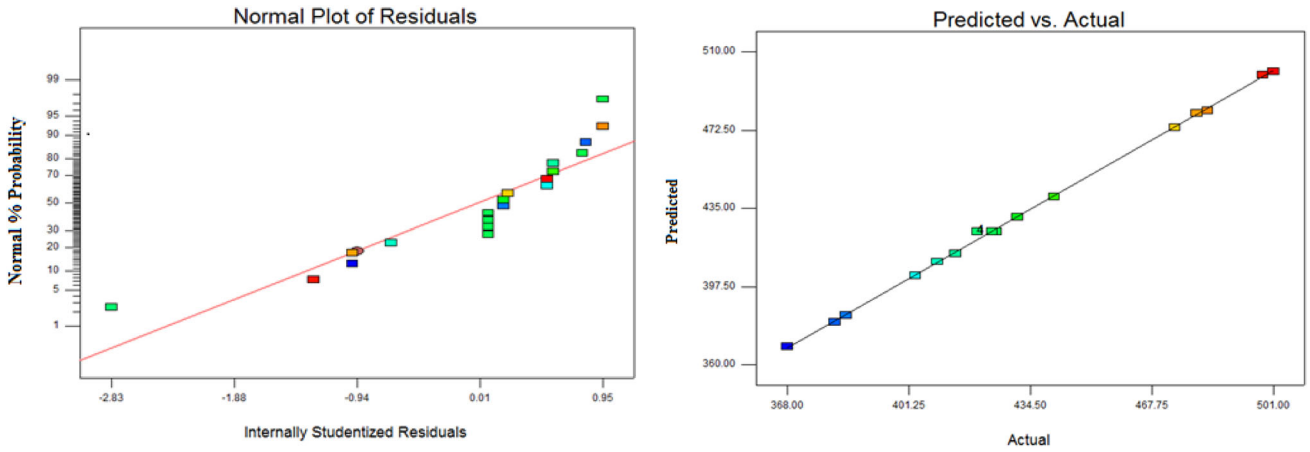


Fig. 4 Predicted v/s actual values

that error distribution is normal. To check the variation of a single input parameter, a graph between the residual and the input parameter can be plotted. From the graph, it can be concluded that variations are within the range (Figs. 4, 5, and 6).

3.2 Welding Time Effects on Tensile Value

When the welding time increases from least value to maximal value, the tensile value of the joint material increases [8]. As welding time increases, more heat transfers to the material from the charcoal to nickel slurry to the material interface. The time required to heat the material depends on the heat and mass transfer characteristics of the material. For better tensile strength, more welding time is always desirable within the permissible range (Tables 4 and 5).

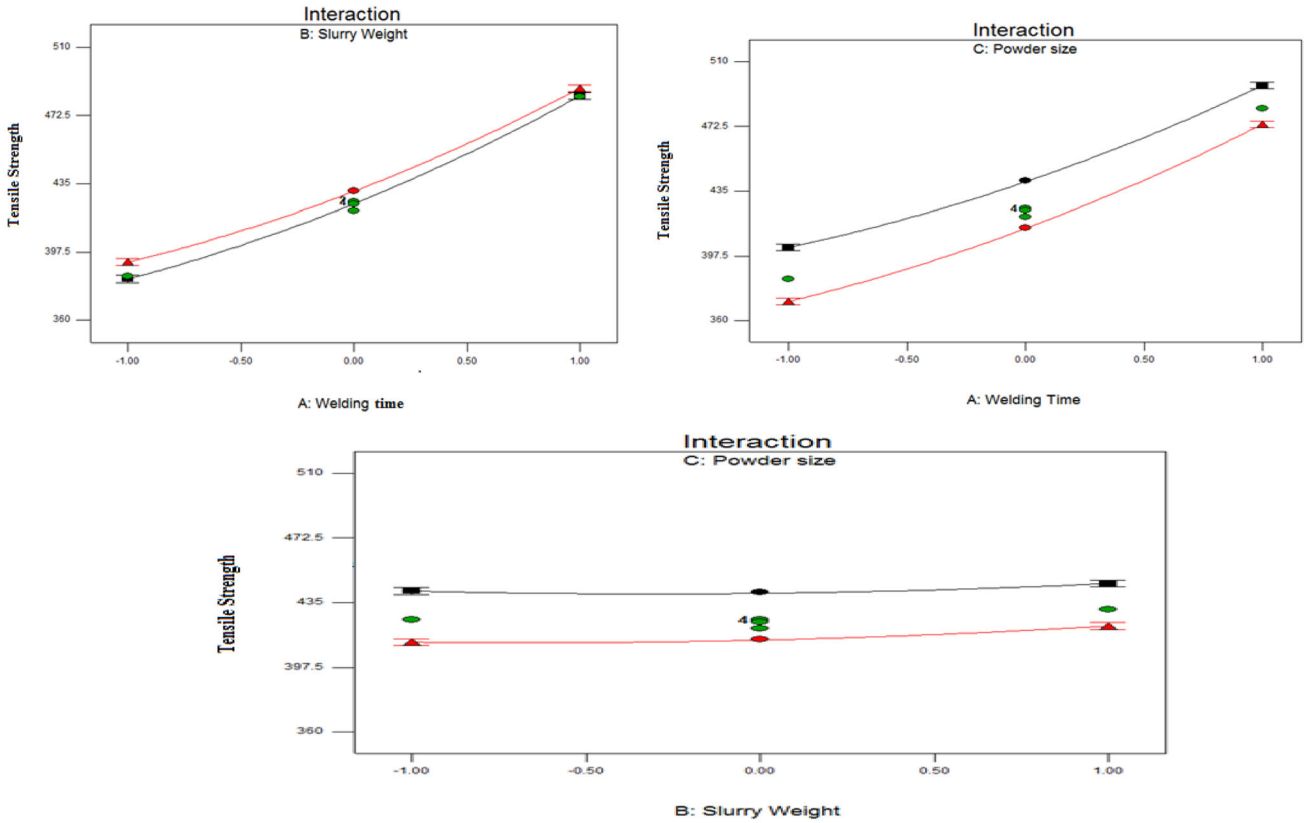


Fig. 5 Effect of interaction of input parameter on Tensile strength

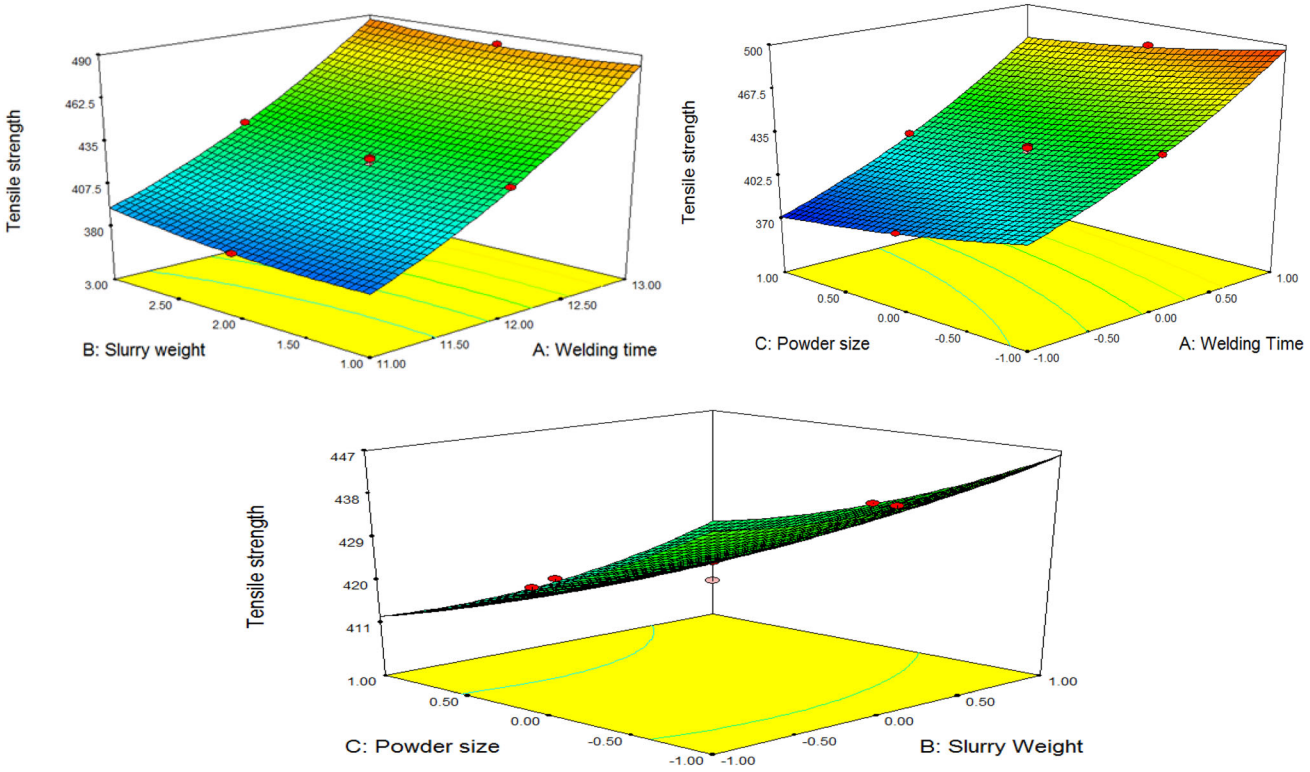


Fig. 6 3-D surface plot of interaction of input and output for tensile strength

Table 4 Anova values for tensile strength

Source	Sum of Squares	df	Mean square	F-value	*p*-value Prob > F	
Model	27,036.95	9	3004.11	1390.79	< 0.0001	Significant
A-Welding time	24,108.10	1	24,108.10	11,161.16	< 0.0001	
B-Slurry weight	115.60	1	115.60	53.52	< 0.0001	
C-Powder size	1822.50	1	1822.50	843.75	< 0.0001	
AB	12.50	1	12.50	5.79	0.0370	
AC	40.50	1	40.50	18.75	0.0015	
BC	12.50	1	12.50	5.79	0.0370	
A^2	222.75	1	222.75	103.12	< 0.0001	
B^2	33.69	1	33.69	15.60	0.0027	
C^2	24.75	1	24.75	11.46	0.0069	
Residual	21.60	10	2.16			
Lack of fit	6.10	5	1.22	0.39	0.8354	Not significant
Pure error	15.50	5	3.10			
Cor total	27,058.55	19				
SD		1.47		R*R		0.9992
Mean		431.65		Adjusted R*R		0.9985
C.V%		0.34		PredictedR*R		0.9968
				Adeq.precision		127.017

Table 5 Anova values for Hardness

Source	Sum of Squares	df	Mean Square	F-value	*p*-value Prob > F	
Model	36,469.15	9	4052.13	1348.36	< 0.0001	Significant
A-Welding time	34,222.50	1	34,222.50	11,387.66	< 0.0001	
B-Slurry weight	72.90	1	72.90	24.26	0.0006	
C-Powder size	1664.10	1	1664.10	553.74	< 0.0001	
AB	15.13	1	15.13	5.03	0.0487	
AC	91.13	1	91.13	30.32	0.0003	
BC	15.13	1	15.13	5.03	0.0487	
A^2	218.27	1	218.27	72.63	< 0.0001	
B^2	23.27	1	23.27	7.74	0.0194	
C^2	46.02	1	46.02	15.31	0.0029	
Residual	30.05	10	3.01			
Lack of fit	7.22	5	1.44	0.32	0.8840	Not significant
Pure error	22.83	5	4.57			
Cor total	36,499.20	19				
SD		1.73		R*R	0.9992	
Mean		373.80		Adjusted R*R	0.9984	
C.V%		0.46		predicted R*R	0.9963	
				Adeq. Precision	120.90	

3.3 Effect of Powder Size on Tensile Value

From figure [6], it is clear that as the powder size enlarges, the tensile strength of the material drops off [16]. More heat is absorbed at less time and this characteristic can be observed as particle size decreases from its maximum value to its minimum value. This leads to fineness in the microstructure of the welded joint, which increases the tensile strength.

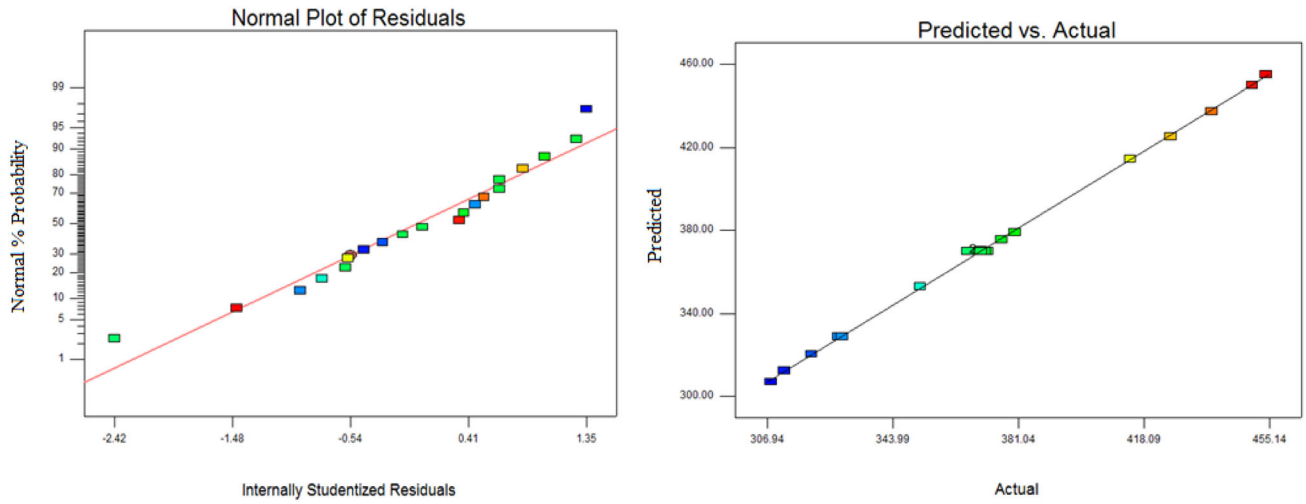


Fig. 7 Predicted v/s Actual graph for Hardness

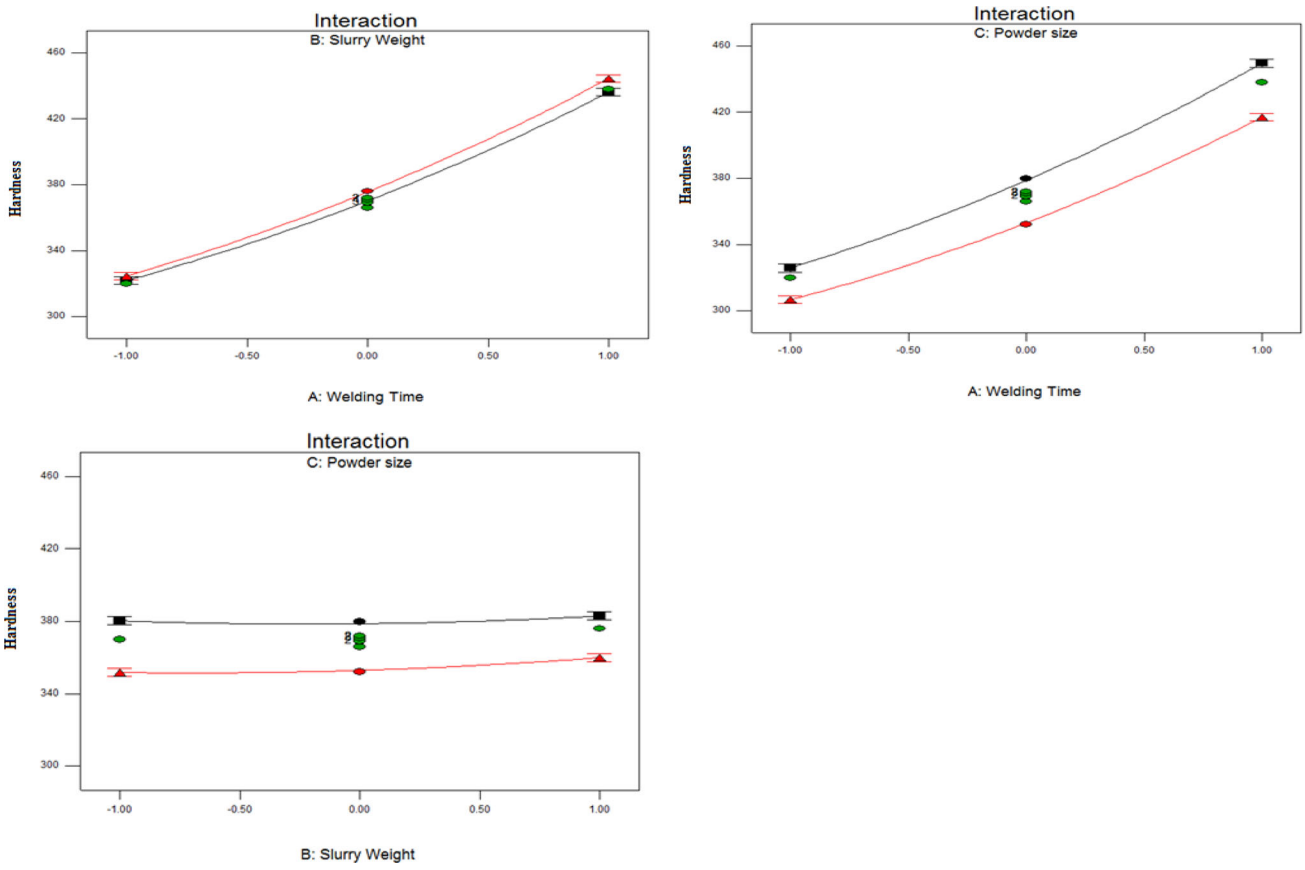


Fig. 8 Interaction curve for Hardness

3.4 Effect of Slurry Weight on Tensile Value

From figure [6], as slurry weight increases, there is a slight increment in the value of tensile strength of the welded specimen because as the weight of the interfacing material increases, more nickel particles are observed inside the joint area that are accountable for the increased value of tensile strength.

3.5 Interactive Effect of Input Variable on Tensile Value

A graph is plotted which connects the welding time, powder size, slurry weight and tensile strength. From the graph, it can be concluded that as welding time and slurry weight increase, tensile strength also increases. More effect of welding time can be predicted on tensile strength than slurry weight.

3.6 Analysis of Hardness

The F-value of model 1348.36 indicates the significance of the model. For the model to be acceptable, the P values have to be below 0.0500. From the model, all values (A, B, C, AB, AC, BC, A^2, B^2, C^2) are below 0.0500, which justifies the significant model, and if having more insignificant model terms, then it may require model reduction, which justifies the model. The F-value of Lack of Fit is 0.32, which hints that the model is fit because lack of fit shows significant terms relative to pure error. There is an acceptable range between the Pred R-Squared (0.9963) and the Adjusted R-Squared (0.9984). Adequate precision computes signal–noise ratio, which should be more than 4, and this model shows a value of 120.900, which justifies the adequacy of the signal.

3.7 Final Equation for Hardness:

$$\begin{aligned}\text{Hardness} = {} & +924.01818 - 151.31818 * \text{Welding time} \\ & - 28.18636 * \text{Slurry weight} \\ & + 8.24273 * \text{Powder size} + 1.37500 * \text{Welding time} \\ & * \text{Slurry weight} \\ & - 0.67500 * \text{Welding time} * \text{Powder size} \\ & + 0.27500 * \text{Slurry weight} * \text{Powder size} \\ & + 8.90909 * \text{Welding time}^2 \\ & + 2.90909 * \text{Slurry weight}^2 \\ & - 0.16364 * \text{Powder size}^2\end{aligned}$$

3.8 Welding Time Effects on Hardness

Welding time increases from the lowest value to the highest value, the hardness of the welded material also increases. As welding time increases, more heat is

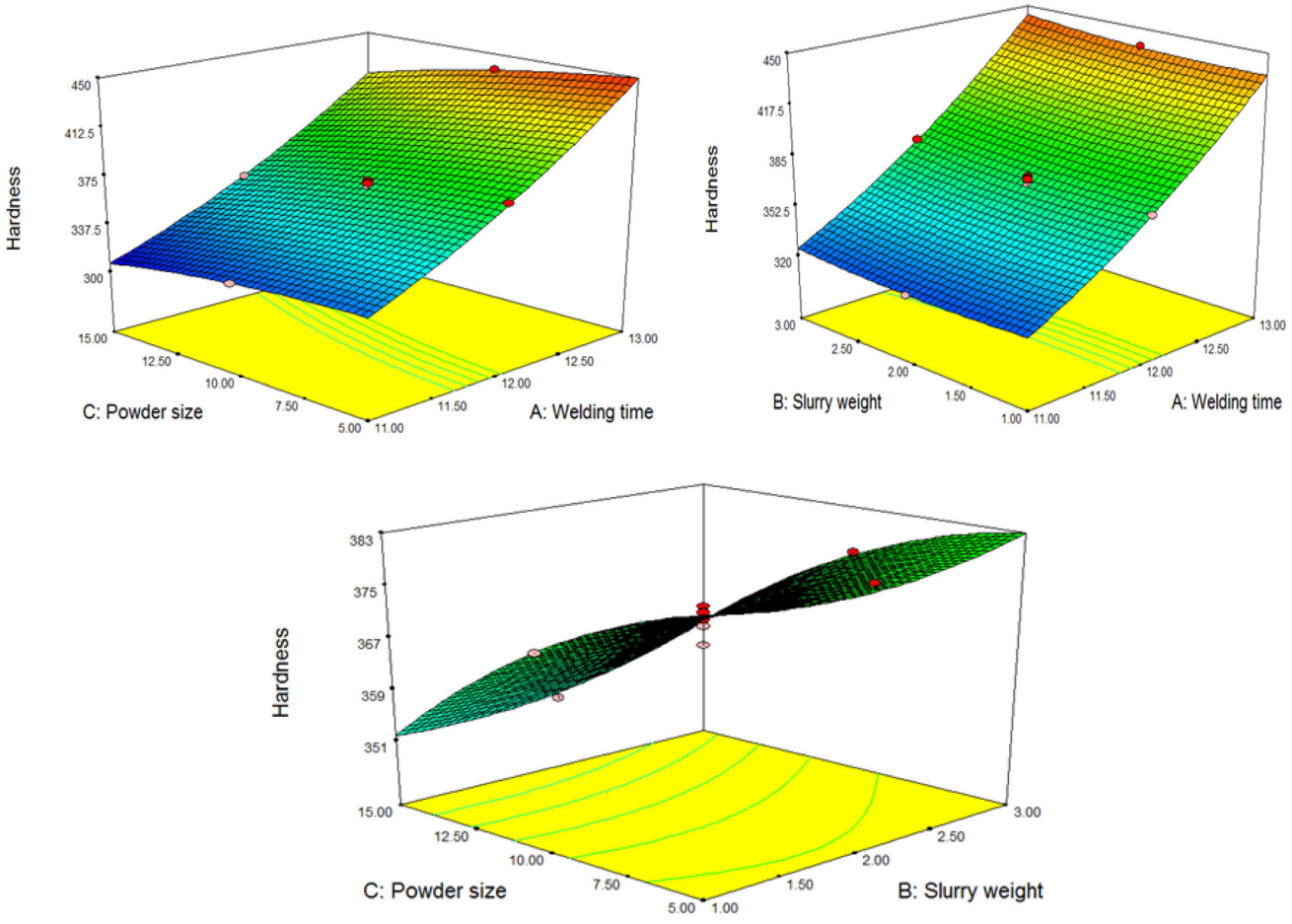


Fig. 9 3D plot curve of interaction for Hardness

transferred to the Ni powder, which increases the melting capacity of Ni particles and, due to this, the strength of the joint increases (Figs. 7, 8, and 9).

3.9 Effect of Powder Size on Hardness

From the figure [9], it is predicted that as the powder size grows, the joint area hardness will reduce [16]. More heat is absorbed at less time and these characteristics can be observed as particle size decreases from its maximum value to its minimum value. This leads to fineness in the microstructure of the welded joint, which increases the hardness value of the joint area.

3.10 Effect of Slurry Weight on Hardness

As slurry weight increases, there is an increment in the value of hardness of the welded specimen because more Ni particles are melted, which increases the value of hardness in the joint area.

3.11 Interactive Effect of Input Variable on Tensile Value

An interactive graph is plotted linking the welding time, slurry weight and particle size. From the graph [9], it can be concluded that as welding time and slurry weight increase, hardness also increases.

4 Conclusion

- By analysing the above data, concept of hybrid heating using microwave can be used for uniting SS 304 using Ni powder at the micron level.
- From the Anova, welding time is the most dominant factor among the three input parameters, followed by powder particle size and slurry weight.
- The values of tensile strength and microhardness of welded material increase as welding time and slurry weight increase, but both the output values decrease as particle size increases.
- Experimental settings for input parameters are 13 min of welding time, 3 g of slurry weight, and 5 micron m of powder size, which results in a maximum value of tensile strength (501), and microhardness (454) in the welded area.

References

1. M.S. Srinath, Apurbba Kumar Sharma, and Pradeep Kumar, Investigation on microstructural and mechanical properties of microwave processed dissimilar joints, Journal of Manufacturing Processes 13(2011), pg no.141–146
2. Osepchuk John M., A history of microwave heating applications, IEEE Trans Microwave Theory Tech., MIT (1984); 32(9):1200–24
3. Siores E, Rego D. Microwave applications in materials joining, J Mater Process Technology 1995; 48:619–25
4. Satnam Singh, Dheeraj Gupta and Vivek Jain, Recent applications of microwaves in materials joining and surface coatings, Journal of Engineering Manufacture, DOI: https://doi.org/10.1177/0954405414560778 –15 (2015) IMechE
5. D. Gupta, P.M. Bhovi, A.K. Sharma,and S. Dutta, Development and characterization of microwave composite cladding. J. Manuf. Process. (2015) 14, 243–249
6. Sharma AK, Srinath MS, and Pradeep K, Microwave joining of metallic materials. Indian Patent. (2009). Application no. 1994/Del/2009.
7. A Designer handbook series N 9014, Design guidelines for selection and use of stainless steel, distributed by Nickel Development Institute.
8. Shashi Prakash Dwivedi and Satpal Sharma, Effect of Process Parameters on Tensile Strength of 1018 Mild Steel Joints Fabricated by Microwave Welding, Metallogr. Microstruct. Anal. DOI https://doi.org/10.1007/s13632-013-0109-1 (2014) 3:58–69
9. Uma G, Vipin V, *Overview of welding using microwave radiation, international conference on advanced production and industrial engineering*, (2018) paper id ICAPIE-2018-PE MT 226.
10. Ms Uma Gautam & Vipin, Joining of Metal using Microwave Energy, International Conference on Emerging Trends in Electro-Mechanical Technologies and Management *Advances in Electromechanical Technologies* Lecture Notes in Mechanical Engineering (2019) https://doi.org/https://doi.org/10.1007/978-981-15-5463-6_92
11. Ms Uma Gautam et al, Microwave welding of Mild Steel in *Advances in Electromechanical Technologies* Lecture Notes in Mechanical Engineering (2019) https://doi.org/https://doi.org/10.1007/978-981-15-5463-6_79
12. Kanwarjeet S, Satpal S, Development and characterizations of co based microwave clad. *ELK Asia Pacific J* (2017) 978–93–85537–06–6.
13. Kanwarjeet S, Satpal S (2019) Fabrication and investigation of Co-based and CeO2-modified microwave coatings. Prot Met Phys Chem 55(2):352–358
14. Kanwarjeet S, Satpal S, Development of Ni-based and CeO2 modified coatings by microwave heating, materials and manufacturing processes (2016)
15. Yashwant Koli, N. Yuvaraj, S. Aravindan & Vipin, Multi-response Mathematical Modeling for Prediction of Weld Bead Geometry of AA6061-T6 Using Response Surface Methodology Transactions of the Indian Institute of Metals (2020) volume 73, pages 645–666
16. M. Pal, V. Kumar, S. Sehgal, H. Kumar, K. Saxena, A. Bagha, Microwave hybrid heating based optimized joining of SS304/SS316, Materials and Manufacturing Processes, (2020) DOI: https://doi.org/10.1080/10426914.2020.1854469

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Physica Scripta

PAPER

Temperature dependency and linearity assessment of dual-metal gate stack junctionless accumulation-mode cylindrical surrounding gate (DMGS-JAM-CSG) MOSFET

To cite this article: Sumedha Gupta *et al* 2021 *Phys. Scr.* **96** 124055

View the article online for updates and enhancements.

You may also like

- A 2-D compact DC model for engineered nanowire JAM-MOSFETs valid for all operating regimes
Kamalaksha Baral, Prince Kr Singh, Sanjay Kumar et al.
- Conical surrounding gate MOSFET: a possibility in gate-all-around family
B Jena, B S Ramkrishna, S Dash et al.
- The Black Hole in the Most Massive Ultracompact Dwarf Galaxy M59-UCD3
Christopher P. Ahn, Anil C. Seth, Michele Cappellari et al.

This content was downloaded from IP address 14.139.251.100 on 06/12/2021 at 06:53

Physica Scripta

PAPER

Temperature dependency and linearity assessment of dual-metal gate stack junctionless accumulation-mode cylindrical surrounding gate (DMGS-JAM-CSG) MOSFET

Sumedha Gupta[1,*], **Neeta Pandey**[1] and **R S Gupta**[2]

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Shahbad Daulatpur, Delhi 110042, India
[2] Department of Electronics and Communication Engineering, Maharaja Agrasen Institute of Technology, Rohini, Delhi 110086, India
[*] Author to whom any correspondence should be addressed.

E-mail: sumedhagpt5@gmail.com, neetapandey@dce.ac.in and rsgupta1943@gmail.com

Keywords: cylindrical surrounding gate, dual gates, junctionless MOSFETs, linearity, temperature

RECEIVED
8 July 2021

REVISED
22 October 2021

ACCEPTED FOR PUBLICATION
27 October 2021

PUBLISHED
8 November 2021

Abstract

A physics-based temperature-dependent analytical model for a Dual- Metal Gate Stack Junctionless Accumulation-Mode Cylindrical Surrounding Gate (DMGS-JAM-CSG) MOSFET has been presented in this paper by solving the 2D Poisson's equation utilizing the suitable boundary conditions. The device performance is examined at various temperatures (T = 100K, 300K and 500K) by observing several parameters like potential, electric field, electron concentration and electron velocity. The obtained outcomes are also contrasted with the outcomes acquired after simulation. Further, the device is also analyzed for its analog performance. Lastly, to check the relevancy of the device for RFIC applications, linearity assessment is performed by evaluating its several figure of merits (FOMs) like second-order and third-order Voltage Intercept Points (VIP2, VIP3), third-order Intercept Input Power (IIP3), third-order Intermodulation Distortion (IMD3) and various higher order transconductances. The obtained results are contrasted with an analogous device and it is discovered that DMGS-JAM- CSG MOSFET device acquires better electrical and linearity characteristics. ATLAS 3D device simulator has been utilized to conduct the simulations.

1. Introduction

Nowadays, MOSFETs are approaching nanometer scale and due to scaling, a problem called as Short Channel Effects (SCEs) arises. SCEs brings about the degradation in the performance of the MOSFETs [1, 2]. To overcome this problem, several multigate MOSFET topologies, that is, Double-Gate (DG), Tri-Gate (TG), Cylindrical Surrounding Gate (CSG) MOSFETs [3–5] were introduced. These topologies provide improved behavior when compared with the conventional one. CSG is the finest among all these because of the presence of gate all over the silicon film, which offers extended gate control. But it also holds a drawback of development of sharp source-drain junctions. For elimination of the above-stated problem, Junctionless Transistor (JLT) was introduced in 2010 [6]. In JLT, uniform doping is done all over the MOSFET, because of which no source-drain junctions are present. This makes the fabrication process simpler and less expensive. But JLT also possesses lower drain current and transconductance because of the Carrier Mobility Degradation (CMD) [7]. So, for improved drain current and transconductance, another MOSFET called Junctionless Accumulation Mode (JAM) MOSFET came. In this MOSFET, non-uniform doping is done, i.e., channel is slightly less doped as compared to source/drain [8, 9]. JAM MOSFET possesses better electrical characteristics, more conductivity and less SCEs in contrast to JLT. To decrease the SCEs further, double gates and a gate stack [10–12] are introduced to JAM MOSFET, proposing a new structure called Dual- Metal Gate Stack (DMGS) JAM- CSG MOSFET [13].

When MOSFET is utilized for the applications including space communication, satellite communication, infrared detectors, etc, temperature plays a significant role. As, the change in temperature varies the

© 2021 IOP Publishing Ltd

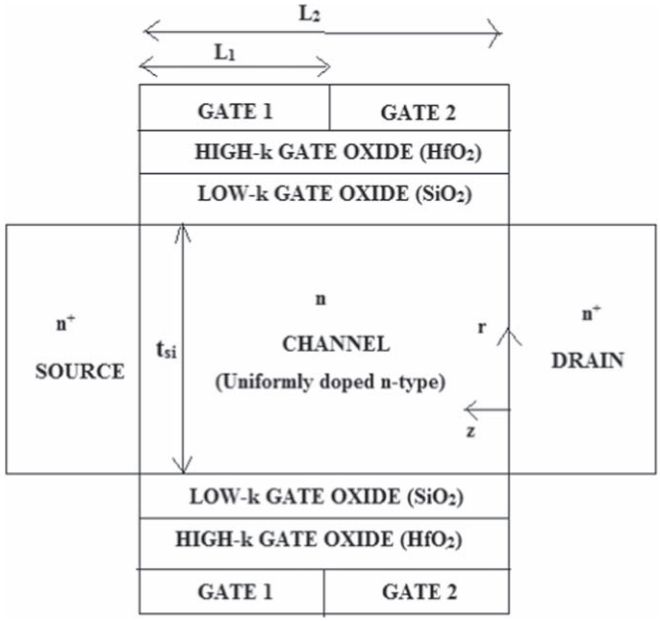


Figure 1. 2D view of DMGS-JAM-CSG MOSFET.

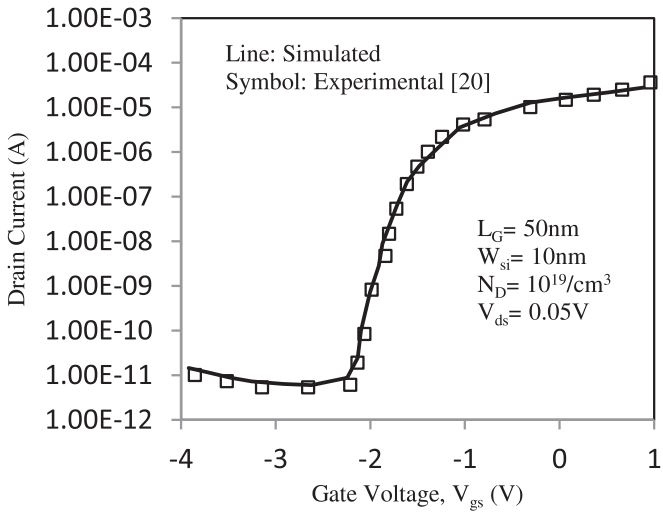


Figure 2. Calibration result.

performance of the MOSFET, therefore, it becomes important to examine the impact of temperature upon the various characteristic aspects of the MOSFET [14]. Therefore, to study the influence of temperature [15] on our structure i.e., DMDG-JAM-CSG MOSFET, we have developed this structure's temperature- dependent physics-based analytical model using the applicable boundary conditions in this paper. Also, we have examined the improved performance of DMGS-JAM-CSG MOSFET over JAM-CSG MOSFET by analyzing its various analog performance parameters.

In addition to the analog performance analysis, we have also performed the linearity assessment of the device. Nowadays, semiconductors are being utilized for numerous RF applications [16–18] including wireless networks, radio broadcasting, navigation, etc For improved performance of these applications, they require intermodulation and higher order harmonics to be minimum because if distortion is present, then it will lead to the generation of undesirable (noise) signals which may alter the output signal [17]. Therefore, linearity analysis is very essential for RF systems. Various techniques are there to confirm high linearity. Determination of linearity characteristics is also performed through experiments but this is a high cost and time consuming affair. Another method which is very effective and less time consuming is by evaluating several FOMs. FOM includes VIP2, VIP3, IIP3, IMD3 and higher order transconductances. So, we have performed the linearity analysis of the device by determining these various FOMs.

Table 1. Structural parameters used.

Parameter	*Jam-csg*	*Dmgs-jam-csg*
Channel length	50 nm	50 nm
Channel doping	10^{18} cm^{-3}	10^{18} cm^{-3}
source/drain doping	10^{20} cm^{-3}	$10^{20}/cm^{3}$
gate oxide material	SiO_2	SiO_2 and HfO_2
gate stack oxide thickness	SiO_2–2 nm	SiO_2–1 nmHfO_2–1nm
silicon pillar thickness	20 nm	20 nm
work- function	$\Phi = 4.8$ eV	$\phi_1 = 5.0$ eV (Rhodium) $\phi_2 = 4.8$ eV (Ruthenium)

Table 2. Physical models used.

Physical models	*Description*
Recombination Model	The Shockley– Read– Hall recombination model (with carrier lifetime 1×10^{-7} s) is included in simulation to incorporate minority recombination effects. The SRH model considers carrier lifetimes.
Concentration Dependent Model	Considers SRH recombination along with their lifetimes. CONMOB relates the low-field mobility to the impurity concentration.
Band To Band Tunneling (BTBT) Model	For direct transitions. Required with very high fields. BTBT model inculcates tunnelling effect of the charge carriers.
Mobility Model	Field-Dependent Mobility Model (FLDMOB) has been used in the analysis to accommodate the velocity saturation effect.
Statistics	Boltzmann Model is used because it considers the Carrier Statistics.
Methods	The Newton and Gummel methods are simultaneously invoked for the numerical solution.

For performing the simulation, ATLAS-3D device simulator has been employed with the following models: Recombination Model for considering minority recombination effects, Concentration Dependent Mobility Model for relating the low-field mobility to the impurity concentration, Band To Band Tunneling Model for inculcating tunnelling effect, Boltzmann Model is used to consider the Carrier Statistics, Newton and Gummel methods for the numerical solution.

2. Device description, simulation methodology and calibration

2.1. Device description

Figure 1 represents the 2D structural view of the device DMGS-JAM-CSG MOSFET. The channel is slightly less doped in comparison to source/drain with n-type doping. It consists of dual gates (gate1 and gate2) of different workfunctions, ϕ_1 and ϕ_2 respectively such that $\phi_1 > \phi_2$. Also, a high-k material HfO_2 along with SiO_2 is also present as a gate stack. Here, thickness of SiO_2, t_{ox} is 1nm and thickness of HfO_2, t_{hk} is 1nm. Thus, the effective oxide thickness, t_{oxeff} is calculated as:

$$t_{oxeff} = t_{ox} + \frac{\varepsilon_{ox} \cdot t_{hk}}{\varepsilon_{hk}} \quad (1)$$

where, ε_{ox}, ε_{hk} are the permittivities of SiO_2 and HfO_2 respectively and are equal to 3.9 and 25 respectively.

Table 1 summarizes the various structural parameters employed for the device DMDG-JAM-CSG MOSFET.

2.2. Models utilized

For performing the device simulation, DMDG-JAM-CSG MOSFET with SiO_2 and HfO_2 present as the low-k and high-k gate oxide respectively, ATLAS 3D device simulator has been employed. Table 2 enumerates several physical models and methods utilized while performing simulation on ATLAS 3D simulator [19].

2.3. Device calibration

Figure 2 shows the calibration of the model parameters utilized during the simulation process with the experimental work carried out by Choi *et al* [20], considering the same cross-sectional area. Extraction of the experimental data [20] has been carried out by utilizing the Graph Digitizer tool and then using the extracted data, the experimental graph has been plotted. Also, from the figure 2, it is noticed that both the simulated and experimental results are corresponding well with each other.

3. Analytical model

For investigating the effect of temperature variation upon various device's characteristics, potential distribution model of the device is developed using the 2D Poisson's equation, stated as:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi_i(r,\,z,\,T)}{\partial r}\right)+\frac{\partial^2\psi_i(r,\,z,\,T)}{\partial z^2}=-\frac{qN_D}{\varepsilon_{si}},\quad i=1,\,2 \tag{2}$$

where, T be the temperature measured in Kelvin (K), $\psi_i(r,\,z,\,T)$ represents the potential distribution, r and z are the distances across the radial and vertical axis respectively, q is the electronic charge, N_D be the channel doping and ε_{si} is the permittivity of SiO_2.

Using parabolic potential approximation (PPA) technique, the potential is calculated as:

$$\psi_i(r,\,z,\,T)=P_{i0}(z)+\,P_{i1}(z)\,r+\,P_{i2}(z)\,r^2,\quad 0\leqslant z\leqslant L \tag{3}$$

Solution of the above potential equation is acquired by administering the subsequent boundary conditions:

(1) Center potential

$$\psi_i(r=0,\,z,\,T)=\psi_c \tag{4}$$

(2) Field is zero at the center:

$$\frac{\partial\psi_i(r,\,z,\,T)}{\partial r}|_{\,r=0}=0 \tag{5}$$

(3) Surface potential:

$$\psi_i\left(r=\frac{t_{si}}{2},\,z,\,T\right)=\psi_s\left(\frac{t_{si}}{2},\,z,\,T\right) \tag{6}$$

(4) Surface field:

$$\frac{\partial\psi_i\left(\frac{t_{si}}{2},\,z,\,T\right)}{\partial r}=\frac{C_{oxeff}}{\varepsilon_{si}}\left[V_{gs}-V_{fb}-\psi_i\left(\frac{t_{si}}{2},\,z,\,T\right)\right] \tag{7}$$

(5) Potential present towards source side:

$$\psi_1(r,\,0,\,T)=V_{bi} \tag{8}$$

(6) Potential present towards drain side:

$$\psi_2(r,\,L_2,\,T)=V_{bi}+V_{ds} \tag{9}$$

(7) Potential is continous at the interface of both the gates:

$$\psi_{s_1}(r,\,L_1,\,T)=\psi_{s_2}(r,\,L_1,\,T) \tag{10}$$

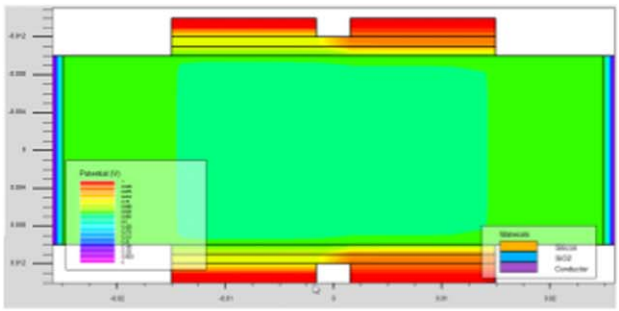
(8) Continuity of electric field at the interface:

$$\frac{\partial\psi_{s_1(r,L_1,T)}}{\partial r}=\frac{\partial\psi_{s_2(r,L_1,T)}}{\partial r} \tag{11}$$

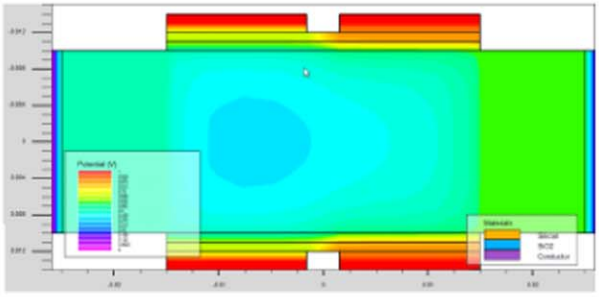
where

$$C_{oxeff}=\frac{2\varepsilon_{ox}}{t_{si}\,\ln\left[1+\left(\frac{2t_{oxeff}}{t_{si}}\right)\right]} \tag{12}$$

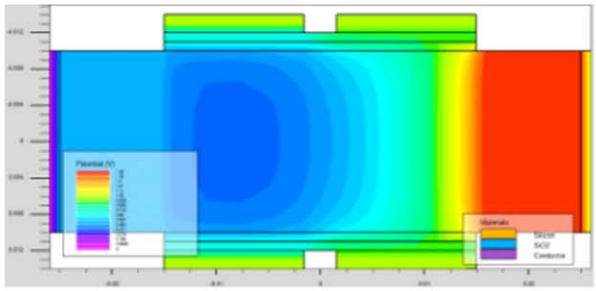
V_{bi} represents the built-in potential; t_{si} be the width of the silicon film.



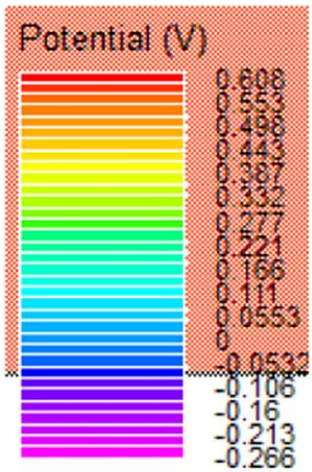
(a)



(b)

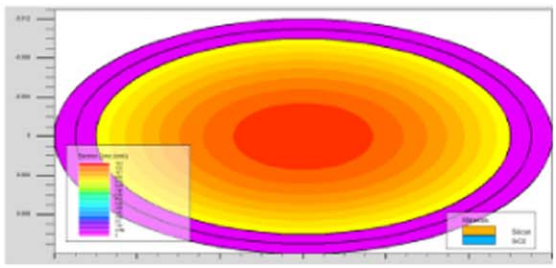


(c)

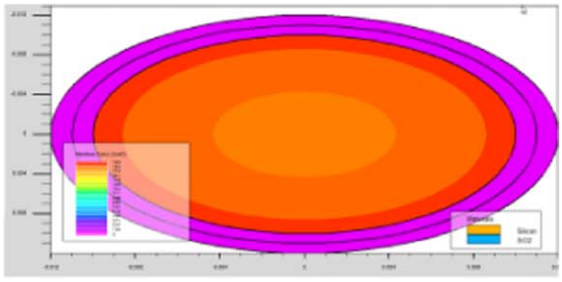


(d)

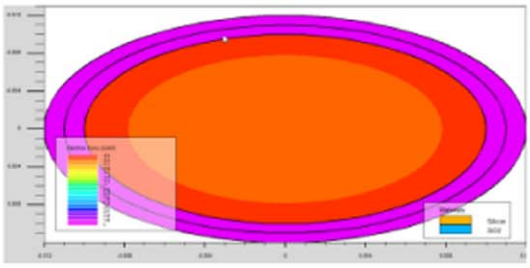
Figure 3. Potential contour plot for DMGS-JAM-CSG MOSFET at $V_{gs} = 1.0V$ and $V_{ds} = 0.1V$ for various temperatures (a) 100K (b) 300K (c) 500K and (d) color coding scale.



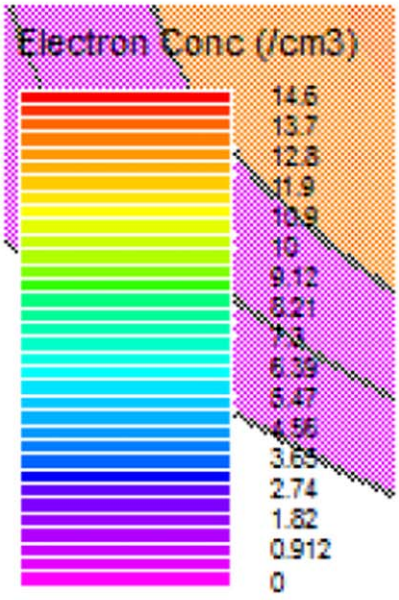
(a)



(b)



(c)



(d)

Figure 4. Contour plot for electron concentration at the channel's surface at $V_{gs} = 1.0V$ and $V_{ds} = 0.1V$ for various temperatures (a) 100K (b) 300K (c) 500K and (d) color coding scale.

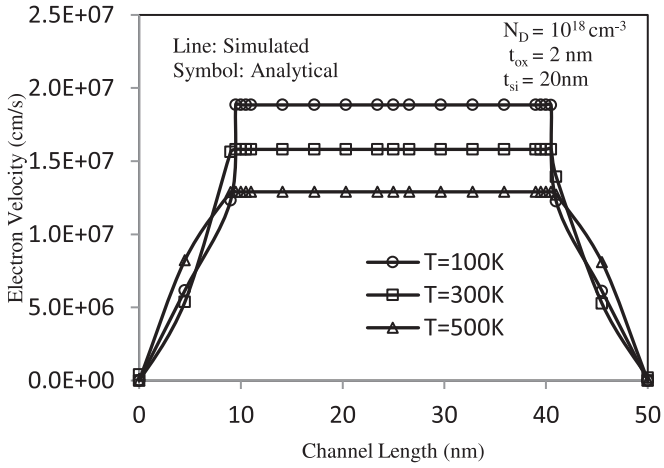


Figure 5. Electron velocity for DMGS-JAM-CSG MOSFET at T = 100K, 300K and 500K.

On applying the various boundary conditions, solution of center potential, $\psi_{ci}(r,\ z,\ T)$ was obtained:

$$\psi_{ci}(r,\ z,\ T) = \begin{cases} \psi_{c1}(r,\ z,\ T)\ \ for\ 0 \leqslant z \leqslant L_1 \\ \psi_{c2}(r,\ z,\ T) for\ L_1 \leqslant z \leqslant L_2 \end{cases} \tag{13}$$

where,

$$\psi_{C1}(z) = A\ e^{z/\lambda} + Be^{-z/\lambda} + \eta_1 \tag{14}$$

$$\psi_{C2}(z) = Ce^{z/\lambda} + De^{-z/\lambda} + \eta_2 \tag{15}$$

Coefficients A, B, C, D are obtained using the given boundary conditions [21].

Temperature affects the parameters and performance of the device [22]. Various temperature sensitive parameters are:

Band Gap Energy (E_g):

$$E_g(T) = Eg(0) - \frac{\alpha T^2}{T + \beta} \tag{16}$$

where, $Eg(0) = 1.166\ eV$, $\alpha = 4.73\ x10^{-4} eV/K$, $\beta = 636K$.

Intrinsic carrier concentration (n_i):

$$n_i(T) = 1.706 \times 10^{25}\left(\frac{T}{300}\right)^{3/2} exp\left(\frac{-qE_g(T)}{kT}\right) \tag{17}$$

Permittivity of silicon:

$$\varepsilon_{si}(T) = 11.4 + (1 + 1.2 \times 10^{-4}T) \tag{18}$$

Fermi Potential:

$$\phi_f(T) = \frac{kT}{q} ln\left(\frac{N_D}{n_i(T)}\right) \tag{19}$$

Flat- band voltage:

$$V_{fb}(T) = \phi_m - \phi_s(T) \tag{20}$$

where, surface potential $\phi_s(T)$ is given as:

$$\phi_s(T) = \frac{E_g(T)}{2} + \chi - q\phi_f(T) \tag{21}$$

Another important parameter affected due to temperature is mobility. Caughey and Thomas function [23] for modelling mobility, considering the effect of temperature is expressed as:

$$\mu_{Caug} = \mu_1 + \frac{\mu_2\left(\frac{T}{300}\right)^{\beta} - \mu_1}{1 + \left(\frac{T}{300}\right)^{\gamma}\left(\frac{N_a}{N_{crit}}\right)^{\delta}} \tag{22}$$

where, $\mu_1 = 55.24\ cm^2\ Vs^{-1}$, $\mu_2 = 240\ cm^2\ Vs^{-1}$, $\beta = -2.3$, $\gamma = -3.8$, $N_{crit} = 1.072 \times 10^{17}\ cm^{-3}$.

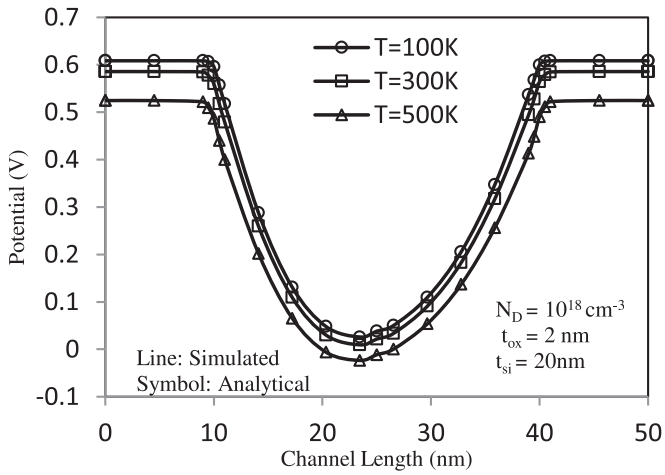


Figure 6. Potential variation for DMGS-JAM-CSG MOSFET for different temperatures (T = 100K, 300K and 500K).

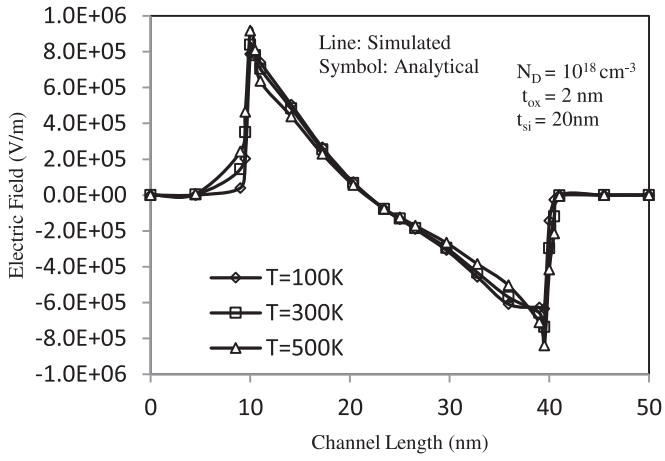


Figure 7. Electric field variation for DMGS-JAM-CSG MOSFET for different temperatures (T = 100K, 300K and 500K).

Electric Field distribution is given as:

$$E_i(r,\, z,\, T) = -\frac{d\psi_i(r,\, z,\, T)}{dz} \tag{23}$$

4. Results and discussion

4.1. DMGS-JAM-CSG mosfet model

Figures 3(a)–(c) displays the contour plots of the potential for DMGS-JAM-CSG MOSFET with $V_{gs} = 1.0V$ and $V_{ds} = 0.1V$ at T = 100K, 300K and 500K respectively. From the figure, it is seen that as the temperature is increasing from 100K to 500K, the potential is getting lowered down. With the rise in temperature, the intrinsic carrier concentration increases which results in the reduction of the potential across the channel [24]. This is due to increase in the mobility of the carriers because of the Fermi level shift towards the bandgap.

Contour plots for electron concentration for DMGS-JAM-CSG MOSFET when $V_{gs} = 1.0V$ and $V_{ds} = 0.1V$ at (a) T = 100K (b) T = 300K and (b) T = 500K is depicted in figures 4(a)–(c) respectively. From the figures, it is noted that when temperature is increasing from 100K to 500K, the concentration of the electrons is also increasing [24]. The increase in temperature brings about movement of Fermi level towards the bandgap. This results in inflation in the mobility of the carriers and thus, increment in electron concentration.

Figure 5 depicts the electron velocity for DMGS-JAM-CSG MOSFET across the channel length at distinct temperatures (T = 100K, 300K and 500K). It is noted from the figure 5 that as the temperature is increased, the electron velocity decreases. For T = 100K, the electron velocity reaches the maximum value of 1.88E + 07 cm s^{-1} and for T = 500K, it reaches 1.29E + 07 cm s^{-1}. The reason behind this is that the mobility of the electrons

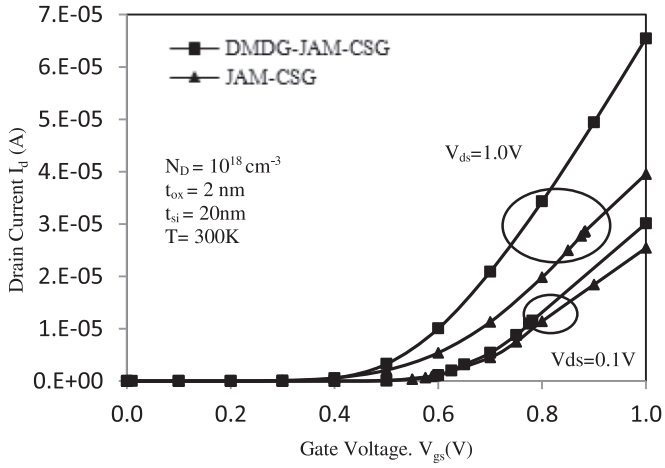


Figure 8. Transfer characteristics (I_d versus V_{gs}).

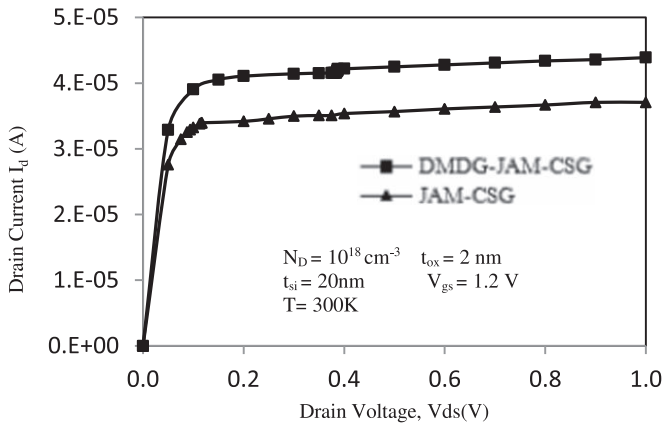


Figure 9. Output characteristics (I_d versus Vds).

increases due to the increment in temperature. Thereby, the number of collisions of electrons increases, leading towards the decrease of electron velocity.

Variation of the potential for DMGS-JAM-CSG MOSFET across the channel at various temperatures (T = 100K, 300K and 500K) is depicted in figure 6. From the figure 6, it is discovered that as the temperature is incremented, the center potential is lowered down. The potential is lowered down from 25mV (at T = 100K) to −23mV (at T = 500K). This lowering down of potential is owing to the inflation in the carrier concentration with the increment in temperature. This reduction in potential indicates lowering down of the threshold voltage, which makes the device to turn ON faster.

Electric field along the channel at distinct temperatures (T = 100K, 300K and 500K) for DMGS-JAM-CSG MOSFET is depicted in figure 7. It is observed that as the increment in the temperature takes place, an elevation in the electric field occurs towards the source side and decrease towards the drain side. This leads to minimized SCEs and Hot Carrier Effects. The reason behind this is the increment in the amount of charge carriers, when there is a rise in temperature.

4.2. Comparison with analogous device

Transfer characteristics (I_d versus V_{gs}) for DMDG-JAM-CSG and JAM-CSG MOSFET at V_{ds} = 0.1V and 1.0V is shown in figure 8 [25, 26]. From the figure, we can make out that DMGS-JAM-CSG MOSFET owns better transfer characteristics. Further, from the figure 8, it is visible that when V_{ds} is changed from 0.1V to 1.0 V, threshold voltage is getting lowered down. This indicates the DIBL effect. Output characteristics (I_d versus V_{ds}) for DMGS-JAM-CSG and JAM-CSG at V_{gs} = 1.2 V is depicted in figure 9. This figure shows that the DMGS-JAM-CSG MOSFET obtains higher drain current wrt JAM-CSG MOSFET. The reason behind this increased drain current is that when gate voltage is applied, an electric field is produced inside the channel and when this applied voltage overruns the flat band voltage, the charge carriers starts accumulating which makes the current

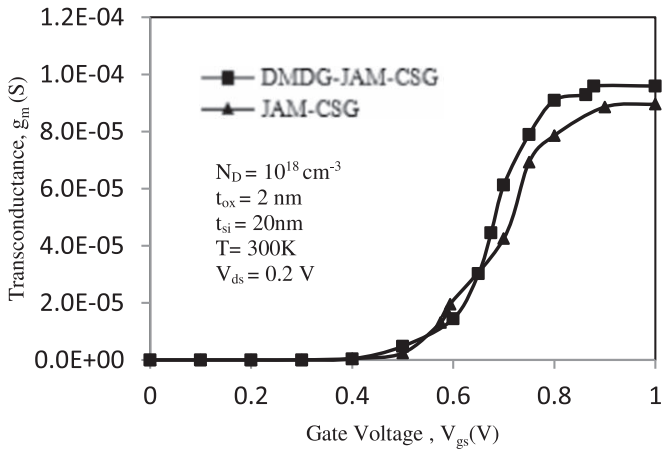


Figure 10. Transconductance (g_m) with V_{gs}.

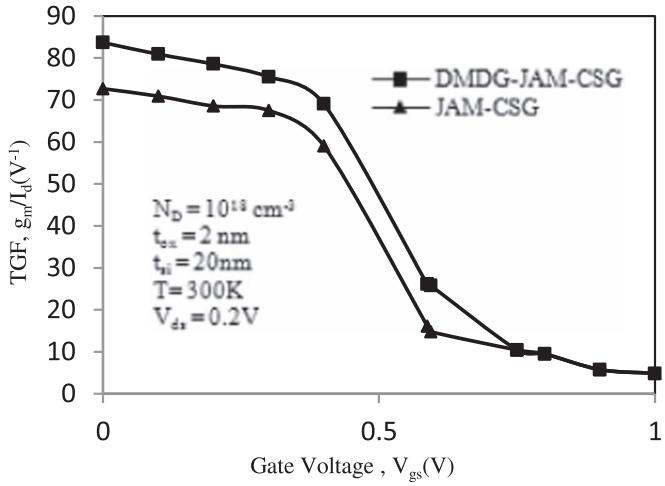


Figure 11. TGF wrt V_{gs}.

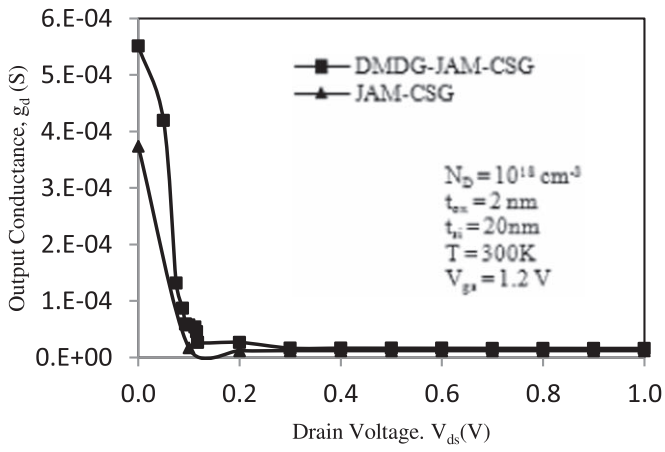


Figure 12. Output Conductance.

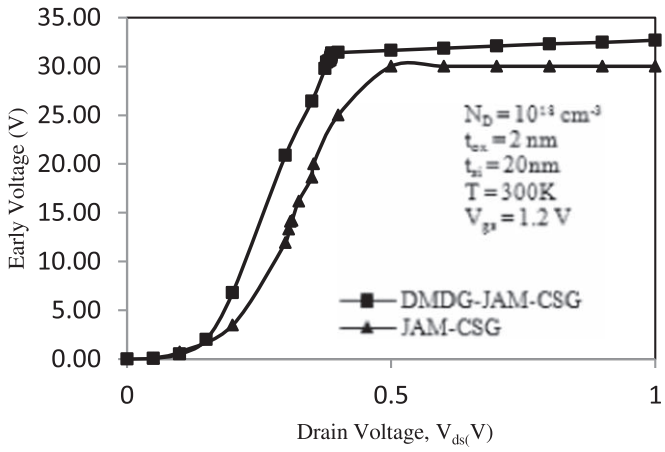


Figure13. Early voltage.

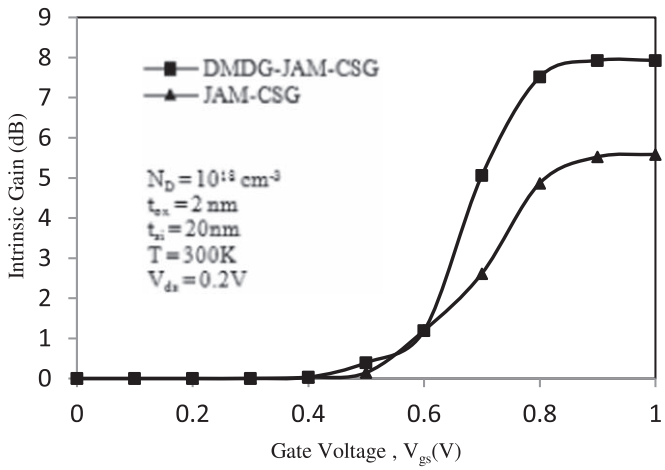


Figure 14. Intrinsic Gain.

to pass through the center of the channel. DMGS-JAM-CSG MOSFET possesses increased gate control over the channel, thereby resulting in faster accumulation of the charge carriers causing high flow of drain current.

Transconductance (g_m) is another vital parameter given by [27]:

$$g_m = (\Delta I_d / \Delta V_{gs})|_{V_{ds}=constant} \tag{24}$$

Figure 10 represents the g_m wrt V_{gs} at $V_{ds} = 0.2$ V. It is noted that this parameter is higher for the DMDG-JAM-CSG MOSFET and therefore, it can be used for applications requiring high gain amplification. Next parameter is Transconductance Generation Factor (TGF) or device efficiency, given by [28]:

$$TGF = g_m / I_d \tag{25}$$

TGF wrt gate voltage is indicated in figure 11. From the figure 11, higher TGF is seen for DMGS-JAM-CSG MOSFET, signifying its capability to amplify a signal more effectively. Thus, DMGS-JAM-CSG MOSFET proves to be more suitable for high amplification applications.

Output conductance is given as [27]:

$$g_d = (\Delta I_d / \Delta V_{ds})|_{V_{gs}=constant} \tag{26}$$

Figure 12 represents the output conductance, g_d with V_{ds} for both the devices. Output conductance is more for DMGS-JAM-CSG MOSFET in comparison to JAM-CSG MOSFET. Early voltage is also major parameter for a MOSFET. It refers to the immunity of the MOSFET towards the Channel Length Modulation (CLM) effect and is stated as [28]:

$$V_A = I_d / g_d \tag{27}$$

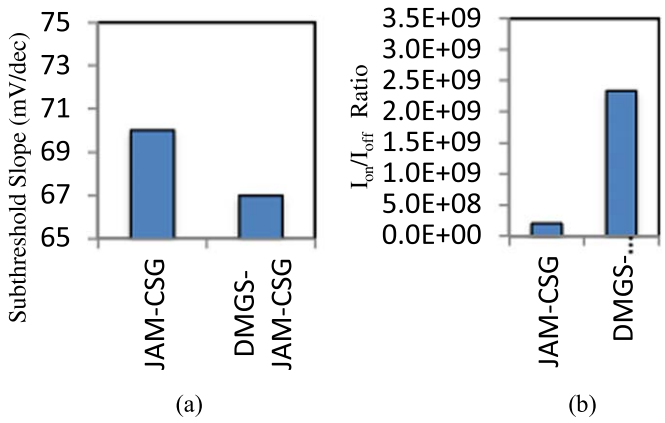


Figure 15. (a) Subthreshold Slope and (b). I_{on}/I_{off} ratio.

Early voltage wrt V_{ds} is depicted in figure 13. For DMGS-JAM-CSG MOSFET, the increase in the drain current is more as compared to the decrease in the output conductance. Therefore, from the figure 13 also, it can be seen that early voltage for DMGS-JAM-CSG MOSFET has been improved and thus is more immune towards CLM effect, which is dominant in short-channel MOSFETs.

Intrinsic gain (A_v) is stated as [28]:

$$A_v = g_m/g_d \tag{28}$$

Intrinsic gain (A_v)wrt V_{gs} is shown in figure 14. DMGS-JAM-CSG MOSFET is having higher intrinsic gain than JAM-CSG MOSFET due to the higher transconductance and early voltage [26]. Thus, DMGS-JAM-CSG MOSFET is more relevant for applications requiring high-gain amplification.

Subthreshold slope (SS) [29] is given by:

$$SS = 2.3\ V_T \left[\frac{d\psi(r,\ z)_{\min}}{dV_{GS}}\right]^{-1} \tag{29}$$

here, V_T represents the threshold voltage, $\psi(r,\ z)_{\min}$ represents the minimum potential distribution of MOSFET. SS for both the MOSFETs is depicted in figure 15(a). Better subthreshold slope can be seen for DMGS-JAM-CSG MOSFET from the figure 15(a) as compared to the JAM-CSG MOSFET.

I_{on}/I_{off} ratio defines the digital performance of the MOSFET and is given as:

$$\frac{I_{on}}{I_{off}} = \frac{I_d(on)\ at\ V_{gs} = 1.0V}{I_d(off)\ at\ V_{gs} = 0.0V} \tag{30}$$

I_{on}/I_{off} ratio is depicted in figure 15(b) for both the devices. From the figure 15(b), DMGS-JAM-CSG MOSFET is noticed to be having higher I_{on}/I_{off} ratio, thereby making it more useful for high- switching applications. This also indicates lower leakage current for DMGS-JAM-CSG MOSFET. Thus, all the above results infers that the DMGS-JAM-CSG MOSFET possesses overall better performance and enhanced capability for digital applications as well when contrasted with JAM-CSG MOSFET, thus verifying the excellence of the device.

4.3. Linearity assessment

Minimum signal distortion is a very important requirement for the device for performing various analog/ RF applications with more accuracy. For this, linearity and various intermodulation parameters are evaluated [17]. These parameters are higher-order transconductances (g_{m2} and g_{m3}) and several FOMs (VIP2, VIP3, IIP3 and IMD3) and are given by:

$$\boldsymbol{g_{mn}} = \frac{\partial^n \boldsymbol{I_d}}{\partial \boldsymbol{V_{gs}}^n} \tag{31}$$

$$\text{VIP2} = 4 \times \frac{g_{m1}}{g_{m2}} \tag{32}$$

$$\text{VIP3} = \sqrt{24 \times \frac{g_{m1}}{g_{m3}}} \tag{33}$$

$$\text{IIP3} = \frac{2}{3} \times \frac{g_{m1}}{g_{m3}\ \text{x}\ R_s} \tag{34}$$

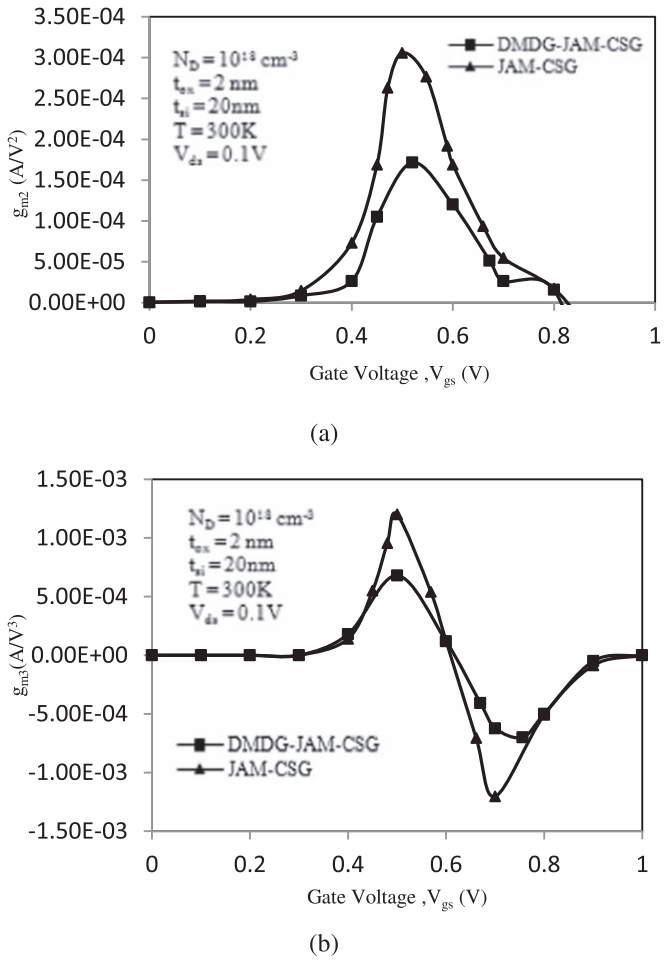


Figure 16. Higher-order transconductances (a) g_{m2} wrt gate voltage (b) g_{m3} wrt gate voltage.

$$\text{IMD3} = \left[\frac{9}{2} \times (\text{VIP3})^2 \times g_{m3} \right]^2 \times R_s \tag{35}$$

where, $R_s = 50\,\Omega$ during RF designs [16].

Figures 16(a) and (b) represents the higher order transconductances, g_{m2} and g_{m3} variation against the V_{gs}, respectively. These coefficients should be ideally zero or of minimal value for the device to be of high-linearity. From the figure, it can be seen that DMGS-JAM-CSG MOSFET possesses lower higher- order transconductances as compared to JAM-CSG MOSFET, which indicates the lowered harmonic distortions and improved linearity behaviour. The reason is the increased gate control provided in the case of DMGS-JAM-CSG MOSFET.

Figures 17(a) to (d) represents the numerous FOMs (VIP2, VIP3, IIP3 and IMD3). VIP2 (Second-order Voltage Intercept Point) is the extrapolated input voltage at which first order harmonic voltage is equal to the second order harmonic voltage. Likewise, VIP3 (Third- order Voltage Intercept Point) is the extrapolated input voltage at which first order harmonic is equal to the third order harmonic voltage. For improved linearity of the device, elevated values of VIP2 and VIP3 is preferable. Figures 17(a) represents the VIP2 against V_{gs} and higher value of VIP2 is being observed for DMGS-JAM-CSG as contrasted to the JAM-CSG MOSFET, which indicates its lower distortion and better linearity characteristics over JAM-CSG MOSFET.

Figures 17(b) shows VIP3 against the gate voltage for both the devices under consideration. A peak of VIP3 at lower gate bias for DMGS-JAM-CSG MOSFET is observed from the figure. This peak of VIP3 defines cancellation of 3rd order non-linear coefficient around 2nd order non-linear coefficient by internal feedback of the device. This occurs on account of the lower metal workfunction gate present towards the drain side.

IIP3 (Third-order Intercept Input Power) is the extrapolated input power at which 3rd order intermodulation voltage and the 1st order harmonic voltage are equal. IIP3 for the device must be high for better linearity characteristics [30]. Figures 17(c) shows IIP3 against the gate bias voltage. DMGS-JAM-CSG MOSFET possesses higher value of IIP3 which indicates its better linearity as compared to JAM-CSG MOSFET. This enhancement in IIP3 for DMGS-JAM-CSG MOSFET is due to the dual-metal architecture which provides

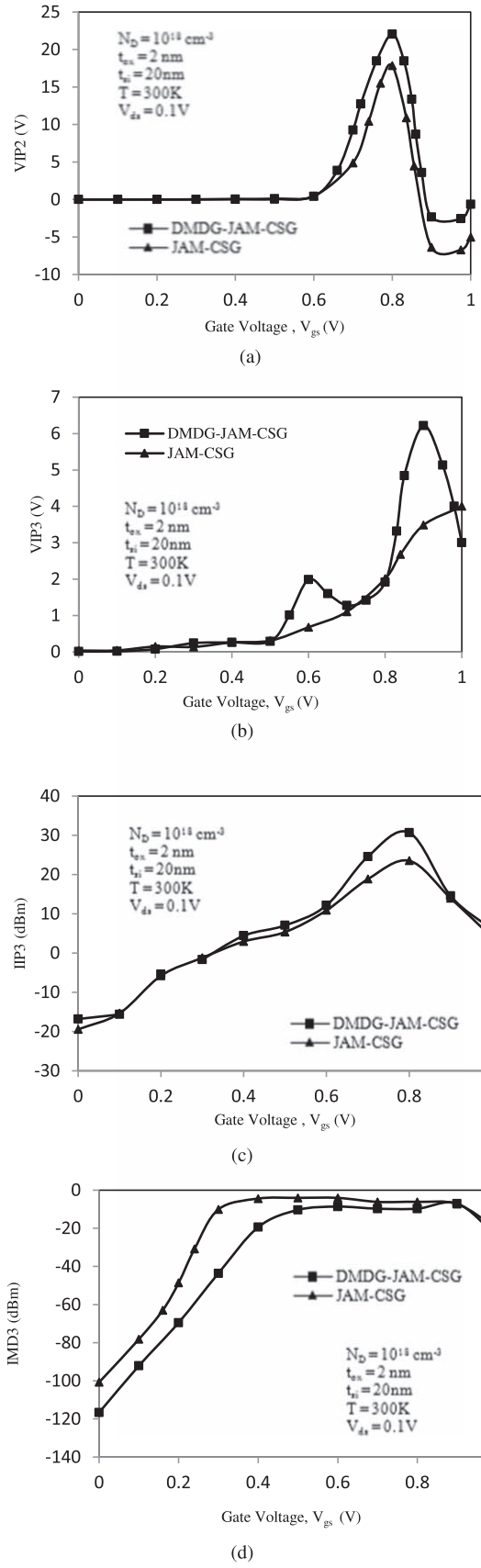


Figure 17. Various FOMs (a) VIP2 (b) VIP3 (c) IIP3 (d) IMD3.

distinct workfunctions for the two metal gates. Thus, providing a step- profile for the device's potential. This higher IIP3 yields better gate control and also provides higher carrier transport over the channel.

Next relevant linearity specification is IMD3 (Third- order Intermodulation Distortion). It is the extrapolated intermodulation power at which the 1st order harmonic power and 3rd order intermodulation harmonic power are equal. Figures 17(d) shows IMD3 for DMGS-JAM-CSG MOSFET and JAM-CSG MOSFET. For any device to be distortionless, this parameter should be low [30]. From the Figures 17(d), we can make out that the DMGS-JAM-CSG MOSFET possesses lower value of IMD3 as compared to JAM-CSG MOSFET. This indicates superior linearity of DMGS-JAM-CSG MOSFET and its higher efficiency.

5. Conclusion

Analysis of the influence of temperature (T = 100K, 300K and 500K) upon DMGS-JAM-CSG MOSFET has been performed in this paper by developing an analytical model and examining its various parameters such as potential, electric field, electron velocity and electron concentration at these temperatures. An improvement is seen in all these parameters as the temperature increases. As the temperature is increasing from 100K to 500K, the potential is lowered by 48 mV and electron velocity is getting lowered down by 59 cm s^{-1}, thus, resulting in reduction of SCEs and HCEs. Also, analytical results agree well with the simulated ones. Further, the device has been examined for its analog performance and higher values for drain current, transconductance, output conductance, early voltage and intrinsic gain is noticed for the DMGS-JAM-CSG MOSFET. Thus, the device stands out to be more applicable for high-frequency and high-gain amplification applications. I_{on}/I_{off} ratio has also increased by 11.6 times in comparison to JAM-CSG MOSFET, thus, the device possesses lower leakage current and more useful for high-switching applications. Linearity metrics g_{m2}, g_{m3}, VIP2, VIP3, IMD3 and IIP3 were also evaluated and found to be more improved for DMGS-JAM-CSG MOSFET, thus, marking the device to be more distortion free and applicable for RFIC applications.

Acknowledgments

Authors are thankful to the Director of Maharaja Agrasen Institute of Technology (MAIT), Delhi for providing the resources needed for accomplishing this work.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

Sumedha Gupta https://orcid.org/0000-0002-1580-5587
R S Gupta https://orcid.org/0000-0003-1031-8621

References

[1] Jimenez D, Saenz J J, Iniguez B, Sune J, Marsal L F and Pallares J 2004 Modeling of nanoscale gate-all-around MOSFETs *IEEE Electron Device Lett.* **25** 314–6
[2] Wang R, Zhuge J, Huang R, Tian Y, Xiao H, Zhang L, Li C, Zhang X and Wang Y 2007 Analog/RF performance of Si nanowire MOSFETs and the impact of process variation *IEEE Trans. Electron Devices* **54** 1288–94
[3] Colinge J P 2007 Multi-gate soi mosfets *Microelectron. Eng.* **84** 2071–6
[4] Sarkar A, Das A K, De S and Sarkar C K 2012 Effect of gate engineering in double-gate MOSFETs for analog/RF applications *Microelectron. J.* **43** 873–82
[5] Wu Y-S and Su P 2008 Sensitivity of gate-all-around nanowire MOSFETs to process variations—A comparison with multigate MOSFETs *IEEE Trans. Electron Devices* **55** 3042–7
[6] Colinge J-P *et al* 2010 Nanowire transistors without junctions *Nat. Nanotechnol.* **5** 225–9
[7] Cho S, Kim K R, Park B-G and Kang M 2011 RF performance and small-signal parameter extraction of junctionless silicon nanowire MOSFETs *IEEE Trans. Electron Devices* **58** 1388–96
[8] Kim T *et al* 2013 First demonstration of junctionless accumulation-mode bulk FinFETs with robust junction isolation *IEEE Electron Device Lett.* **34** 1479–81
[9] Holtij T, Graef M, Hain F M, Kloes A and Iñíguez B 2013 Compact model for short-channel junctionless accumulation mode double gate MOSFETs *IEEE Trans. Electron Devices* **61** 288–99
[10] Chau R, Datta S, Doczy M, Doyle B, Kavalieros J and Metz M 2004 High-/spl kappa//metal-gate stack and its MOSFET characteristics *IEEE Electron Device Lett.* **25** 408–10
[11] Dhiman G, Pourush R and Ghosh P K 2018 Performance analysis of high-κ material gate stack based nanoscale junction less double gate MOSFET *Materials Focus* **7** 259–67

[12] Darwin S and Arun Samuel T S 2020 A holistic approach on Junctionless dual material double gate (DMDG) MOSFET with high k gate stack for low power digital applications *silicon* **12** 393–403
[13] Gupta S, Pandey N and Gupta R S 2020 Investigation of dual-material double gate junctionless accumulation-mode cylindrical gate all around (dmdg-jlam-cgaa) mosfet with high-k gate stack for low power digital applications *2020 IEEE 17th India Council Int. Conf. (INDICON)* pp 1–4 *(Piscataway, NJ)* (IEEE)
[14] Skotnicki T, James A H and King T-J 2005 H-SP Wong, and Frederic Boeuf. 'The end of CMOS scaling: toward the introduction of new materials and structural changes to improve MOSFET performance *IEEE Circuits Devices Mag.* **21** 16–26
[15] Gautam R, Saxena M, Gupta R S and Gupta M 2014 Temperature dependent subthreshold model of long channel GAA MOSFET including localized charges to study variations in its temperature sensitivity *Microelectron. Reliab.* **54** 37–43
[16] Saha R, Bhowmick B and Baishya S 2018 Temperature effect on RF/analog and linearity parameters in DMG FinFET *Appl. Phys. A* **124** 1–10
[17] Pratap Y, Haldar S, Gupta R S and Gupta M 2014 Performance evaluation and reliability issues of junctionless CSG MOSFET for RFIC design *IEEE Trans. Device Mater. Reliab.* **14** 418–25
[18] Kaya S and Ma W 2004 Optimization of RF linearity in DG-MOSFETs *IEEE Electron Device Lett.* **25** 308–10
[19] ATLAS, Device Simulator 2015 (Santa Clara: Silvaco International)
[20] Choi S-J, Moon D-I, Kim S, Duarte J P and Choi Y-K 2010 Sensitivity of threshold voltage to nanowire width variation in junctionless transistors *IEEE Electron Device Lett.* **32** 125–7
[21] Gupta S, Pandey N and Gupta R S 2021 Analytical modeling of dual-metal gate stack engineered junctionless accumulation-mode cylindrical surrounding gate (DMGSE-JAM-CSG) MOSFET *Appl. Phys. A* **127** 520
[22] Goel A, Sonam Rewari S V and Gupta R S 2019 Temperature-dependent gate-induced drain leakages assessment of dual-metal nanowire field-effect transistor—analytical model *IEEE Trans. Electron Devices* **66** 2437–45
[23] Caughey D, Mo and Thomas R E 1967 Carrier mobilities in silicon empirically related to doping and field *Proc. IEEE* **55** 2192–3
[24] Hasanuzzaman M, Syed K I and Leon M T 2004 Effects of temperature variation (300–600 K) in MOSFET modeling in 6H–silicon carbide *Solid-State Electronics* **48** 125–32
[25] Gnudi A, Reggiani S, Gnani E and Baccarani G 2012 Analysis of threshold voltage variability due to random dopant fluctuations in junctionless FETs *IEEE Electron Device Lett.* **33** 336–8
[26] Duarte J P, Choi S-J and Choi Y-K 2011 A full-range drain current model for double-gate junctionless transistors *IEEE transactions on electron devices* **58.12** 4219–25
[27] Arora N D 2012 *MOSFET Models for VLSI Circuit Simulation: Theory and Practice* (Berlin: Springer)
[28] Rewari S, Nath Vandana, SHaldar S S Deswal and Gupta R S 2016 Improved analog and AC performance with increased noise immunity using nanotube junctionless field effect transistor (NJLFET) *Appl. Phys. A* **122** 1–10
[29] Colinge J-P 1986 Subthreshold slope of thin-film SOI MOSFET's *IEEE Electron Device Lett.* **7** 244–6
[30] Wang T, Liang L and Chengkuo L 2013 A junctionless gate-all-around silicon nanowire FET of high linearity and its potential applications *IEEE Electron Device Lett.* **34** 478–80

PAPER

Evaluation of tribological, oxidation and corrosion behavior of HVOF sprayed sustainable temperature-dependent carbon coating

To cite this article: Ankit Tyagi *et al* 2021 *Surf. Topogr.: Metrol. Prop.* **9** 045024

View the article online for updates and enhancements.

You may also like

- Mechanical and Tribological behaviour of velocity oxygen fuel thermal spray coating: A Review
Shivani Jha and R.S. Mishra
- Microstructural characterization and electrochemical corrosion behaviour of HVOF sprayed Alloy718-nanoAl_2O_3 composite coatings
Hitesh Vasudev, Gaurav Prashar, Lalit Thakur et al.
- Development of dense corrosion resistant coatings by an improved HVOF spraying process
Jin Kawakita, Seiji Kuroda, Takeshi Fukushima et al.



IOP ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection–download the first chapter of every title for free.

This content was downloaded from IP address 14.139.251.100 on 09/12/2021 at 10:55

Surface Topography: Metrology and Properties

PAPER

Evaluation of tribological, oxidation and corrosion behavior of HVOF sprayed sustainable temperature-dependent carbon coating

Ankit Tyagi[1,2], **Shubhangi Chourasia**[1,2], **Qasim Murtaza**[2] and **R S Walia**[3]

[1] Department of Mechanical Engineering, SGT University, Haryana, India
[2] Department of Mechanical Engineering, Delhi Technological University, Delhi, India
[3] Department of production & Industrial engineering, PEC, Chandigarh, India

E-mail: tyagiankit10506@yahoo.com

RECEIVED
22 June 2021

REVISED
2 October 2021

ACCEPTED FOR PUBLICATION
27 October 2021

PUBLISHED
10 November 2021

Keywords: oxidation test, corrosion test, wear test, raman spectra test

Abstract

To abate frictional loss, advanced coatings are vital for piston rings, thick low friction environment-friendly carbon coatings have been deposited using HVOF spray (High Velocity Oxy-fuel) technique. The COF (Coefficient of Friction) was found in the range of 0.3 to 0.45 whereas the wearing range was about 45–150 μm during the tested conditions. After wear test at high temperature, the developed coating shows that micro-hardness was increased about 11.5%, 90% reduction in residual stress, COF decreases to about 50%, and wear decreases to about 70% under test conditions. The experimental result of the conducted corrosion test illustrates that the overall measured mass loss per unit area of the as deposited sample was decreased to be about 85.7% at an exposure time of 1 h, 63.6%, at an exposure time of 3 h, and 55% at an exposure time of 5 h.

1. Introduction

Nowadays vital improvements in the tribological performance of IC engines are directly associated with optimization of surface geometry, improvement in lubrication, and also application of unconventional coatings materials. Amid engine constituents, piston rings are of special concern, as 40%–45% of overall energy is lost owed to friction. The tribological performance of piston assembly has a substantial effect concerning power loss, harmful exhaust emissions, and fuel consumption. In literature, enhancement of piston components performance is generally sought over and done with optimization of lubricant regime [1], surface topography [2, 3], and surface modification [4]. Surface modification and engineering are some of the most important approaches in the reduction of friction and wear. In extensive work, surface modification of nano HVOF [5], thermal and plasma spray [6], ceramic or metallic alloying, nitriding [7], hard chrome plating [8] have been recommended to improve tribology in piston rings. Amid these techniques, thermal spray and HVOF deposition are well-adhered coating techniques for tailoring chemical, mechanical, and tribological properties, keeping the geometry of the piston ring unaffected.

Recently, the reciprocating and rotary test is used for showing tribological properties (friction) of numbers of researcher coatings which includes thermally sprayed nanocermets (WC86-Co10Cr4) [9], nanocomposites (NiP-Si_3N_4, TiSiCN), multilayer coatings [10], and carbon coatings. A comparative study between the plasma and flame spraying process shows plasma spraying process offers high particle velocity with low porosity and good microstructural integrity. Further enhancement in tribological and mechanical properties can be achieved by inculpating hard ceramic, diamond, SiC particles, etc in the coating [11, 12]. Plasma sprayed CRC coatings because of higher hardness, excellent wear resistance, and low internal stresses have also attracted more attention up to 700 °C temperature [13–16]. The reason why HVOF coating is the most desirable coating technique which has less than 1% oxide content even for reactive metals like CRC–NiCr and WC and that produces highly adherent coatings and low porosity [17–19]. Taking traditional Chromium coating as standard, friction behavior was enhanced by tungsten doped carbon coatings, while wear resistance was enhanced by bilayer and multilayer coatings. The experimental study of Derelizade *et al* [20] observed that HVOF developed GNPs incorporated on WC-Co coating

© 2021 IOP Publishing Ltd

shows less deformation and improved wear rate. Another study of Tyagi *et al* [21] shows that HVOF developed carbon coating shows improved wear, hardness, residual stress and COF at elevated temperature. Another experimental study shows that HVOF developed coating shows lower wear, lesser mass loss on comparing with APS developed coating [22]. It was also noted the HVOF developed composite coating shows anti-wear properties with higher hardness [23].

In today's world, realistic cost-effective coating, which causes less pollution is in strong demand. The aim of the study focused to develop an environment-friendly coating deposited in the presence of nitrogen, oxygen, and LPG using high-velocity oxy-flame (HVOF). This present work of research scholar ids dedicated to assessed the grain microstructure, surface texture,surface morphology, tribological properties for piston rings, by aid of high-temperature tribometer, at experimental conditions. The wear behavior of the engineered piston rings will be discussed through analysis of wear characteristics between rubbing parts, which will potentially benefit to enhance the tribology of piston assembly in IC engines.

1.1. Need of research

In modern era, piston rings are of special concern, as 40%–45% of overall energy is lost owed to friction. The tribological performance of piston assembly has a substantial effect concerning power loss, harmful exhaust emissions, and fuel consumption. The objective of contemporary study is to produce environment-friendly coating from agricultural waste, which can resist wear, and corrosion.

1.2. Research questions and outcome of present work

The present research will addresses:

1. To deposit Carbon coating for piston ring applications.
2. To study characterization (FESEM (Field Emission Scanning Electron Microscopy) + EDS (Energy Dispersive x-ray Spectroscopy), Raman Spectrum, HRXRD (High resolution x-ray Diffraction)) of as-deposited coating.
3. To evaluate sliding wear and wear mechanism, residual stress, micro-hardness, oxidation and corrosion test results of carbon coating.

2. The experimental procedure

2.1. Coating deposition process

Initially, 70*70*5 mm size mild steel plate were cut and ground using 220–2500 grit paper. After that, the MEC-made PR-9182 model pressure blaster was used to deposit additive layers of SiC. The produced carbon powder was sprayed using HIPOJET-2700M HVOF system (Metallizing Equipment Co. Pvt. Ltd, Jodhpur) at supersonic speed jet. The powder spray system and steel plate were encumbered with a continuous supply of N_2 gases, for the growth of carbon coating. The agricultural waste produced carbon can be used for absorption of chromium, and wastewater treatment [24], also helps in the reduction of harmful gas emissions and removal of chlorine [25]. Table 1 demonstrate the deposition conditions, used to grow environment-friendly HVOF coating from agricultural waste.

Table 1. Details of coating deposition.

No.	Parameters used	Standards
1.	Deposit efficiency (%)	55
2.	Coating Powder	carbon powder
3.	Powder feed rate	4 Kg h^{-1}
4.	Fuel flow rate	10 ± 1 (/h
5.	Surface pre-treatment	Acetone
6.	Carrier gases	N_2
7.	Treatment time	10 min

2.2. Coating characterization and wear test

The JEOL made FESEM attached with EDS (available at IUAC, Delhi,) has used to characterize the surface morphology, elemental as well as microstructural composition of developed. The micro-Raman spectrometer (Renishaw), was used to characterize & confirms the peaks of the element present in the developed coatings. The HRXRD (PANalytical systems) technique was used for crystal and phase identification by using conventional Bragg mode.

The microhardness of self-developed based carbon coating was calculated with the use of Vickers hardness tester. The internal residual stresses are evaluated with the residual stress analyzer (μ-X360, XRD, (Pulstec manufacturer)) on the carbon coatings. During the corrosion test the mass loss of as-deposited carbon coatings after high-temperature wear test was evaluated by using electrochemical process was employed using an electrolytic solution. The mass gain during oxidation test of as-deposited carbon coatings after high-temperature wear test was evaluated and employed using an electric muffle furnace.

The tribometer (DUCOM, DHM850 model, available at DTU) was used to evaluate and analyze COF and wear of carbon coatings. During the tribological test, one cylindrical pin (Ø8*25 mm) made up of steel and one cuboidal plate (70*70*5 mm) having carbon coating was used and samples were prepared according to G 99 standards. The particulars of an experiment for wear test are given in table 2.

Table 2. The wear test experimental details.

S. no.	Parameter	CC1	CC2	CC3	CC4
1	Temp. (C)	100	150	200	250
2	SV ($m\,s^{-1}$)	1	1	1	1
3	Load (N)	50	50	50	50

3. Results and discussion

3.1. Coating characterization

Figures 1(a)–(c) shows surface morphology of coating and at X 250 resolution that carbon powder shows a needle-like structure along with flakes and flatter. The presence of pebbles-like structure was also seen with a variable size of 10–100 μm. At a higher resolution of x1500 and 2.50kx flake sort of structures are seen along with a breakdown of flakes during the milling operation. The powder shows the polycrystalline structure at a higher resolution. Figure 1(d) depicts the HRXRD spectra with elemental peaks at $2\theta = 13.9°$, $2\theta = 16.8°$, $2\theta = 25.2°$ and at $2\theta = 44.6°$ respectively. Figure 1(e) display the presence of carbon hybridized peak on the self-developed coatings which was captured by Raman spectra peak. In the existing research work, research scholar has employed Raman spectroscopy to studying the structural properties. In the existing research work, noticeable peaks of carbon coating has observed at 1140, 1340–1350, and 1550–1600 cm^{-1}. It is recorded that peaks around 1340–1350 cm^{-1} and at 1140 cm^{-1} show sp^1 or sp^2 hybridized carbon deposition, while peaks around 1550–1600 cm^{-1} show the presence of sp^3 hybridization with the presence of DLC coating [26–28].

3.2. Carbon coating mechanical characterization

3.2.1. Micro-hardness

Figure 2(a) shows the micro-harness of coatings deposited in the presence of nitrogen, oxygen, and LPG using high-velocity oxy-flame (HVOF). The experimental results after wear test depicts that as the experimental conditions increases the micro-hardness rapidly increased from 520 to 580 HV and hardness stabilized ~580 HV. The micro-hardness test showed an 11.5% increase in micro-hardness at test conditions of 1 m s^{-1} sliding velocities, 250 °C temperature, and 50 N load. It has been found in conducted study by Tyagi *et al* [26], that as temperature ,load and velocity increases the carbon coating hardness is also increases significantly as obtained in present work. The rapid increase in micro-hardness is due to strong adhesion between coating particles, agglomerate size of powders, carbides layer, and formation of oxides for piston rings [26]. It has been also investigated and concluded by Matikainen *et al* [29] in his experimental work that high rate velocity travelling particles enhance the hardness of the developed coating. Moreover, it is also seen that the presence of hard phased structured particles works as an impediment contrary to deformation, tends to results in an increase in hardness [30].

3.2.2. Residual stress

Figure 2(b) shows the residual stress of coatings deposited using HVOF. For the measurement of residual stress, initially, the x-ray will incidence on the work-piece at 2θ angle from the x-ray unit and after striking the sample, the x-ray diffracted back at an angle of η and ψ and thereby forming the Debye ring. The small- single x ray exposure of the incidence ray has formed Debye ring, which clearly shows the concentration of residual stress in the substrate materials. Residual stresses are generated on the coated samples due to development of thermal stresses and applied biaxial load on the respective body is shown in figure 3 and can be evaluated by using the following equations:

$$\sigma = \sigma_{\text{thermal}} + \sigma_{\text{n}} \tag{1}$$

Where,

$$\sigma_{\text{n}} = \frac{1}{2}\{(\sigma_{\text{x}} + \sigma_{\text{y}}) + (\sigma_{\text{x}} - \sigma_{\text{y}})\cos 2\theta\} + \tau_{\text{xy}} \sin 2\theta \tag{2}$$

$$\sigma_{\text{thermal}} = \left\{\frac{\alpha \text{TL} - \lambda}{L}\right\} \text{E} \tag{3}$$

Therefore the total residual stress developed on the material will be:

$$\sigma_{\text{x}} = -\left(\frac{\text{E}}{1 + \text{V}}\right)\left(\frac{1}{\sin 2\eta}\right)\left(\frac{1}{\sin 2\psi}\right)\left(\frac{\partial \varepsilon \alpha 1}{\partial \cos \alpha}\right) \tag{4}$$

The experimental results after wear test show that as the test conditions increases the residual stress rapidly decreased in the range of −50 to −5 MPa. The residual stress test showed a 90% decrease in residual stress, at test conditions of the sliding velocity of 1 m s^{-1}, 250 °C temperature, and 50N load. Due to the bombardment of high velocity, particles on to the substrate materials results to development of compressive residual stresses. More increase in, load, temperature, and sliding velocity during tribo test, tensile stresses are developed which stabilities the compressive stresses [21, 26]. The increase in temperature during testing conditions results in an increase in thermal stresses and negligible structural stress. The increase in thermal stress results in decreases in residual stress. According to Skordaris *et al* [31], if the degree of temperature increases, value of thermal stresses are also tends to increase but structural stresses remain constant up to 400 °C. The various different colour combinations are used to represent the residual stresses concentrations like red colour and blue colour as shown in figure 4.

3.3. Tribological characterization of coating

In the present study COF is a function of temperature, load, velocity, and property of materials as shown in the equation:

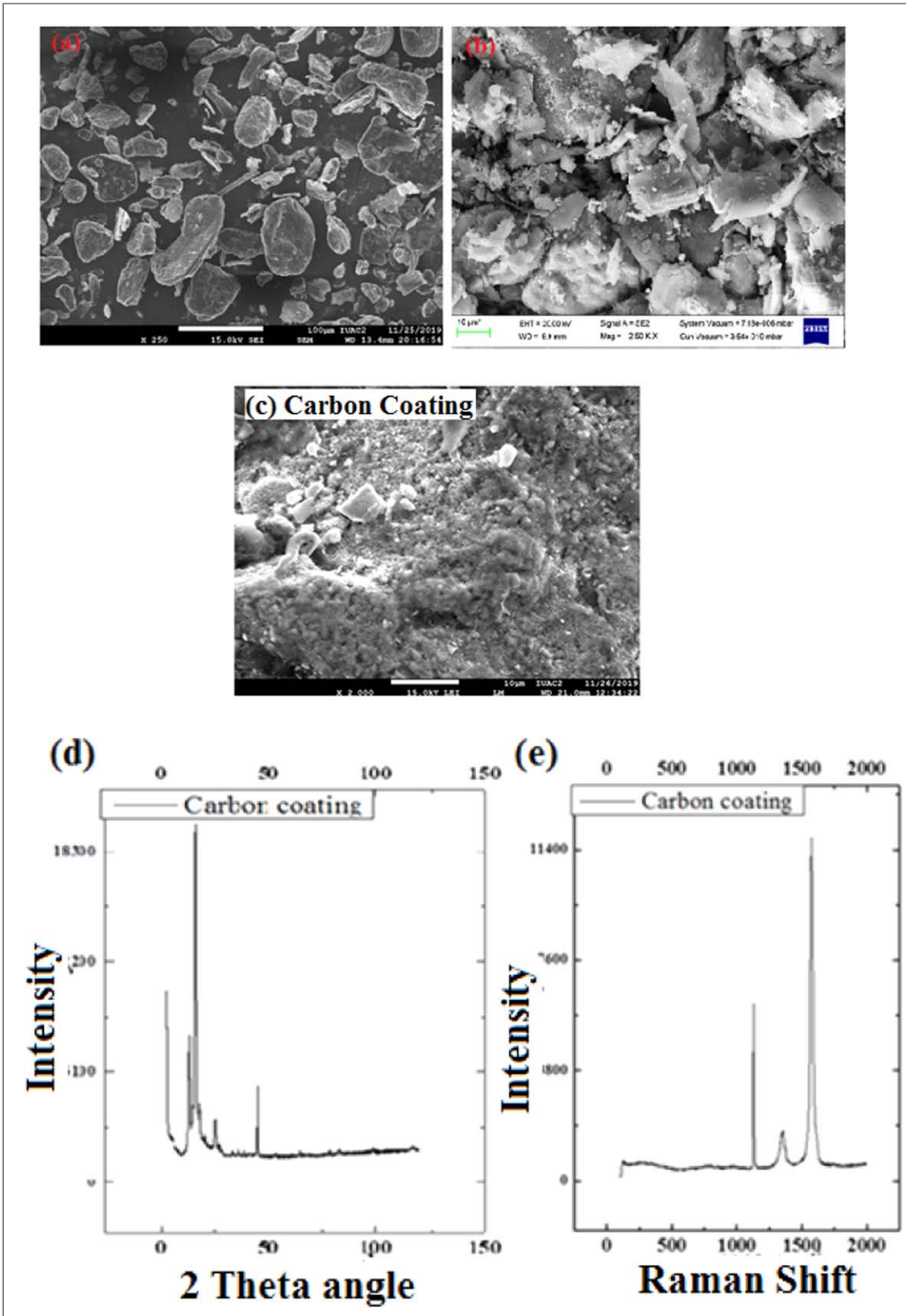


Figure 1. (a)–(c) FESEM + EDS image (d) HRXRD (e) Raman Spectra image of carbon coating.

$$\mu_{\text{cof}} = fT,\ V,\ L,\ \Upsilon\} \tag{5}$$

$$\text{Wear volume} = \frac{\pi(r - \sqrt{r * r - d * d/4})}{6(3d * d/(4 + h\,h * \text{h}))} \tag{6}$$

Theoretically wear volume can be calculated using equation (6) and volume loss of material can be calculated using equation (7)

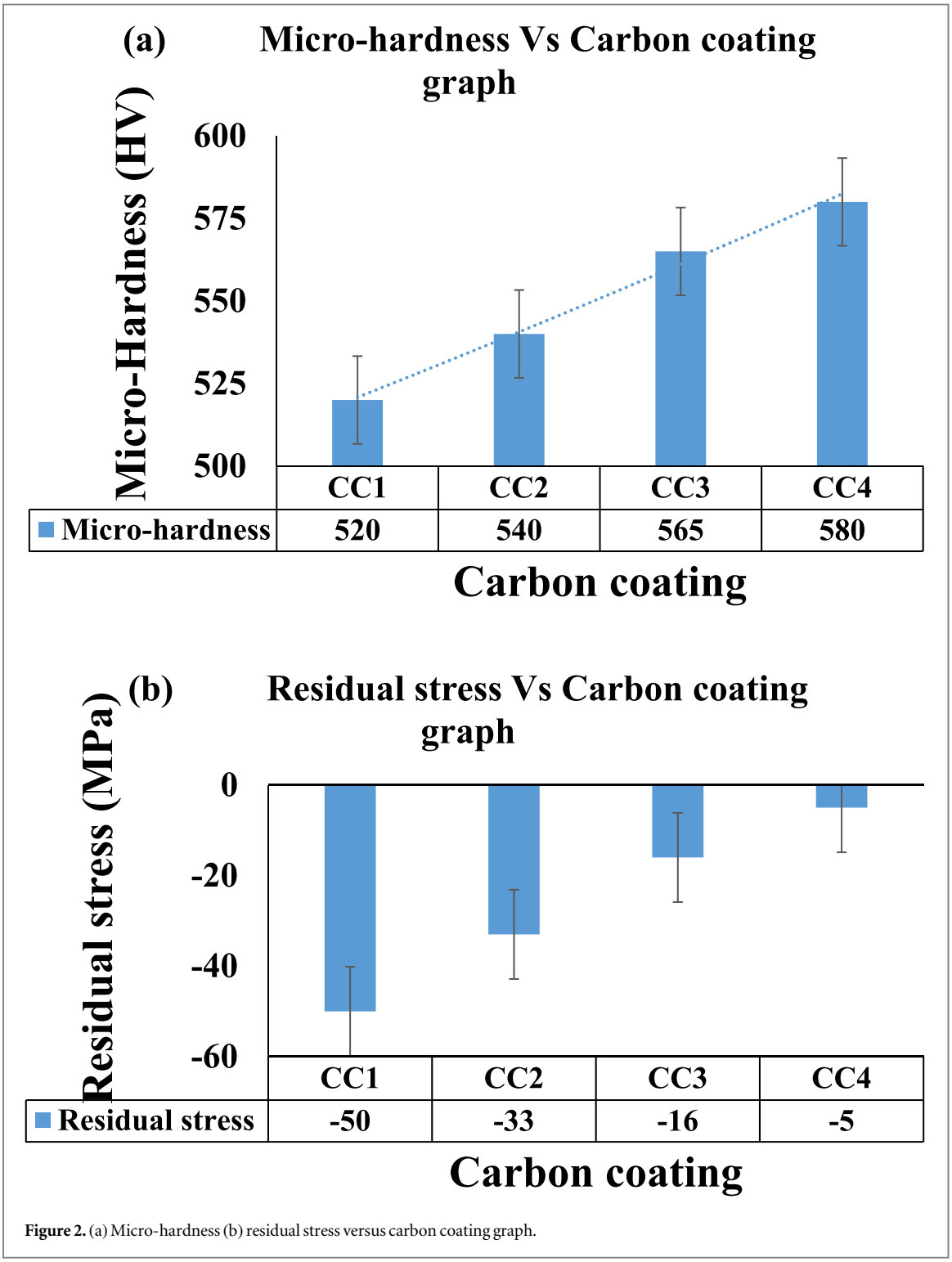


Figure 2. (a) Micro-hardness (b) residual stress versus carbon coating graph.

$$\text{Volume loss}(\text{mm}^3) = \frac{\textit{Mass loss}\ (\text{gm}) * 10000}{\textit{Density}\left(\frac{\text{gm}}{\text{cm}^3}\right)} \quad (7)$$

The pin-on-disk tribo-meter is used for evaluating the tribological performance such as coefficient of friction. The value of COF calculated by using pin-on-disk tribometer are in the range of 0.3 to 0.45 as shown in figure 5(a), whereas figure 5(b) clearly illustrates the range of wear which are in between of 45–150 μm, at experimental tested condition of temperature range of 100 to 250 °C, applied load of 50N and sliding velocity of 1 m s^{-1} respectively. In this research work, tribological analysis of self-made carbon coating has to be performed to calculating the wear behavior for the applications of IC engine embedded piston rings. The conducted tribological test on to the coated samples has shown a 70% decrease in the wear at tested condition of 1 m s^{-1} sliding velocities, applied load of 50N and 250 °C temperature. It has been perceived in present study that during the wear testing, the value of COF and wear are increase in the starting and become came into the stable condition during the running

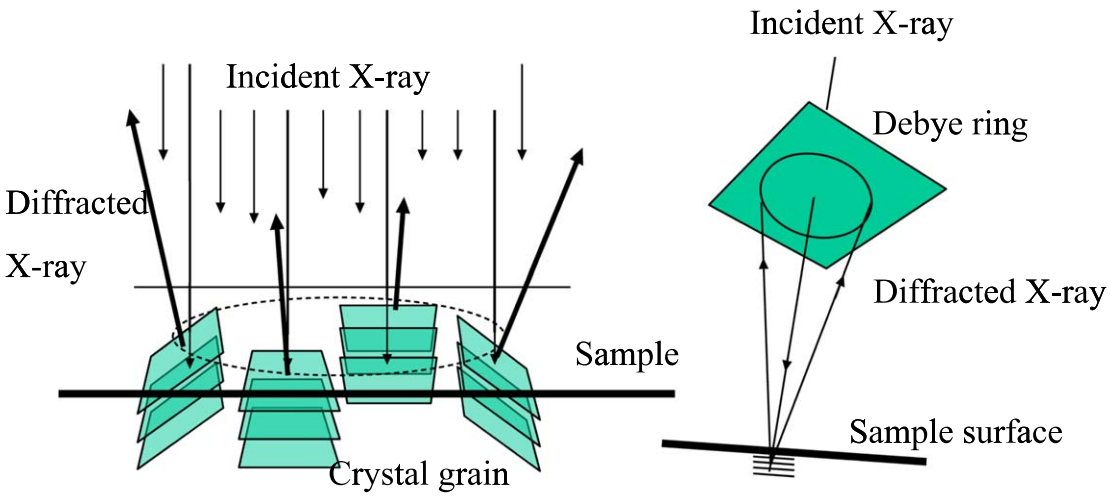


Figure 3. Systematic diagram of the residual stress of carbon coating.

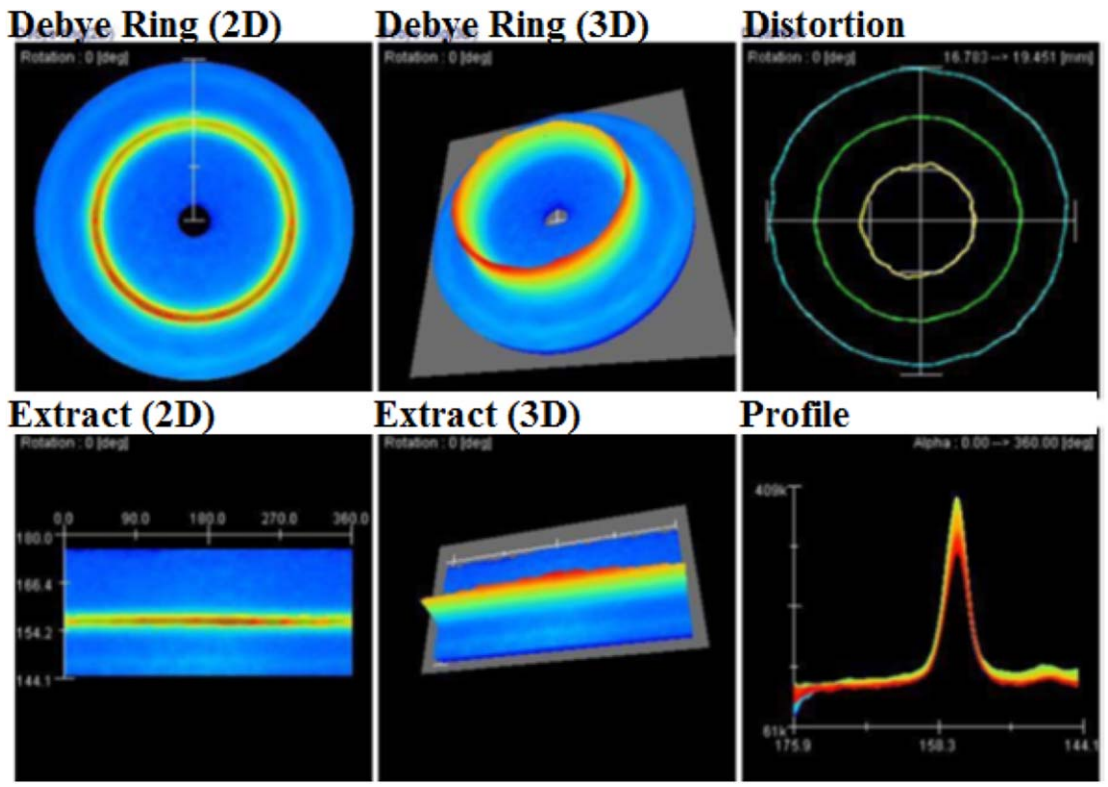


Figure 4. Residual stress result of the composite coating.

process between the counterpart and applied carbon coating, because of excellent interfacial strength as well as adhesive strength. The graphitization effect has noted due to the presence of carbon and hydrogen contents and built a lean solid lubricating layer [21, 26, 32]. The study conducted by Tyagi *et al* [21] and Matikainen *et al* [29] have also recognized that wear resistance properties of self-developed coatings are enhanced due to the impingement of higher velocity particles[21, 26, 29],and the same results are also be seen in the present conducted tribological studies. According to Strobeck theory [33] and performed experimental results it has been seen that as the value of load,degree of hotness and velocity increases, carbon coating wear behaviour decreases, as attained in present study. The deviation of friction & wear shows approximately 500 m run-in-stage, after that a stable stage attain during the test [34, 35]. During the running stage, carbon coating shows severe fluctuation in COF and wear, but later on, became stable after the introduction of graphitization effect in the coating. It was quite clear from the experimental analysis that as the carbon contents increase friction and wear of carbon-based coating decreases, since generally, carbon

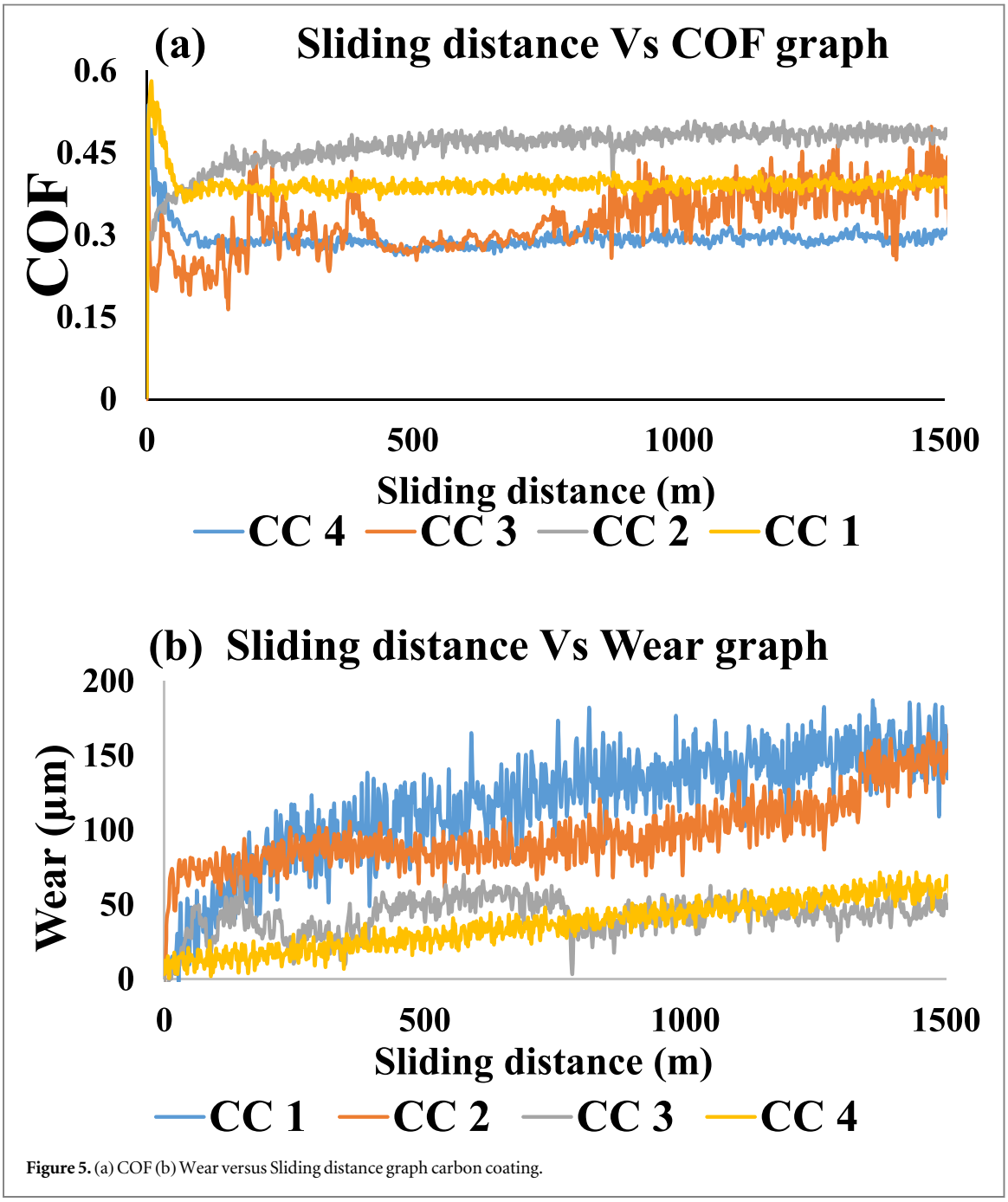


Figure 5. (a) COF (b) Wear versus Sliding distance graph carbon coating.

phases act as a good lubricant (solid lubricated layer) [21, 26]. The changes in phase structure and stoichiometry composition, formation of the additive layer, and oleophobic behavior of developed carbon coatings played an important role that outcomes in a decrease in friction and wear.

3.4. Wear mechanism of carbon coating

The coated surfaces of piston rings and cylinder bore follows Gaussian asperities after the termination of running wear without altering surface topography significantly [36]. The total load carrying capacity between the cylinder bore and piston rings can be calculated using equation [37] and as shown in figure 6.

$$W_t = \int_0^R (dWa + pdA) \tag{8}$$

During tribological test, rolling, sticking and sliding friction occurs on the symmetrical surface. Considering the equilibrium of forces during tribological testing:

$$(\sigma_r + d\sigma_r)(r + \mathrm{Dr})\mathrm{d}\theta\ \mathrm{h} - \sigma_r \mathrm{rd}\theta\mathrm{h} - 2\sigma_\theta \sin\frac{\mathrm{d}\theta}{2}\mathrm{drh} - 2\tau_r(\mathrm{rd}\theta\mathrm{Dr}) = 0 \tag{9}$$

$$\frac{\mathrm{d}\sigma_r}{\mathrm{dr}} + \left[\frac{\sigma_r - \sigma_\theta}{r}\right] - 2\frac{\tau_r}{\mathrm{h}} = 0 \tag{10}$$

Considering sliding friction all over the surface and sliding force will be acting on the surface during radius R. The pressure distribution on the entire

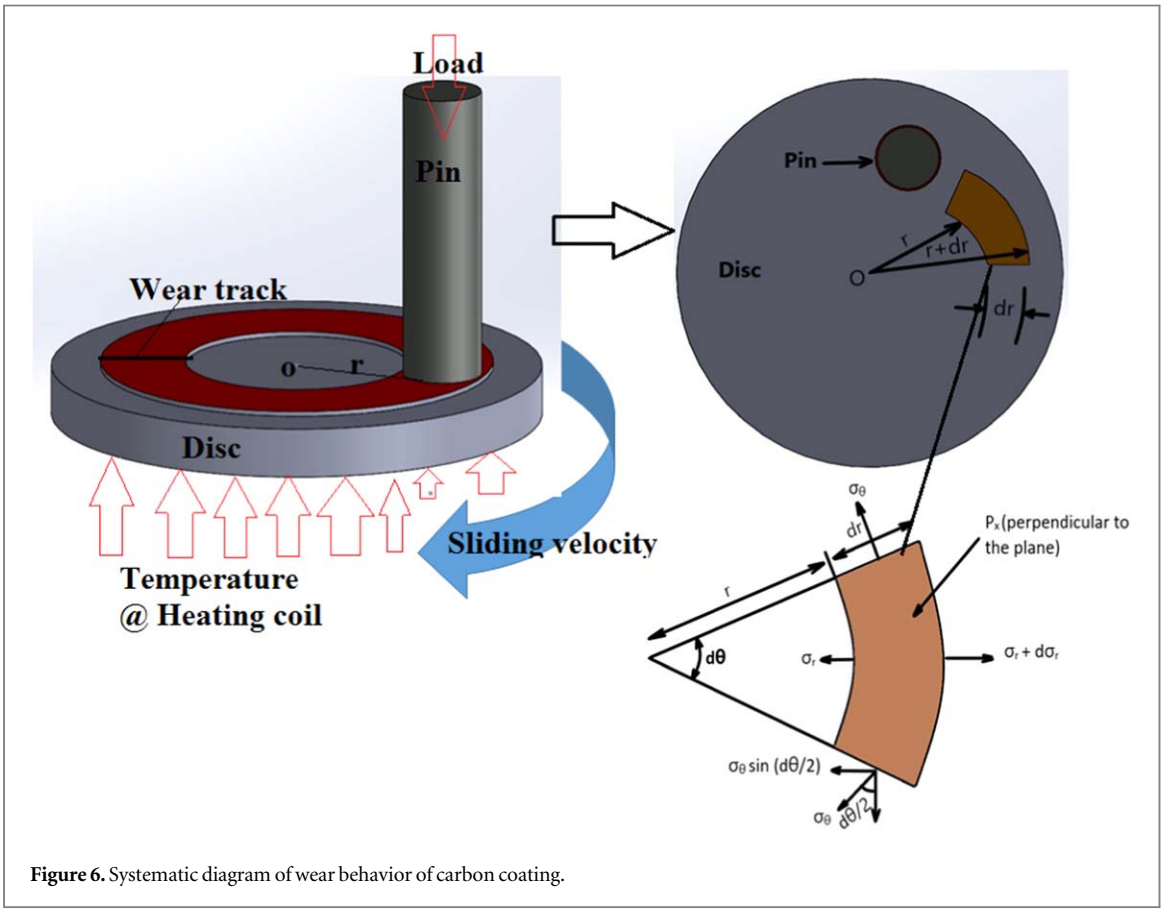


Figure 6. Systematic diagram of wear behavior of carbon coating.

surface without sticking will be given the equation:

$$\int \frac{-dp}{p} = -\frac{2\mu}{h} r \int dr \tag{11}$$

$$\ln \mathrm{P} = -\frac{2\mu}{h} r + \frac{2\mu}{h} R + \ln \sigma_o \tag{12}$$

$$\mathrm{P_s} = \sigma_\theta e^{\frac{2\mu}{h}(\mathrm{R}-\mathrm{r})} \tag{13}$$

Therefore the total sliding forces without sticking on the surfaces will be:

$$\mathrm{dF} = \mathrm{P_s} * 2\pi r dr \tag{14}$$

$$\mathrm{F_s} = \int_R^0 \left\{ \sigma o + \frac{2K}{h}(R - r) \right\} 2\pi r dr \tag{15}$$

Considering sticking friction all over the surface and sticking force will be acting on the surface during radius R_s. The pressure distribution on the entire surface without sliding will be given the equation

$$\int -dp = -\frac{2\mu}{h} \int dr \tag{16}$$

$$\mathrm{P} = -\frac{2\mu}{h} r + \frac{2\mu}{h} R + \sigma_o \tag{17}$$

$$\mathrm{P_{st}} = \mathrm{P_s} + \frac{2K}{h} \mathrm{R_{st}} - \mathrm{r}\frac{2K}{h} \tag{18}$$

$$\mathrm{P_{st}} = \mathrm{P_s} + \frac{2K}{h}(\mathrm{R_{st}} - \mathrm{r}) \tag{19}$$

Therefore the total sticking forces without sliding on the surfaces will be:

$$\mathrm{dF} = P_{\mathrm{st}} * 2\pi r dr \tag{20}$$

$$\mathrm{F_{st}} = \int_{Rs}^0 \left\{ \mathrm{Ps} + \frac{2K}{h}(Rs - r) \right\} 2\pi r dr \tag{21}$$

During tribological testing, both rolling and forging forces (sticking + sliding) will occur on the symmetrical surface. Initially sticking friction will occur on the surface from radius 0 to R_s and sliding force will be acting on the surface during radius R_s to R while rolling force will be acting on the entire surface. Therefore the total forces on the surface:

$$\text{Total Force} = \int_0^{Rs} \left\{ \mathrm{Ps} + \frac{2K}{h}(Rs - r) \right\} 2\pi r dr + \int_{Rs}^{R} \{\sigma\theta e^{\frac{2\mu}{h}}(R - \mathrm{r})\} 2\pi r dr + \sigma_\theta b \tag{22}$$

Where μ is the coefficient of friction, P is the pressure exerted, σ is the stress generated during the process.

Different stages of wear behavior of carbon coating are illustrated as:

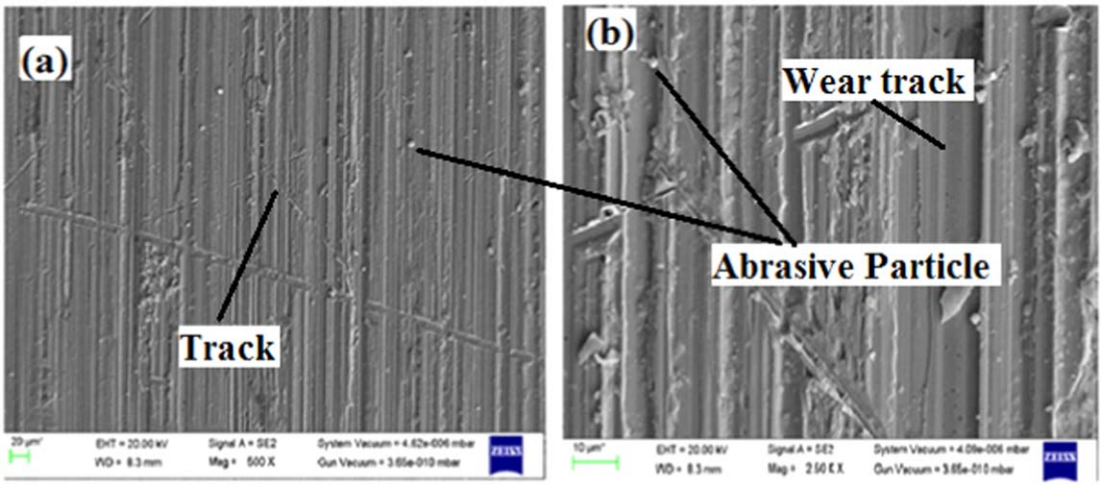
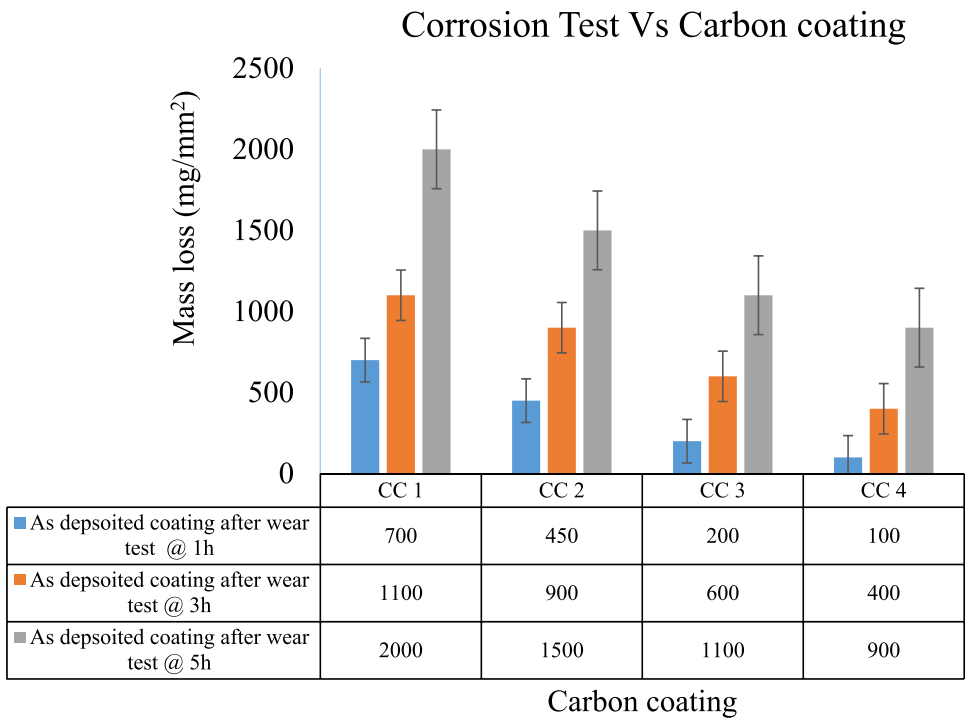


Figure 7. FESEM images of wear characteristics of carbon coating.



	CC 1	CC 2	CC 3	CC 4
As depsoited coating after wear test @ 1h	700	450	200	100
As depsoited coating after wear test @ 3h	1100	900	600	400
As depsoited coating after wear test @ 5h	2000	1500	1100	900

Figure 8. Corrosion test result of the carbon coating.

1st stage:

Initially, asperities of developed coating shot to penetrate disc during the running condition, when variable load, sliding velocity, and temperature is applied, results in plastic deformation.

2nd stage:

Plowing of the disc by carbon coating asperities results in the formation of tribofilm formation (due to carbon diffusion and oxide layer formation), which results in the weakening of asperities.

3rd stage:

An increase in oxides, carbides layer formation, and carbon diffusion shows smoothening of coating surface and lastly consequences in COF and wear reduction.

The figures 7(a)–(b). Illustrates the FESEM images, which clearly show the wear characterization of the distorted-out surface of developed carbon coatings. The FESEM images obtained after conducting the wear test by using a Pin-on-disk tribometer have shown the brittle structure of grains besides the transformation of the polycrystalline structure. The coating shows abrasive wear with the presence of plowing action of chips-off, rubbing with worn-out particles [32]. It has also been experienced during the test that as the diffusion of carbon particles increases and

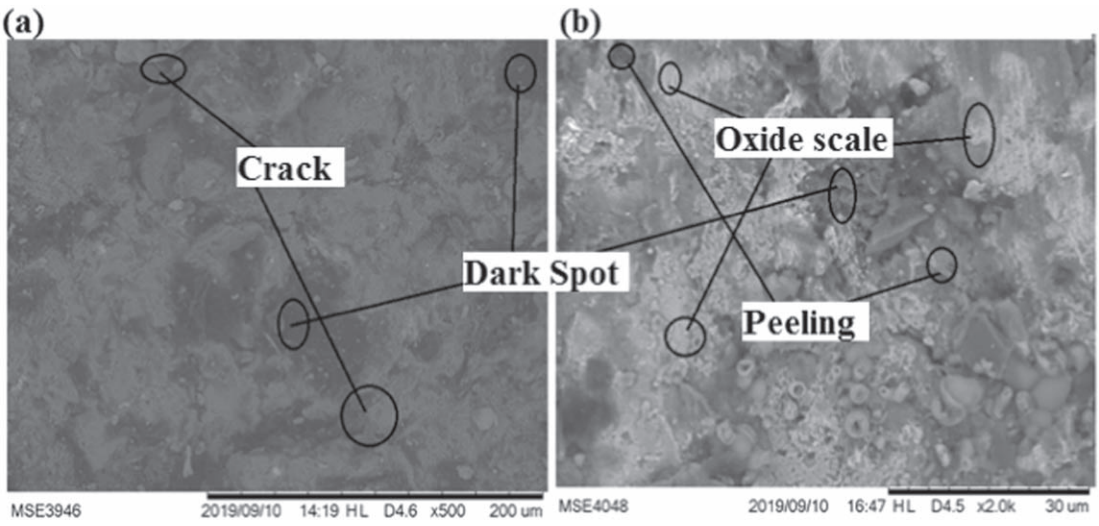


Figure 9. (a)–(b) FESEM images of coated samples after corrosion test.

formation of tribo-film increases as experimental tested conditions increases [21, 26].

3.5. Corrosion test

The mass loss during corrosion test of as-deposited carbon coatings after high-temperature wear test was evaluated by using electrochemical process was employed using an electrolytic solution of 3.5 wt% of NaCl powder in 1000 ml distilled water. The measured mass loss per unit area against the exposure time results of all the test combinations is presented in figure 8 and mathematically average mass loss can be calculated by using equation (23):

$$\text{Average Mass loss} = \Delta W/A \tag{23}$$

The experimental test results after wear test shows that mass loss after corrosion test showed as-deposited samples CC 1, CC 2, CC 3 and CC 4 was observed to be about 700, 450, 200 and 100 mg mm^{-2} at an exposure time of 1 h, while at an exposure of 3 h the mass loss per unit area was increased to 1100, 900, 600, 400 mg mm^{-2} respectively for CC 1, CC 2, CC 3 and CC 4 samples. On further increase in exposure time at 5 h, the mass loss per unit area further increases to 2000, 1500, 1100 and 900 respectively for CC 1, CC 2, CC 3 and CC4 samples. It is noted from the experimental test results of as-deposited carbon coatings after wear test that as the temperature increases from 100 °C–250 °C, the mass loss per unit area also decreases due to the presence of hard phased structured micron size carbon particles works as an impediment contrary to deformation [38]. Hukovic *et al* [39], Ali *et al* [40], and Singh *et al* [41], the crystalline structure delivers better corrosion resistance on comparing with the amorphous structure. The experimental result of the corrosion test illustrates that the overall measured mass loss per unit area of the as deposited sample was decreased to be about 85.7% at an exposure time of 1 h, 63.6%, at an exposure time of 3 h, and 55% at an exposure time of 5 h.

Figure 9 shows the FESEM images of samples subjected to corrosion testing. Some white shiny sparkles along with swelling and fragile scale, dark spot with peeling off exhibits on the corrosive samples. However, fewer cracks and spallation were observed on annealed heat-treated samples. It was also noted that oxides scales or protective oxides are formed which results in a reduction of mass loss per unit area [42]. It was also noted that coating remains intact which means less spallation tendency at an evaluated temperature [43].

3.6. Oxidation test

The mass gain during oxidation test of as-deposited carbon coatings after high-temperature wear test was evaluated and employed using an electric muffle furnace. The measured mass gain per unit area against the exposure time results of all the test combinations is presented in figure 10. The experimental test results after wear test shows that mass gain after oxidation test showed as-deposited samples CC 1, CC 2, CC 3 and CC 4 was observed to be about 900, 700, 500 and 300 mg mm^{-2} at an exposure time of 1 h, while at an exposure of 3 h the mass gain per unit area was increased to 1300, 1050, 750, 375 mg mm^{-2} respectively for CC 1, CC 2, CC 3 and CC 4 samples. On further increase in exposure time at 5 h, the mass gain per unit area further increases to 1500, 1700, 1250 and 500 respectively for CC 1, CC 2, CC 3 and CC4 samples. It is noted from the experimental test results of as-deposited carbon coatings after wear test that as the temperature increases from 100–250°C, the mass gain per unit area also increases due to the presence of hard phased structured micron size carbon particles works as an impediment contrary to deformation [38]. According to Hukovic *et al* [39], Ali *et al* [40], and

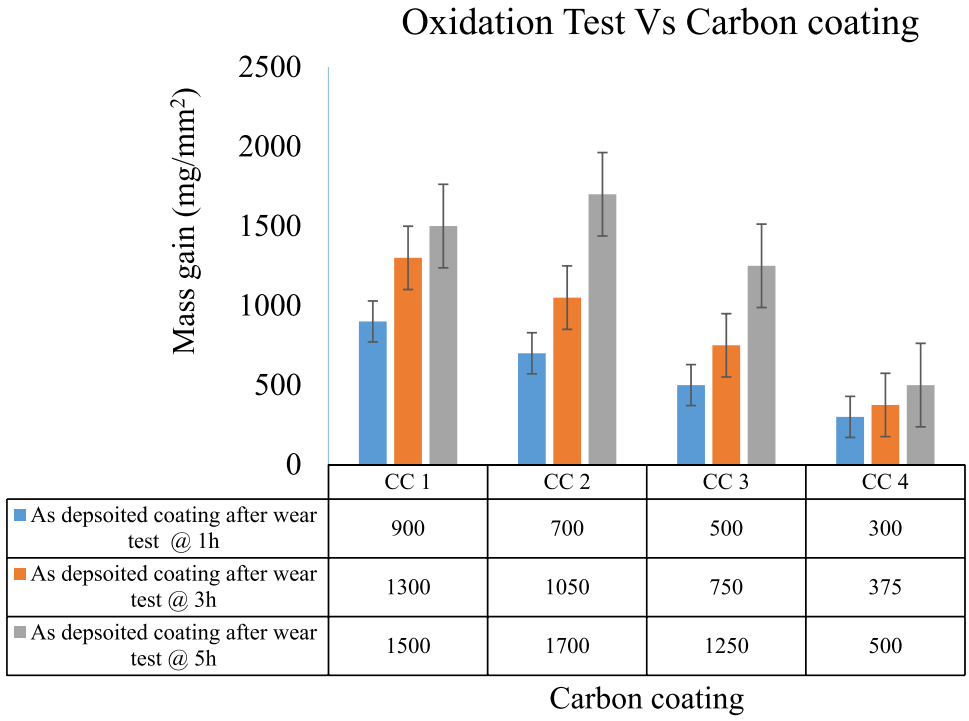


Figure 10. Corrosion test result of the carbon coating.

Singh *et al* [41], the crystalline structure provides better oxidation resistance on comparing with the amorphous structure. The experimental result of the oxidation test shows that the overall measured mass gain per unit area of the as deposited sample was increased to be about 66.6% at an exposure time of 1 h, 71.1%, at an exposure time of 3 h, and 66% at an exposure time of 5 h.

4. Conclusions

To abate frictional loss, emission and optimization of advanced coatings are vital for piston rings, thick low friction environment-friendly carbon coatings have been successfully deposited using high-velocity oxy-flame (HVOF). The major results of carbon coatings are:

1. The test result of carbon coating for micro-hardness is in the range of 520 to 580 HV and residual stress in the range of −5 to −50 MPa.
2. The value of COF calculated by using pin-on-disk tribometer are in the range of 0.3 to 0.45 while the range of wear which are in between of 45–150 μm, at experimental tested condition of temperature range of 100 to 250 °C, applied load of 50N and sliding velocity of 1 m s^{-1} respectively.
3. The tribological test showed an 11.5% growth in micro-hardness, 90% diminution in residual stress, 50% reduction in COF, and 70% reduction in the wear at test conditions of 250°C temperature, 50N load, and 1 m s^{-1} sliding velocities.
4. The experimental result of the corrosion test shows that the overall measured mass loss per unit area of the as deposited sample was decreased to be about 85.7% at an exposure time of 1 h, 63.6%, at an exposure time of 3 h, and 55% at an exposure time of 5 h.
5. The experimental result of the oxidation test shows that the overall measured mass gain per unit area of the as deposited sample was increased to be about 66.6% at an exposure time of 1 h, 71.1%, at an exposure time of 3 h, and 66% at an exposure time of 5 h.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

Ankit Tyagi https://orcid.org/0000-0001-9153-1402
Qasim Murtaza https://orcid.org/0000-0001-8784-2304
R S Walia https://orcid.org/0000-0002-1201-7892

References

[1] Nevshupa R, Conte M, del Campo A and Roman E 2016 Analysis of tribochemical decomposition of two imidazolium ionic liquids on Ti–6Al–4V through mechanically stimulated gas emission spectrometry *Tribol. Int.* **102** 19–27

[2] Profito F J, Tomanik E and Zachariadis D C 2016 Effect of cylinder liner wear on the mixed lubrication regime of TLOCRs *Tribol. Int.* **93** 723–32

[3] Zavos A B and Nikolakopoulos P G 2015 Simulation of piston ring tribology with surface texturing for internal combustion engines *Lubr. Sci.* **27** 151–76

[4] Igartua A, Nevshupa R, Fernandez X, Conte M, Zabala R and Bernaola J 2011 Alternative eco-friendly lubes for clean two-stroke engines *Tribol. Int.* **44** 727–36

[5] Picas J A, Forn A and Matthäus G 2006 HVOF coatings as an alternative to hard chrome for pistons and valves *Wear* **261** 477–84

[6] Skopp A, Kelling N, Woydt M and Berger L M 2007 Thermally sprayed titanium suboxide coatings for piston ring/cylinder liners under mixed lubrication and dry-running conditions *Wear* **262** 1061–70

[7] Borghi A, Gualtieri E, Marchetto D, Moretti L and Valeri S 2008 Tribological effects of surface texturing on nitriding steel for high-performance engine applications *Wear* **265** 1046–51

[8] Friedrich C, Berg G, Broszeit E, Rick F and Holland J 1997 PVD CrxN coatings for tribological application on piston rings *Surf. Coat. Technol.* **97** 661–8

[9] Igartua A, Fernandez X, Woydt M, Luther R and Illaramendi I 2011 Tribological tests to simulate wear on piston rings ed P A Lakshminarayanan and N S Nayak *Critical Component Wear in Heavy-Duty Engines* (Singapore: Wiley) pp 167–95

[10] Lima LGDBS, Nunes L C S, Souza R M, Fukumasu N K and Ferrarese A 2013 Numerical analysis of the influence of film thickness and properties on the stress state of thin film-coated piston rings under contact loads *Surf. Coat. Technol.* **215** 327–33

[11] Wielage B, Wilden J, Schnick T, Wank A, Beczkowiak J, Schülein R and Ren H Z H Mechanically alloyed SiC composite powders for HVOF applications *Int. Therm. Spray Conf., n.d* pp 4–8

[12] Wank A and Wielage B 2003 High energy ball milling-a promising route for the production of tailored thermal spray consumables *Conf. Mod. Wear Corros. Resist. Coatings, Warsaw*

[13] Broszeit E, Friedrich C and Berg G 1999 Deposition, properties and applications of PVD Cr N coatings *Surf Coatings Technol* **115** 9–16

[14] Chatha S S, Sidhu H S and Sidhu B S 2012 Characterisation and corrosion-erosion behaviour of carbide based thermal spray coatings *J Miner Mater Charact Eng* **11** 569–86

[15] Bolelli G, Berger L M, Bonetti M and Lusvarghi L 2014 Comparative study of the dry sliding wear behavior of HVOF-sprayed WC-(W, Cr)2C-Ni and WC-CoCr hard metal coatings *Wear* **309** 96–111

[16] Costa M Y P, Cioffi M O H, Voorwald H J C and Guimares V A 2010 An investigation on sliding wear behavior of PVD coatings *Tribol. Int.* **43** 2196–202

[17] Barbezat G 2005 Advanced thermal spray technology and coating for lightweight engine blocks for the automotive industry *Surf Coatings Technol* **200** 1990–3

[18] Liu S, Sun D, Fan Z, Yu H and Meng H M 2008 The influence of HVAF powder feedstock characteristics on the sliding wear behavior of WC-NiCr coatings *Surf Coatings Technol* **202** 4893–900

[19] Sidhu H S, Sidhu B S and Prakash S 2010 Wear characteristics of Cr_3C_2-NiCr and WC-Co coatings deposited by LPG fueled HVOF *Tribol. Int.* **43** 887–90

[20] Derelizade K, Venturi F, Wellman R G, Kholobysov A and Hussain T 2021 Wear performance of graphene nano platelets incorporated WC-Co coatings deposited by hybrid high velocity oxy fuel thermal spray *Wear 482-* **483** 203974

[21] Tyagi A, Walia R S and Murtaza Q Tribological behavior of HVOF carbon coating for wear resistance applications *Mater. Res. Express* **6**

[22] Bhosale D G and Rathod W S 2020 Tribological behaviour of atmospheric plasma and high velocity oxy-fuel sprayed WCCr3C2-Ni coatings at elevated temperatures *Ceram. Int.* (https://doi.org/10.1016/j.ceramint.2020.01.288)

[23] Bhosale D G, Rathod W S and Rukhande S W 2019 Sliding wear behavior of high velocity oxy-fuel sprayed WC-Cr_3C_2-Ni coating for automotive applications *Materials Today: Proceedings xxx (xxxx) xxx* **19** 339-343

[24] Lin J, Wei R, Bitsis D C and Lee Peter M. 2016 Development and evaluation of low friction TiSiCN nanocomposite coatings for piston ring applications *Surface & Coatings Technology* **298** 121–31

[25] Jaguaribe E F, Medeiros L L, Barreto M C S and Araujo L P 2005 The performance of activated carbons from sugarcane bagasse, babassu, and coconut shells in removing residual chlorine *Braz. J. Chem. Eng.* **22** 41–7

[26] Tyagi A, Walia R S and Murtaza Q 2019 Tribological behavior of temperature-dependent environment-friendly thermal CVD diamond coating *Diamond & Related Materials* **96** 148–59

[27] Manninen N, Ribeiro F, Escudeiro A, Polcar T, Carvalho S and Cavaleiro A 2013 Influence of Ag content on mechanical and tribological behavior of DLC coatings *Surf. Coat. Technol.* **232** 440–446

[28] Casiraghi C, Ferrari A and Robertson J 2005 Raman spectroscopy of hydrogenated amorphous carbons *Phys. Rev.; B* **72** 085401

[29] Matikainen V, Koivuluoto H, Vuoristo P, Schubert J and Houdková Š 2018 Effect of nozzle geometry on the microstructure and properties of HVAF-sprayed WC-10Co_4Cr and Cr_3 $C_{2\text{-}25}$NiCr coatings *J. Therm. Spray Technol.* **27** 680–94

[30] Wu H, Lan X D, Liu Y, Li F, Zhang W D, Chen Z J, Zai X F and Zeng H 2016 Fabrication, tribological and corrosion behaviors of detonation gun sprayed Fe-based metallic glass coating *Trans. Nonferrous Metals Soc. China* **26** 1629–37

[31] Skordaris G, Bouzakis K-D, Charalampous P, Kotsanis T, Bouzakis E and Lemmer O 2016 Effect of structure and residual stresses of diamond-coated cemented carbide tools on the film adhesion and developed wear mechanisms in milling *CIRP Annals - Manufacturing Technology* **65** (1) 101–104

[32] Tyagi A, Pandey S M, Gupta K, Walia R S, Murtaza Q and Krishen K 2019 Tribological behavior of sustainable carbon-based composite coating for wear resistance applications' *Mater. Res. Express* **6** Mater. Res. Express 6 (2019) 125601

[33] Berger J and Füßer H-J Development of a test method for a realistic, single parameter-dependent analysis of piston ring versus cylinder liner contacts with a rotational tribometer' *Tribology International* **113** 111–24

[34] Tyagi A, Pandey S M, Murtaza Q, Walia R S and Tyagi M 2019 Tribological behavior of carbon coating for piston ring applications using Taguchi approach' *Materials Today Proceeding* **25** 759–64

[35] Tyagi Ankit, Mani Pandey S, Walia R S and Murtaza Q 2019 Characterization and parametric optimization of tribological properties of Mo blend composite coating', Material research express *Mater. Res. Express* **6** 086428

[36] Gore M, Morris N, Rahmani R, Rahnejat H, King P D and Howell-Smith S 2017 A combined analytical-experimental investigation of friction in a cylinder bore inserts under mixed and boundary regimes of lubrication *Lubr. Sci.* **29** 293–316

[37] Dolatabad N, Forder M, Morris N, Rahmani R, Rahneja H and Howell-Smith S 2020 'Influence of advanced cylinder coatings on vehicular fuel economy and emissions in piston compression ring conjunction *Appl. Energy* **259** 114129

[38] Wu H, Lan X D, Liu Y, Li F, Zhang W D, Chen Z J, Zai X F and Zeng H 2016 Fabrication, tribological and corrosion behaviors of detonation gun sprayed Fe-based metallic glass coating *Trans. Nonferrous Metals Soc. China* **26** 1629–37

[39] Metikos-Hukovic M, Tkalčec E, Kwok A and Piljac J 2003 An *in vitro* study of Ti and Ti- alloys coated with sol-gel derived hydroxyapatite coatings *Surf. Coat. Technol.* **165** 40–50
[40] Ali M S, Song S and Xiao P 2002 Evaluation of degradation of thermal barrier coatings using impedance spectroscopy *J. Eur. Ceram. Soc.* **22** 101–7
[41] Singh A, Singh G and Chawla V 2018 Influence of post coating heat treatment on microstructural, mechanical and electrochemical corrosion behavior of vacuum plasma sprayed reinforced hydroxyapatite coatings *J. Mech. Behav. Biomed. Mater.* **85** 20–36
[42] Krzyzanowski M, Beynon J H and Farrugia D C J 2010 *Oxide Scale Behavior in High-Temperature Metal Processing* (New York: Wiley)
[43] Bai M, Song B, Reddy L and Hussain T 2019 Preparation of MCrAlY-Al2O3 composite coatings with enhanced oxidation resistance through a novel powder manufacturing process *J Thermal Spray Tech* **28** 433–443

Sustainable Energy Technologies and Assessments 49 (2022) 101714

Contents lists available at ScienceDirect

Sustainable Energy Technologies and Assessments

journal homepage: www.elsevier.com/locate/seta

Exergy and energy analysis of sensible heat storage based double pass hybrid solar air heater

Om Prakash [a,*], Anil Kumar [b,c,*], Samsher [b,c,d], Kumaresh Dey [e], Ankish Aman [a]

[a] Department of Mechanical Engineering, Birla Institute of Technology, Mesra, Ranchi, India
[b] Department of Mechanical Engineering, Delhi Technological University, Delhi, India
[c] Centre for Energy and Environment, Delhi Technological University, Delhi, India
[d] Harcourt Butler Technical University, Kanpur, India
[e] Department of Mechanical Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, India

ARTICLE INFO

Keywords:
Exergy
Energy
Sensible heat storage
Metco material

ABSTRACT

This research work discusses the exergy and energy analysis of an innovative double pass solar air heater with sensible heat storage. The system is evaluated in three different operating conditions; namely, natural convection mode, induced forced convection mode, and induced forced convection mode with a reflector. Performance analysis in all operating conditions evaluated and the best operating condition selected to apply crop drying and space heating application. Metco material and aluminum scrap mixture were used as sensible heat storage material below the bottom plate of the system. Experiments were conducted on three consecutive days for the different operating conditions. Average energy and exergy efficiency under natural convection mode is 11.47% and 2%, while 56% and 10.43% for a system operated under induced forced convection mode; however, 86.19% and 17.617% for a system operated under induced forced convection mode with a reflector. The third condition (system under induced forced convection mode with a reflector) has superior heat performance than the other two cases. Therefore, the proposed system is highly suitable for low-temperature applications.

Introduction

Environmental pollution has increased in the last 20 years. Poisonous gas emissions like CO, NOx, etc have increased linearly due to the overuse of nonrenewable energy sources. Therefore, regular planning and research are needed to replace the application of exhaustible energy resources. Renewable energy has a broad scope of research and can replace nonrenewable sources of energy in the future. Solar technology and its applications are the current topics of research and innovation. Initiative and planning are being carried forward to harness maximum energy from solar panels and solar power plants to meet the energy demand and reduce energy stress on coal power plants.

One of the eco-friendly applications of solar technology is the solar air heater (SAH). It is a cost-effective renewable device used to heat air for proper conditioning in commercial and industrial buildings. Solar air heater (SAH) is application of solar thermal technology that absorbs solar radiation through an absorber plate and heating the air flowing through the ducts. SAH uses solar energy for space heating, crop drying and other industrial usages. There are various types of solar air heaters designed depending upon the requirement. A double pass solar air heater (DPSAH) is one of such kind with superior thermal performance compared to single-pass SAH. DPSAH utilizes heat losses from the absorber plate to heat the air passed from the bottom collector [1]. However, to increase efficiency, some design criteria should be maintained, including the absorber area, the absorptance of the absorber, and airflow.

Many researchers have worked towards the concept of DPSAH. Bahrehmand et al. [2] studied the exergy and energy analysis of different types of solar air collectors using tin absorber plate. Experimental research concludes that the double glass cover SAHs have better performance than single glass cover SAH. Research also found that triangular fins in the collector are more energy-efficient than rectangular fins. Velmurugan et al. [3] analyzed that triple-pass SAH gives better exergy and energy performance than single pass and DPSAH. Ghiami and Ghiami [4] conducted a comparative study based on energy and exergy analyses of a baffled solar air heater with latent storage collector. Rise in air temperature, exergy gain, and efficiency of a SAH

* Corresponding authors at: Department of Mechanical Engineering, Delhi Technological University, Delhi-110042 (India) (A. Kumar), Department of Mechanical Engineering, Birla Institute of Technology, Mesra, Ranchi (O.Prakash).
E-mail addresses: omprakash@bitmesra.ac.in (O. Prakash), anilkumar76@dtu.ac.in (A. Kumar).

https://doi.org/10.1016/j.seta.2021.101714
Received 21 December 2020; Received in revised form 27 July 2021; Accepted 15 October 2021
Available online 17 November 2021
2213-1388/© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Q_A	Total heat (W)
Q_u	Heat utilized (W)
Q_{loss}	Heat loss (W)
Q_{st}	Incident heat energy stored in the heater area (W)
m	Mass flow rate (MFR), kg/s
C_p	Specific Heat in constant pressure (J/kg.K)
T_{out}	Outlet Temperature (^{o} C)
T_{in}	Inlet Temperature (^{o} C)
V_{ai}	Velocity of air (m/s)
S	Area of the duct (m^2)
A_p	Area of Absorber Plate (m^2)
I_i	Solar radiation (W/m^2)
U_{loss}	Overall heat loss coefficient [$W/(m^2K)$]
T_p	Absorber plate temperature (°C)
T_a	Ambient Temperature (°C)
U_e	Side heat loss coefficient [$W/(m^2K)$]
U_b	Bottom heat loss coefficient [$W/(m^2K)$]
U_t	Top heat loss coefficient [$W/(m^2K)$]
$h_{r,Pg}$	Radiative heat transfer coefficient ($Wm^{-2}K^{-1}$)
$h_{c,Pg}$	Convective heat transfer coefficient ($Wm^{-2}K^{-1}$)
$h_{r,ga}$	Radiative heat transfer coefficient due to wind velocity ($Wm^{-2}K^{-1}$)
h_w	External radiative and convective heat transfer coefficient due to velocity of wind ($Wm^{-2}K^{-1}$)
Nu	Nusselt Number
Ra	Reynolds number
Pr	Prandtl number
t	Temperature (°C)
T_g	Temperature of glass plate (°C)
T_a	Ambient Temperature (°C)
ε_p	Emissivity of the plate
V	Velocity of air (m/s)
T_{sky}	Temperature of the external air at the surrounding (°C)
L_1	Length of Solar air heater (m)
L_2	Width of solar air heater (m)
L_3	Height of the solar air heater (m)
I_T	Solar radiation (W/m^2)
g	Gravitation constant ($m^3\ kg^{-1}\ s^{-2}$)
ΔT	Temperature difference of air, (°C)
L	Length of the solar air heater (m)
v	Kinematic viscosity of air (m^2/s)
D_h	Characteristic distance (m)
Ex_{in}	Exergy at entrance (J)
Ex_d	Exergy destroyed (J)
Ex_{out}	Exergy at outlet (J)
Ex_{st}	Exergy stored (J)
Ex_{los}	Exergy lost (J)
P_{in}	Air Pressure inside solar air heater [$kg/(m^2)$]
P_a	Atmospheric Pressure [$kg/(m^2)$])
I	incident solar radiation (W/m^2)
T_s	Mean temperature of the absorber plate (°C)
r	Distribution of solar radiation
A_c	Area of the collector (m^2)
$Ex_{d(p-s)}$	Exergy destroyed due to pressure drop (J)
η_0	Efficiency (%)
$Ex_{d(\Delta P)}$	Exergy destroyed from absorber plate to surrounding (J)
ΔP	Pressure drop [$kg/(mS^2)$]
$Ex_{d(p-f)}$	Exergy destroyed from absorber plate surface to flowing air (J)
W_X	Uncertainty due to anomalous system
$W_{Temperature}$	Error due to temperature measuring instrument
$W_{thermometer}$	Error due to thermometer
$W_{reading}$	Reading error
W_{Wind}	Error due to wind speed
$W_{anemometer}$	Error due to wind speed and pressure
W_{SR}	System deformities error
$W_{Pyranometer}$	Error in solar irradiance on a planar surface
$W_{Experimental}$	Experimental error

Greek Symbol

α	Absorptivity
ρ	Density of air (kg/m^3)
ε_p	Emissivity of absorber plate
ε_g	Emissivity of glass surface
λ	Insulation thermal conductivity (W/mK)
λ_e	Insulation thermal conductivity at the side (W/mK)
δ_b	Insulation thickness at the bottom (mm)
δ_e	Insulation thickness at the side (mm)
σ	Stefan–Boltzmann Constant ($W/m^2\ K^4$)
τ	Transmissivity
η_o	Overall thermal efficiency (%)
μ	Air viscosity (Ns/m^2)
β	Thermal expansion coefficient

depend on the mass flow rate (MFR) of air, solar radiation intensity, and the absorber plate's surface geometry. Chamoli et al. [5] analytically found that V-down pierced baffles in absorber plate can result in better exergy performance by increasing solar heat absorption and dissipation potential. Khadraoui et al. [6] analyzed the use of paraffin wax as a PCM (phase change material) in SAH. The use of paraffin as PCM can increase the daily (including night) efficiency of a SAH up to 33%. Chabane et al. [7] studied the efficiency of single-pass SAH using longitudinal fins by varying the MFR. Zukowski [8] studied the flow characteristics and efficiency of a novel microjet SAH. Developed device achieved heat conversion efficiency ranging from 66% to 90%, while a single glass cover can improve efficiency from 16% to 24%. Saxena et al. [9] conducted an experiment to study free and forced convection in SAH using a mixture of granular carbon and desert sand as PCM coupled with sensible thermal storage materials. Comparison is drawn between two similar sets of operations in natural convective flow. Others operating in forced convective flow stated that 19–20.7% thermal efficiency was achieved in natural convection and 53–80% thermal efficiency was achieved in forced convection. Siddiqui et al. [10] did a study to determine the role of SAH in design and performance enhancement in the water desalination process by the solar humidification-dehumidification system. Experiment was conducted in Medina, Saudi Arabia on two experimental setups, one without SAH and two electric heaters, while the second was set up with one electric heater and one SAH. The increased daily productivity rate and simultaneous cost decrease in the second experimental set up one with an electric heater and SAH placed before the humidification chamber. Pramanik et al. [11] did an experimental study to evaluate the performance of DPSAH. Better designed SAH, a dc fan and selective coating on the absorber plate were used. Average thermal efficiency and outlet temperature increased, and the difference between the experimental and theoretical values lies between 1.2% and 9.7%. Kalaiarasi et al. [12] designed a type of SAH with absorber plates made up of copper tubes filled with synthetic oil. Exergy analysis of the absorber cum heat storage is performed. Maximum exergy and energy efficiency obtained was 50–59% for higher MFR 0.026 kg/s. Moreover, Edalatpour et al. [13] did exergy and energy analysis of a double glazed SAH, which had paraffin wax as PCM. Experimental study was done on the same set up for three days of summer to evaluate daily exergy and

energy analysis. Daily energy efficiency of the system varied from 58% to 69%, while the exergy efficiency ranged from 14% to 26% in latitude, 37° 28′ N and longitude, 57° 20′ E. A double pass finned SAH studied by Fudholi et al. [14] at different MFRs and solar radiation. Study reveals that variation in MFRs and solar radiation affects energy efficiency, exergy efficiency and improvement potential for solar air heater. Srinivas et al. [15] experimented on photovoltaic SAH where the absorber plate is partially covered by a photovoltaic cell. It had been observed that the overall energy (thermal and electrical) efficiency varied between 29 and 37%, whereas the exergy efficiency varied between 14 and 17%. Abuşka [16] experimented on SAH with attached conical fins at the absorber plate to increase its efficiency and performance. Energy efficiency increased up to 74.6% (MFR.1 kg/s) and exergy up to 19.3% (0.04 kg/s). Heydari et al. [17] found out that DPSAH with helical channeling is 14.7% more efficient than simple ducts. Salih et al. [18] discovered that increased airflow rate results in delaying the melting point during charging of PCM. However, the inverse case was observed during the discharge since the airflow rate was inversely related to freezing time. Satyender et al. [19] did experimental work with converging finned wire mesh packed bed SAH. Experiment concluded that the converging sections improve the convective heat transfer enhancing thermal and thermohydraulic efficiency to 93% and 80%. Rhee et al. [20] studied transpired, corrugated and simple flat plate collectors having Re = 4000–16,000. The corrugated and transpired configuration had better efficiency with optimal pressure drop. Prasad and Sah [21] did an experimental study on SAH collector by modifying it with reflective mirrors to concentrate the solar radiation and circular ribs created for turbulence. The result showed that for Re = 4000–16,000 as solar radiation increased to 40% while performance increased up to 90% when reflective mirror combined with turbulator were used. Manjunath et al. [22] did numerical study of SAH by using wave profiled absorber plate. For Re = 4000–24,000. the average thermal efficiency increased by 12.5% while MFR reduced. Thakur et al. [23] did numerical study on SAH by modifying the absorber plate by using hyperbolic rib profile. Optimum rib performance achieved at Re 6000, roughness 1 mm height and 10 mm pitch. Study concluded to have improvement in thermo-hydraulic performance of SAH. Jaurker et al. [24] stated that the maximum heat transfer occurred at 6 relative roughness pitch. Nu improved by 2.7 times and friction factor rose up to 3.6 times whereas pressure loss reduced through the grooves in the absorber plate and rectangular ribs. Satyender et al. [25] stated that 49 °C outlet air temperature achieved for optimum MFR 0.023 kg/s for double pass converging finned wire mesh pack SAH. Tyagi et al. [26] stated that DPSAH has maximum thermal efficiency up to 33% when aluminum metal used in absorber plate. Pfister et al. [27] did a study of six grids assembly corrugated in a triangular fashion so that the air traverse 16 times through SAH with MFR 0.025–0.095.80% conversion efficiency was achieved for 0.06 kg/s MFR. Krishnananth and Murugavel [28] did experiment in SAH to study the effect of aluminum capsules and paraffin-waxes material. Thermal performance increased due to property of the material to act as a thermal storage. Anup et al. [29] stated that the heat transfer through modified SAH increased with 2.46% increase in Nusselt number and 1.78 times increase in friction factor for the normal duct at 60° twist angle of twisted roughness. Baissi et al. [30] done study on rectangular channel artificially roughed with two longitudinally curved delta-shaped baffles with perforation and without perforations SAH. Highest thermal enhancement factor for perforated configuration (2.26%) while 2.21% for non-perforated configuration. Filiz et al. [31] did study on DPSAH by modifying absorber plate connecting it with aluminum cans. MFR was in the range of 0.03–0.05 kg/s, helped aluminum cans to reduce the dead zones in the collector by better airflow and turbulence on both side of the absorber plate. Jha et al. [32] did mathematical modelling of DPSAH in parallel mode heat transfer for space air conditioning application. Mathematical model predicted 3–7 °C temperature rise. Rajneesh et al. [33] did geometrical modification of the absorber triangular cross section duct with forward facing chamfered rectangular ribs having Re 4000–18,000. For 1.5 rib aspect ratio 2.64 times Nu increased to 17,000 while Re value was 4000. Binguang et al. [34] did study of spiral shaped flow path of the air inside the duct with MFR0.0146–0.0250. Spiral type SAH had efficiency and heat loss coefficient higher in comparison to conventional SAH with inlet and ambient temperature difference of 5 °C and 3% efficiency reduction.

In view of the above literature survey, it is noticed that a lot of experimental and numerical works have already been done to analyze the performance of the double pass solar air heater. However, no literature was found on the energy and exergy analysis of double pass solar air heater with sensible thermal storage assisted with reflector for the composite climate. Hence the motivation of the present study is to design and develop double pass solar air heater with sensible thermal storage coupled with steel reflector for composite climate. A double pass solar air heater was fabricated for experimentation to achieve these objectives. Observed parameters have been used for performance analysis in terms of energy and exergy analysis. Energy efficient, environmentally friendly and sustainable double pass solar air heater is manufactured with locally available materials. Experiments were conducted on three consecutive days under different modes (natural mode, induced forced mode and induced forced mode coupled with reflector). In the second and third sections, energy and exergy analysis had performed based on experimental data. Proposed system under induced forced mode coupled with reflector shows superior results in both analyses. At last, the thermal efficiency of the proposed system is compared with previous competitive literature. The result of the proposed system is found to be comparable with other systems.

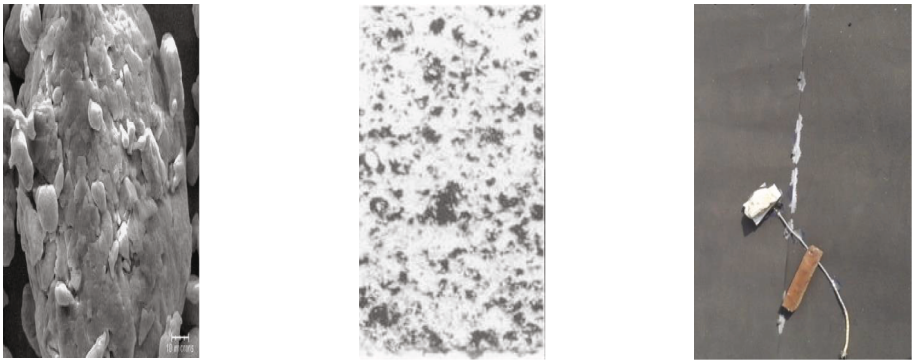
Material and methods

Experimental setup and procedure

A double pass solar air heater with sensible heat storage (DPSAHSHC) has been designed, developed, and experimented with in different conditions in the present study (Figs. 1–4). Fig. 1 shows the microscopic view of Metco material and its coating on the absorber plate. Dark surface having a high absorptance, and hence collector absorber is blackened if not naturally dark. Black surface would absorb over 95% of the incident solar radiation. Also, to avoid heat loss from the air heater wall, ceramic wool type insulation material having 20 mm thickness is used. Sunlight exposed collector led to useful heat towards air and some heat losses to the surrounding. Cumulative heat losses to the surrounding area are in three ways i.e. side, top and edge.

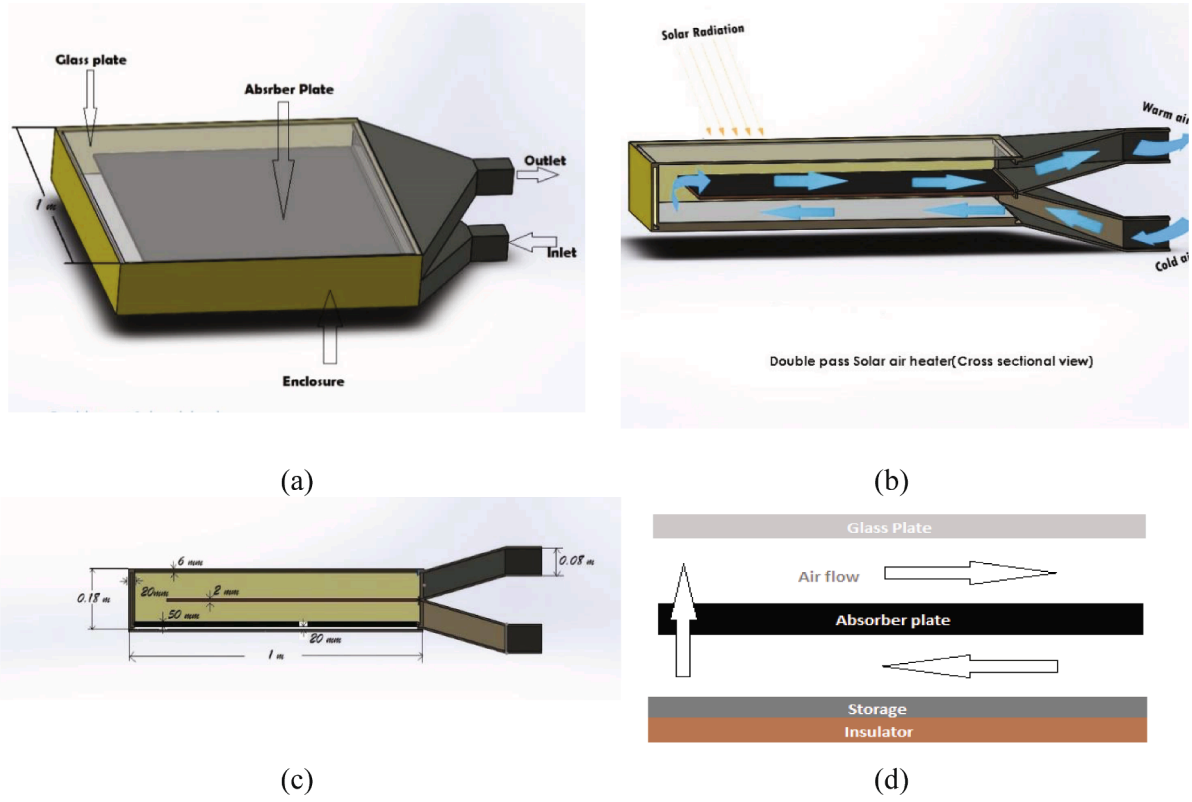
Fig. 2 represents the various section views of the proposed system. System is kept at tilt angle of the location (23°) and south-facing to obtain maximum solar radiation. Absorber plate of the DPSAHSHC is made up of a copper sheet with 2 mm thickness. Absorber surface is coated with a black metco material (Aluminum silicon-boron nitride powder with high absorptivity and low corrosion resistance). Copper or steel material has limitations since it would require a selective coating to maximize the absorption of solar energy and minimize the radiation emitted by the plate. Metco material with aluminum scrap (20 kg) is used as thermal storage material in DPSAHSHC. A 50 mm thick layer of Metco material with aluminum scrap is spread at the bottom inner wall of the SAH. In the natural mode, air comes in the down vent and get exhaust through upper vent. Various items involved in the manufacture of the proposed system and its embodied energy are given in Table 1. Embodied energy is the energy required to manufacture the proposed system from extraction to processing to final construction. There are twelve different components used in the proposed system. Among all, the polycarbonate sheet's contribution is the highest both by weight and embodied energy.

An 18-volt dc fan is fitted at the outlet vent to maintain air circulation inside the proposed system. DC fan can also be put at the inlet vent of the system; however, the number of air exchanges per minute will be



(a) (b) (c)

Fig. 1. (a) Microscopic view of metco material, (b) normal view of metco material, (c) Metco coating on the absorber plate.



(a) (b) (c) (d)

Fig. 2. Sectional view of double pass solar air heater with sensible thermal storage concept.

less than the outlet vent. Due to this reason, in this study, exhaust fan is kept at the outlet vent. A solar panel (single crystalline) with a 40-watt power supply assisted with solar charge controller with battery backup is used to power up dc fan. Due to this arrangement, DC fan get steady power supply despite the fluctuation in solar radiation. Fig. 3 is the representation of the air circulation in the natural and induced forced mode of convection. Due to the application of exhaust fan at the outlet vent, the number of air exchanges in the system under induced forced mode is much higher than the natural mode.

Fig. 4(a-e) represents a detailed experimental setup of the proposed system. Top layer of the proposed system is made of toughen glass to transmit solar radiation with minimum thermal loss to the absorber plate. Thickness of the toughened glass is 3 mm, and its surface area is 1 × 1 m^2. Outer side and bottom body are made of 4 mm thick polycarbonate sheet. Polycarbonates sheet is robust, flexible, and can reduce heat losses. The whole system is mounted on an iron stand.

DPSAHSHC collector consists of two channels, the first channel framed between the glass cover and absorber plate and the second channel formed under the absorber plate and sensible thermal storage bed. Thermal storage material is kept on the bottom of the system. Sides and bottom of the proposed system are insulated with ceramic wool. Area of outlet/inlet vent is 0.0484 m^2. Both vents are in square shape. Space between the glass cover and top absorber plate is 5 cm, while space between the top absorber plate and bottom layer is 3 cm. A polished stainless steel reflector is used in proposed system to receive more solar radiation. Size of reflector is 0.25 m^2 and framed in mild steel iron structure.

Experiments were conducted on three consecutive days starting from 7th - 9th April 2019 with a different configuration of air flow and reflector:

i. 1st day (7th April): On the first day of experiment, proposed system has been experimented in natural convection mode.

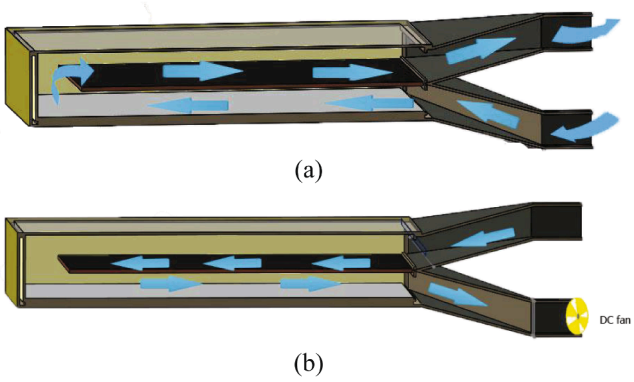


Fig. 3. Air flow arrangement in double pass solar air heater with sensible thermal storage (a) Natural convection mode (b) Induced force convection mode.

ii. 2nd day (8th April): On the second day of the experiment, the fan is fitted in the bottom duct of the system. Thus, airflow allowed entering through the upper duct to perform induced forced convection in the SAH.
iii. 3rd day (9th April): On third day, the experiment was conducted under induced forced convection mode with reflector.

(A) Uncertainty analysis

Every measuring instrument has limitations; none of the measuring devices is 100% accurate. Uncertainty analysis is intended to indicate the possible error in measurement by the instruments. The uncertainty analysis is being calculated with the help of Eqs. (29)–(33). External Uncertainty Percentage Error is given Table 2 [49].

$$W_X = \sqrt{\left(\frac{\partial X}{\partial x_1}W_1\right)^2 + \left(\frac{\partial X}{\partial x_2}W_2\right)^2 + \left(\frac{\partial X}{\partial x_3}W_3\right)^2 + \ldots\ldots + \left(\frac{\partial X}{\partial x_n}W_n\right)^2} \quad (29)$$

$$W_{Temperature} = \sqrt{(W_{thermometer})^2 + (W_{reading})^2} \quad (30)$$

$$W_{Wind} = \sqrt{(W_{anemometer})^2 + (W_{reading})^2} \quad (31)$$

$$W_{SR} = \sqrt{(W_{Pyranometer})^2 + (W_{reading})^2} \quad (32)$$

$$W_{Experimental} = \sqrt{(W_{Wind})^2 + (W_{Temperature})^2 + + (W_{SR})^2} \quad (33)$$

By solving all these equations (Eqs. (30)–(33)) with the help of Table 2, total uncertainty is found to be ±0.2004% which is in actable range.

Methodology

A. Energy analysis

Energy analysis of the proposed system was conducted with the help of experimental data. This analysis is based on the conservation of energy principle. For the present system, the following energy balance equations are used:

$$Q_A = Q_u + Q_{loss} + Q_{st} \quad (1)$$

Now each term can be expressed in the following ways:

Q_u is expressed as the total amount of heat utilized and which is mathematically expressed as follows [35]:

$$Q_u = \dot{m}C_p(T_{out} - T_{in}) \quad (2)$$

where, m is the MFR of the air inside the system. This can be evaluated by the following expression [35]:

$$\dot{m} = \rho V_{ai} S \quad (3)$$

In this system, the primary source of energy is solar radiation. Net solar radiation absorbed by the absorber plate is denoted by Q_A which is evaluated by the following expression [35]:

$$Q_A = A_p(\alpha\tau)I_i \quad (4)$$

In the solar collector, some heat loss occurs on the bottom side and top surface of the system. Therefore, the total amount of heat loss is calculated by the following expression [35]:

$$Q_{loss} = U_{loss}A_p(\bar{T}_p - T_a) \quad (5)$$

U_{loss} is total heat losses coefficient which include losses from the top, bottom and side. Mathematically it is expressed as follow [36]:

$$U_{loss} = U_e + U_b + U_t \quad (6)$$

U_t is calculated as [36]:

$$U_t = \left[1/(h_{r,Pg} + h_{c,Pg}) + 1/(h_{r,ga} + h_w)\right]^{-1} \quad (7)$$

Flow is found to be laminar inside the solar collector. Hence, the Nusselt number is presented as Eq. (8) [20,37-39]:

$$Nu = 0.68 + \frac{(0.67 \times Ra^{1/4})}{\left(1 + (0.492/Pr)^{9/16}\right)^{4/9}} \; [Ra \lesssim 10^9] \quad (8)$$

$$h_{c,pg} = Nu(\lambda/t) \quad (9)$$

$$h_{r,pg} = \left(\frac{\sigma\left(T_g^2 + T_P^2\right)(T_g + T_p)}{\left(\frac{1}{\varepsilon_p} + \frac{1}{\varepsilon_p} - 1\right)}\right) \quad (10)$$

External radiative and convective heat transfer due to the wind velocity evaluated by Hottle and Woertz [20,40] is represented in Eq. (11) and Eq. (12), respectively.

$$h_w = 3.86V + 5.67 \quad (11)$$

$$h_{r,ga} = \varepsilon_g\sigma(T_g^2 + T_{sky}^2)(T_g + T_{sky}) \quad (12)$$

where T_{sky} calculated by the following Eq. (13) [20,40]:

$$T_{sky} = 0.0552T_a^{1.5} \quad (13)$$

Heat loss coefficient for bottom and sidewalls of the SAH are represented in Eq. (14) and Eq. (15) respectively [41].

$$U_b = \lambda_i/\delta_b \quad (14)$$

$$U_e = ((L_1 + L_2)L_3\lambda_i)/(L_1L_2\delta_e) \quad (15)$$

where L_1L_2, and L_3 are length, width and height of the SAH respectively.

According to the first law of thermodynamics, the thermal efficiency of the SAH is the ratio between the useful energy (Q_u) and the solar radiation received on the collector. Thermal efficiency can be calculated by the Eq. (16) [42]:

$$\eta_{Thermal} = Q_u/I_TA_p \quad (16)$$

Dimensionless parameters referred above are given by Eq. (17), Eq. (18) and Eq. (19) [43]:

$$Ra = \frac{g\beta\Delta TL^3}{\alpha v} \quad (17)$$

$$Pr = \frac{v}{\alpha} \quad (18)$$

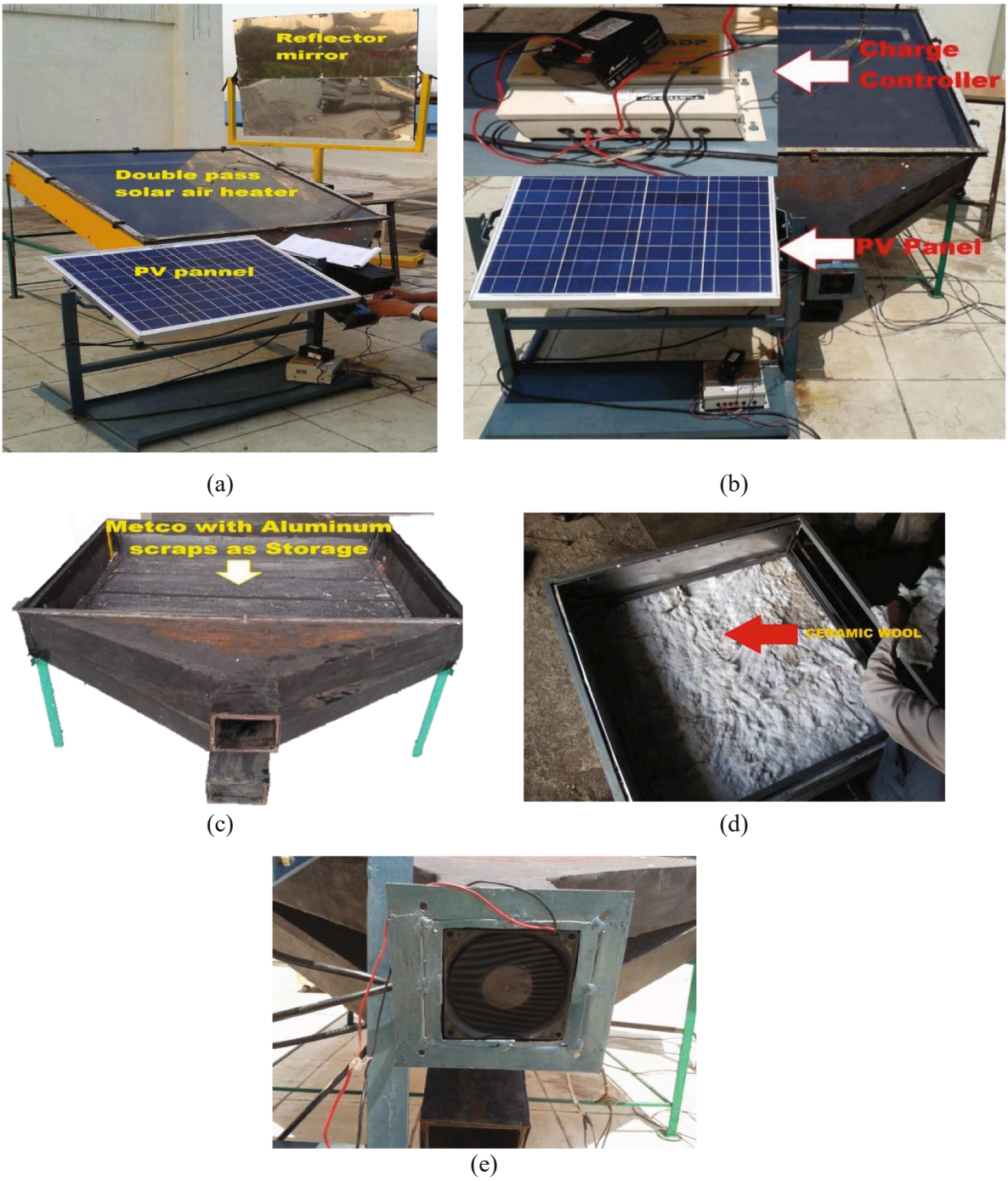


Fig. 4. (a-e): Experimental set up of DPSAH.

$$Re = \frac{VD_h}{v} \tag{19}$$

B. Exergy analysis of the SAH

These are the assumption considered for exergy analysis of the proposed system based on second law of thermodynamics:

i. Steady-state flow operation.
ii. Airflow is an ideal gas and with constant physical property.
iii. Negligible humidity of the air.
iv. No chemical reaction.
v. Although the system is installed with an inclination angle of about 23°, potential and kinetic energy differences are considered negligible.

Exergy balance equation is being developed based on second law of thermodynamics. Exergy at the inlet of the SAH is given as Eq. (20):

$$Ex_{in} = Ex_d + Ex_{out} + Ex_{st} + Ex_{loss} \tag{20}$$

where Ex_{in}, $Ex_d Ex_{out} Ex_{st}$ and Ex_{loss} are the inlet, destroyed, outlet, stored and lost exergy, respectively.

Exergy for inlet airflow is calculated by Eq. (21) [44]:

Table 1
Embodied energy analysis for fabrication of SAH [50].

S. No.	Material	Quantity	Embodied Energy Coefficient	Total (kWh)
1	Glass	5.40 kg	7.28 (kWh/kg)	39.312
2	Polycarbonate sheet	15.60 kg	10.20 (kWh/kg)	159.12
3	Black coating	0.75 m^2	0.28 (kWh/m^2)	0.21
4	Copper Sheet	0.30 kg	19.44 (kWh/kg)	5.832
5	Metco Material	12.00 kg	26.76 (kWh/kg)	321.12
6	Aluminium Scrap	8.00 kg	18.98 (kWh/kg)	151.84
7	Ceramic Wool	1.00 kg	4.05 (kWh/kg)	4.05
8	DC fan			
	(I)Plastic	0.120	19.44 (kWh/kg)	2.332
	(II) Copper wire	0.050	19.61 (kWh/kg)	0.9805
9	Polycrystalline solar cell	0.059 m^2	1130.60 (kWh/m^2)	66.7054
10	Battery			0.046
11	Solar charge controller			0.033
12	Stainless steel	1 kg	55.55(kWh/m^2)	55.55
13	Iron	05 kg	6.94(kWh/m^2)	34.70
	Total Embodied Energy			**841.8309**

Table 2
External uncertainty percentage error

S. No.	Measurement Error	Value
1	Absorber temperature	0.10
2	Air temperature	0.10
3	Relative humidity	0.10
4	Moisture evaporation	0.50
5	Voltage	0.01
6	Current	0.01
7	Wind velocity	0.01
8	Plate Temperature	0.10
Total Percent External Uncertainty		**0.93**

$$Ex_{in}\,(\text{flowing air}) = \dot{m}C_P\left(T_{in} - T_a\left(1 + ln\left(\frac{T_{in}}{T_a}\right)\right)\right) + \dot{m}RT_a ln\left(\frac{P_{in}}{P_a}\right) \quad (21)$$

Exergy at outlet airflow is expressed as Eq. (22) [44]:

$$Ex_{out}(\text{flowingair}) = -\dot{m}C_P(T_{out} - T_a(1 + ln(\frac{T_{out}}{T_a}))) - \dot{m}RT_a ln(\frac{P_{out}}{P_a}) \quad (22)$$

Exergy for absorbed solar radiation can be calculated as Eq. (23) [45]:

$$Ex_{in}\,(\text{solar radiation}) = lr\left[1 - \frac{4}{3}\left(\frac{T_a}{T_s}\right) + \left(\frac{T_a}{T_s}\right)^4\right] \quad (23)$$

where $T_s = 5777$ K.

Lost exergy, Ex_{loss} can be indicated as Eq. (24) [46]:

$$Ex_{loss} = -U_{loss}A_c\left[\left(T_P - T_a\right) - (T_a + 273)ln\frac{(T_P + 273)}{(T_a + 273)}\right] \quad (24)$$

Exergy destruction Ex_d is dependent upon three different terms as:

i. Temperature difference between the absorber plate and sun (shown in Eq. (25)) [47]
ii. Pressure drop at duct as shown in Eq. (26) [47]
iii. Heat transfer to the flowing air from absorber plate to the surface as shown in Eq. (27) [47]

$$Ex_{d(p-s)} = -\eta_0 I_T A_c T_a((\frac{1}{T_p}) - (\frac{1}{T_s})) \quad (25)$$

$$Ex_{d(\Delta P)} = -(\Delta P T_a/\rho)\left(T_a ln\left(\frac{T_{out}}{T_a}\right)\right)\Big/(T_{out} - T_{in}) \quad (26)$$

$$Ex_{d(p-f)} = -\dot{m}C_p T_a\left(\left(ln\left(\frac{T_{out}}{T_{in}}\right)\right) - \left(\frac{T_{out} - T_{in}}{T_p}\right)\right) \quad (27)$$

Daily exergy efficiency of the SAH can be calculated as Eq. (28) [48]:

$$\eta_{Exergy} = (\int Ex_{out}/\int Ex_{in}) \times 100\% \quad (28)$$

Result and discussion

Experiments evaluate the performance of a double pass solar heater in three different modes, namely natural convection, induced forced convection and induced forced convection with reflector mode. Exergy and energy performance of the proposed system are evaluated. The different measuring devices used to generate the data and obtain satisfactory results are listed in Table 3.

Effect of incident solar radiation on temperature difference, thermal and exergy efficiency

Incident solar radiation is the only energy source of the proposed system during three days of experimentation from 7th to 9th April 2019. Average solar radiation noted as 673.33 W/m^2, 671.11 W/m^2 and 665.55 W/m^2, respectively. In natural convection conditions, the solar radiation at 9:00 AM was 400 W/m^2, and the outlet temperature measured was 45 °C. However, the maximum solar radiation found was 950 W/m^2 (at 1:00 PM), and the maximum outlet temperature was 60 °C on that same day. On the second day and third day of the experiment, the solar radiation at 9:00 AM was 410 W/m^2 and 420 W/m^2 and the outlet temperature was 32 °C and 38.2 °C, respectively.

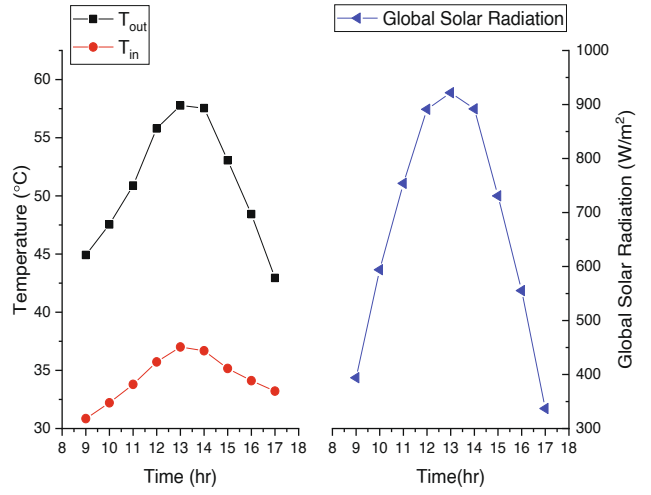
In contrast, the maximum solar radiation observed was 945 W/m^2 and 940 W/m^2 and the maximum outlet temperature was 58.2 °C and 73.2 °C, respectively. Fig. 5(a-c) shows the variation of inlet duct temperature, outlet duct temperature and solar radiation vs time. Higher solar radiation increased the differences between the inlet and outlet temperature and useful energy. Fig. 6 also depicts that the increase in solar radiation leads to increased thermal efficiency and exergy efficiency for all three experiment days. Thus, efficiency of solar air heater depends upon the solar irradiance and area exposed to solar radiation. An increase in solar radiation leads to an increase in temperature, thus reducing the heat loss through the system, thereby increasing the thermal and exergy efficiency. Intensity of solar radiation affects the collector thermal performance, where higher insolation and inlet temperature resulting in higher thermal efficiency.

Effect of temperature difference on thermal and exergy efficiency

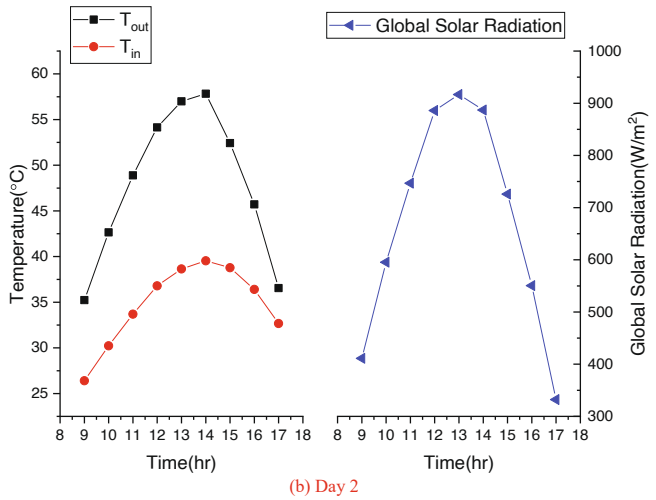
Thermal efficiency is the amount of useful heat energy collected from solar radiation striking on the absorber surface. Exergy efficiency is the adequacy of a system comparative with its performance in reversible conditions. It is characterized as the ratio of the thermal effectiveness of a system compared with a reversible variant of the system. Air temperature that passes through the SAH is determined by subtracting air

Table 3
Technical instruments and measurement device.

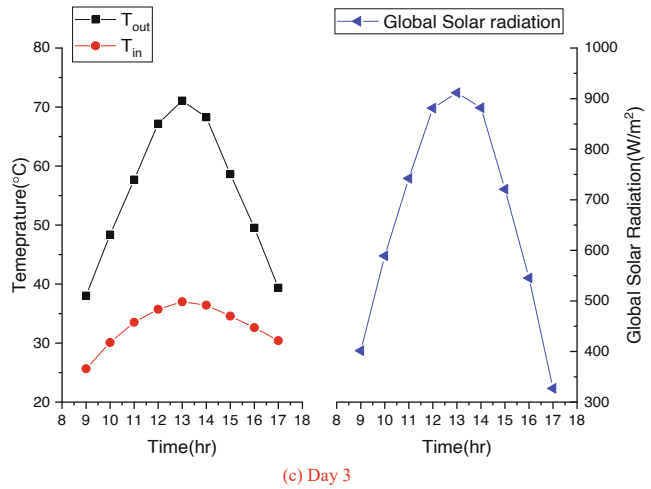
Device	Quantity	Model/Producer	Range	Accuracy
Air blower	2	RS555/Themisto	0–500 RPM	±5 RPM
Anemometer	1	Lutron/AM4202	0.4–30 m/s	±0.1 ms^{-1}
Solar Power Meter	1	CEM/LA-1017	0e1999 W/m^2	±10 W/m^2
Thermocouples	8	K-type/Jain Automation	−200 to 1200 °C	±0.5 °C
Data logger	1	Testo/176T4	−195 to 1200 °C	±3 °C



(a) Day 1



(b) Day 2

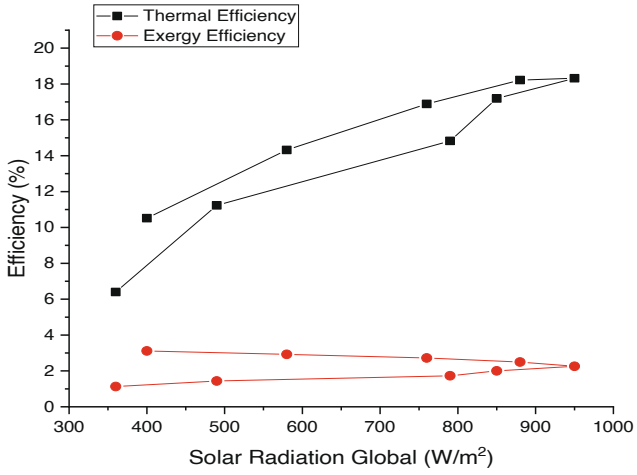


(c) Day 3

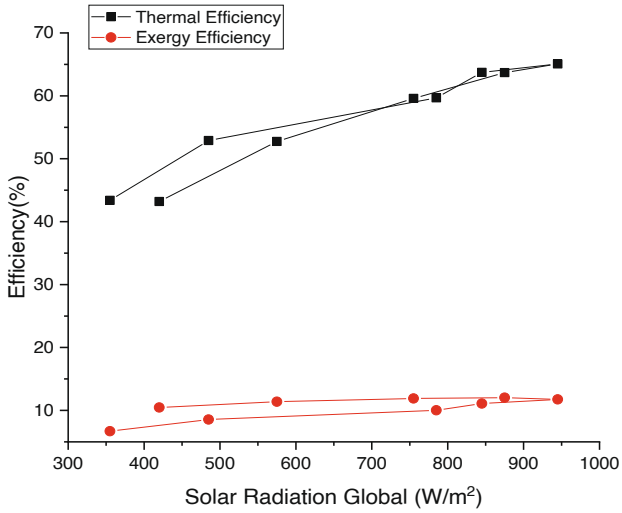
Fig. 5. (a-c): Comparison variation of inlet temperature, outlet temperature and solar radiation with time.

outlet duct temperature with inlet duct temperature, i.e., $\Delta T = T_{out} - T_{in}$. Temperature difference depends upon the operating mode of the DPSAH. The DPSAH under natural convection mode has an average temperature difference of 13.45 °C.

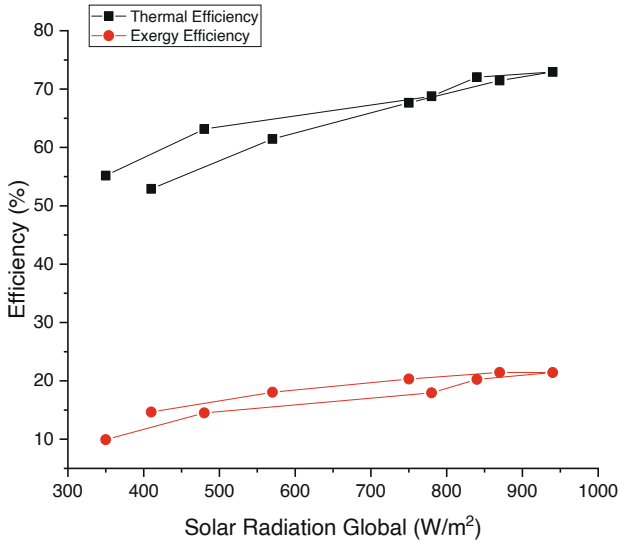
In contrast, when operated in induced forced convection mode with and without reflector, the same SAH has an average temperature



(a) Day 1



(b) Day 2



(c) Day 3

Fig. 6. Effect of solar radiation on thermal efficiency and exergy efficiency.

difference of 22.38 °C and 16.37 °C, respectively. Temperature difference is a major impact on thermal and exergy efficiency. Average thermal efficiency was 56% and 65.05%, respectively, for the 2nd and 3rd-day experiment with the same average MFR of 0.02 kg/s. Average exergy efficiency has been recorded as 13.21% and 16.17%, respectively, for the 2nd and 3rd day experiment. Fig. 7(a-c) compares thermal efficiency and temperature rise on SAH with variation in time. Thermal efficiency increased with temperature difference and further decreases with decreased temperature. Air flowing through the collector carries heat from the sun, naturally raising the air. Thus, cold outside air is heated and delivered, further improving the thermal efficiency. Thermal storage improves the performance of the system by storing and conserving energy to be further utilized by SAH.

Effect of dc fan

The dc fan is used to carry out forced convection in DPSAH. It results in more MFR through the duct of SAH. In the natural mode of operation, no dc fan was required. Hence the average MFR was 0.005 kg/s. It has been observed that both the exergy and thermal efficiency are least when no dc fan is used. In the natural convection mode of operation, the average thermal efficiency was 14.21%, and the exergy efficiency was observed as 3.22%. On 2nd day, fan is fitted at the bottom outlet duct of the DPSAH, resulting in an induced forced air with an average MFR of 0.02 kg/s from the upper duct to the bottom. Induced flow resulted in an average thermal and exergic efficiency of 56% and 13.21%, respectively, on the same SAH configuration. Figs. 8 and 9 show the comparison of thermal and exergy efficiency variation with time in all conditions of natural and induced forced convection mode. Exergy efficiency depends on the inlet and outlet temperature since energy

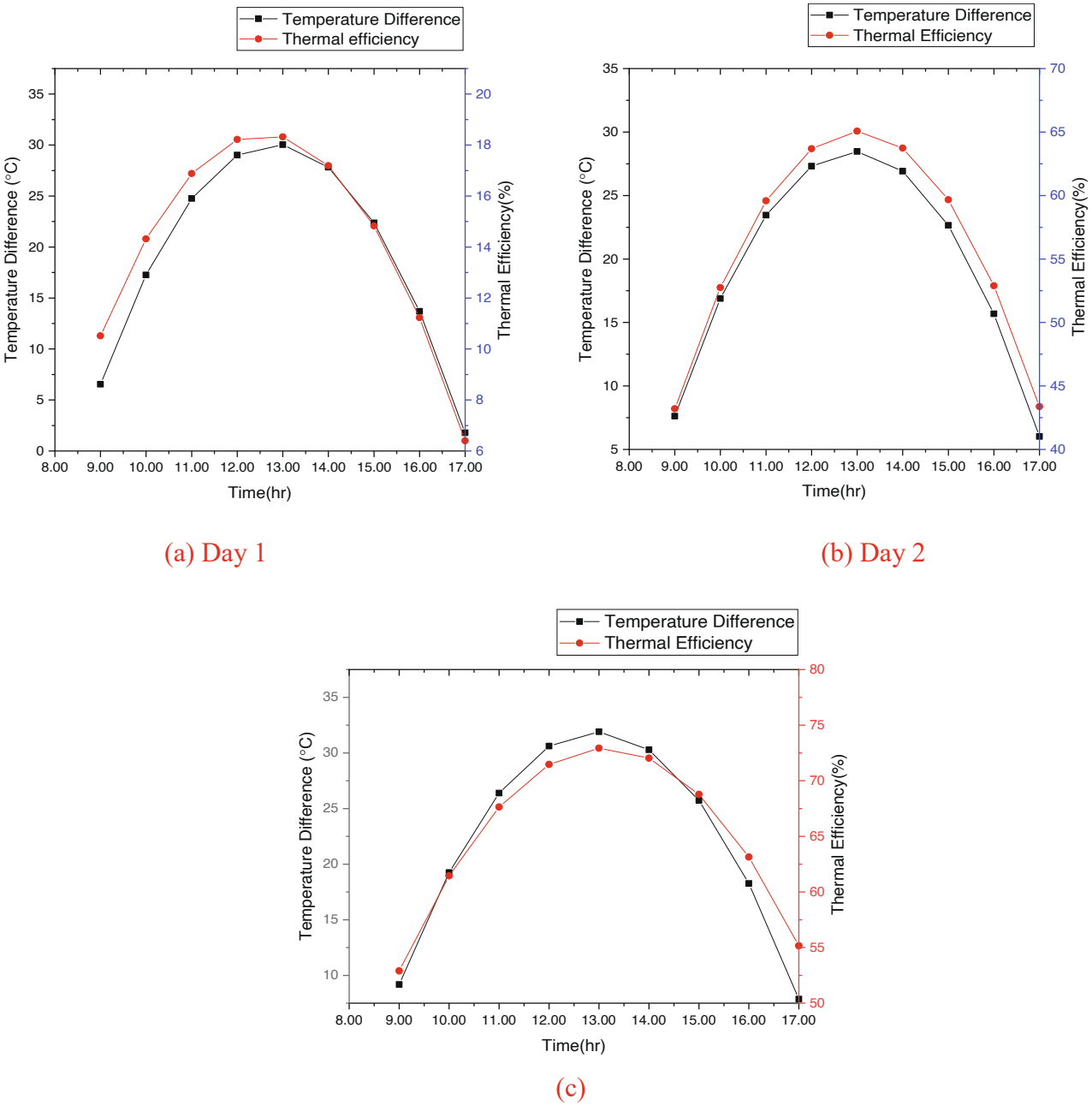


Fig. 7. (a-c): Comparison of temperature difference (black line) and thermal efficiency for time for three days of experiment.

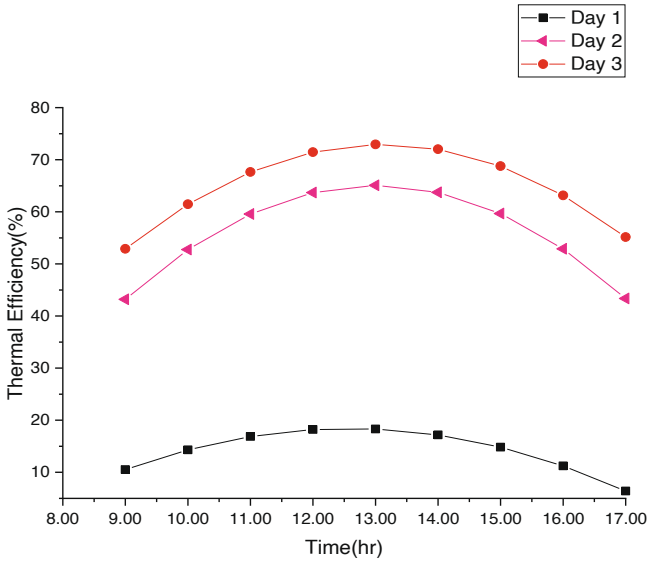


Fig. 8. Thermal Efficiency variation for three days of experiment.

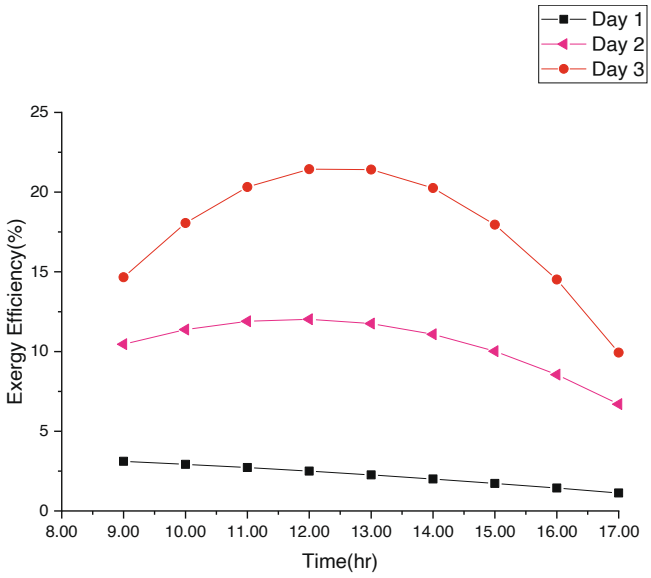


Fig. 9. Exergy efficiency variation for three days of experiment.

increases as the energy utilization temperature reaches inlet energy temperature.

Effect of reflector

Reflector harness solar energy and direct solar radiation to the SAH, leading to increased heat energy through the absorber plate. Reflector assembly was added on the third day of the experiment in induced forced convection mode. Average inlet and outlet temperature found was 32.93 °C and 55.32 °C, respectively. Average thermal and exergy efficiency was evaluated as 65.05% and 16.17%, respectively. Figs. 7 and 8 show the comparison of thermal and exergy efficiency in induced forced convection mode with and without the reflector. Hence the use of reflector with the SAH is fully justified.

Validation

Table 4 presented a comparative study about multi-pass and different solar air heating systems and proposed system designs. Proposed system having thermal efficiency of up to 86.19%; however, double pass air heater with fins cited in Table 4 have thermal efficiency in the range of 61–93%. Therefore, thermal efficiency of solar air heating collector can be enhanced without fins by implementation of reflector and thermal energy storage systems.

Table 4
Summary of selected study about multi-pass and different designs of solar air heating systems.

S. No.	Type	Additional	Thermal Efficiency	Ref.
1	Double Pass	Fins	73.5%	[50]
2	Double Pass	Tubular absorber	86.03%	[51]
3	Double Pass	Finned and wire mesh	93%	[19]
4	Double Pass	Corrugated plate and wire mesh	80%	[52]
5	Double Pass	V-groove absorber	61–77%	[53]
6	Double pass	Aluminum canes	<67%	[31]
7	Double Pass	Mesh absorber	<80.39%	[54]
8	Triple-pass	Thermal energy storage	60%	[55]
9	Triple pass	V groove	<76.39%	[56]
10	Triple-pass	–	67%	[57]
11	Triple-pass and quadruple-pass	V-groove absorber	<91.9%	[58]
12	Quadruple-pass	–	<90.3%	[56]
13	Tube-type	Turbulators	<72.41	[59]
14	Tube-type	Iron mesh modification	<67.69%	[60]
15	Parallel-pass	Plus-shaped perforated baffles	<75.11%	[61]
16	**Double pass**	**Sensible heat storage, Metico coated absorber with reflector**	**<86.19%**	**Present study**

Conclusions

In this work, an innovative DPSAHSHC with reflector is designed and developed and experimentally tested in three consecutive days under similar weather conditions. The main outcomes of this study are summarized as follow:

- A difference in solar radiation is observed for the day, but the difference in each day's daily average solar radiation is comparatively less.
- Outlet temperature of the air flowing out of the DPSAHSHC directly influences incident solar radiation. Average outlet air temperature of 3rd day experiment is 7.52 °C more than that of 2nd day of the experiment, while the inlet air temperature remained almost same. Reflector was used on 3rd-day of the experiment diverted more solar energy on the absorber plate, thus increasing the outlet air temperature.
- Rise in air temperature inside SAH has a direct influence on thermal efficiency. Average thermal efficiency observed was 65.05% and 56% on 3rd and 2nd day. Hence, an increase of 9% thermal efficiency was observed with a temperature difference of 7.52 °C.
- Proposed system has better exergy and thermal efficiency in a bottom outlet than natural convective mode. MFR on 1st day was 0.005 kg/s without dc fan, while on 2nd day the MFR observed was 0.02 kg/s, which is 0.015 kg/s more than 1st day. Average thermal efficiency was 1.79% more in induced forced convection than the natural convection. Exergic efficiency increased by 10% on the 2nd day in comparison to 1st day. The dc fan provides more airflow and reducing heat loss leads to better thermal and exergy performance. Outlet temperature decreases when the airflow rate expands, whereas when MFR increases, the collector efficiency improves, but it causes higher energy extraction. SAH utilizes dark painted material to increment its productivity.

- Reflector in the bottom outlet in induced forced mode SAH had better thermal and exergy performance. Thermal efficiency increased by 9% on 3rd day than 2nd day despite having the same 0.02 kg/s MFR. Likewise, the exergy efficiency was 2.96% more on 3rd day than that on 2nd day.

Experimental and thermal results show the successful design of the proposed system. This system can be used in various low and medium temperature applications such as space heating, desalination, drying, and other industrial and domestic application.

CRediT authorship contribution statement

Om Prakash: Formal analysis, Resources, Validation, Visualization, Writing – review & editing. **Anil Kumar:** Conceptualization, Project administration, Writing – review & editing. **Samsher:** Data curation, Supervision. **Kumaresh Dey:** Investigation. **Ankish Aman:** Methodology, Writing – original draft.

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.

Acknowledgment

Authors would like to express their deepest gratitude towards the Centre for Energy and Environment, Delhi Technological University, Delhi (India), and Birla Institute of Technology, Mesra, Ranchi, India, for the basic facilities provided in preparing this manuscript compilation.

References

[1] Kareem MW, Habib K, Sulaiman SA. Comparative study of single pass collector and double pass solar collector filled with porous media. Asian J Sci Res 2013;6(3):445–55.
[2] Bahrehmand D, Ameri M. Energy and exergy analysis of different solar air collector systems with natural convection. Renewable Energy 2015;74:357–68.
[3] Velmurugan P, Kalaivanan R. Energy and exergy analysis of multi-pass flat plate solar air heater—an analytical approach. Int J Green Energy 2015;12(8):810–20.
[4] Ghiami A, Ghiami S. Comparative study based on energy and exergy analyses of a baffled solar air heater with latent storage collector. Appl Therm Eng 2018;133:797–808.
[5] Chamoli S, Thakur NS. Exergetic performance evaluation of solar air heater having V-down perforated baffles on the absorber plate. J Therm Anal Calorim 2014;117(2):909–23.
[6] El Khadraoui A, Bouadila S, Kooli S, Guizani A, Farhat A. Solar air heater with phase change material: An energy analysis and a comparative study. Appl Therm Eng 2016;107:1057–64.
[7] Chabane F, Moummi N, Benramache S. Experimental study of heat transfer and thermal performance with longitudinal fins of solar air heater. J Adv Res 2014;5(2):183–92.
[8] Zukowski M. Experimental investigations of thermal and flow characteristics of a novel microjet air solar heater. Appl Energy 2015;142:10–20.
[9] Saxena A, Srivastava G, Tirth V. Design and thermal performance evaluation of a novel solar air heater. Renewable Energy 2015;77:501–11.
[10] Siddiqui FR, Elminshawy NAS, Addas MF. Design and performance improvement of a solar desalination system by using solar air heater: Experimental and theoretical approach. Desalination 2016;399:78–87.
[11] Pramanik RN, Sahoo SS, Swain RK, Mohapatra TP, Srivastava AK. Performance analysis of double pass solar air heater with bottom extended surface. Energy Procedia 2017;109:331–7.
[12] Kalaiarasi G, Velraj R, Swami MV. Experimental energy and exergy analysis of a flat plate solar air heater with a new design of integrated sensible heat storage. Energy. 2016;15(111):609–19.
[13] Edalatpour M, Kianifar A, Aryana K, Tiwari GN. Energy, exergy, and cost analyses of a double-glazed solar air heater using phase change material. J Renewable Sustainable Energy 2016;8(1):015101.
[14] Fudholi A, Sopian K, Othman MY, Ruslan MH, Bakhtyar B. Energy analysis and improvement potential of finned double-pass solar collector. Energy Convers Manage 2013;75:234–40.
[15] Srinivas M, Jayaraj S. Energy and exergy analysis of a two pass photovoltaic-thermal (PV/T) air heater. Int J Energy Environ 2013;4:467–80.
[16] Abuşka M. Energy and exergy analysis of solar air heater having new design absorber plate with conical surface. Appl Therm Eng 2018;131:115–24.
[17] Heydari A, Mesgarpour M. Experimental analysis and numerical modeling of solar air heater with helical flow path. Sol Energy 2018;162:278–88.
[18] Salih SM, Jalil JM, Najim SE. Experimental and numerical analysis of double-pass solar air heater utilizing multiple capsules PCM. Renewable Energy 2019;143:1053–66.
[19] Singh S, Dhruw L, Chander S. Experimental investigation of a double pass converging finned wire mesh packed bed solar air heater. J Storage Mater 2019;21:713–23.
[20] Rhee SJ, Edwards DK. Laminar entrance flow in a flat plate duct with asymmetric suction and heating. Num Heat Transf 1981;4(1):85–100.
[21] Prasad BN, Sah GN. Plate temperature and heat transfer characteristics of artificially roughened solar air heater. Energy Procedia 2014;62:256–69.
[22] Manjunath MS, Karanth KV, Sharma NY. Numerical analysis of the influence of spherical turbulence generators on heat transfer enhancement of flat plate solar air heater. Energy 2017;121:616–30.
[23] Thakur DS, Khan MK, Pathak M. Performance evaluation of solar air heater with novel hyperbolic rib geometry. Renewable Energy 2017;1(105):786–97.
[24] Jaurker AR, Saini JS, Gandhi BK. Heat transfer and friction characteristics of rectangular solar air heater duct using rib-grooved artificial roughness. Sol Energy 2006;80(8):895–907.
[25] Singh S, Dhruw L, Chander S. Experimental investigation of a double pass converging finned wire mesh packed bed solar air heater. J Storage Mater 2019;1(21):713–23.
[26] Tyagi RK, Ranjan R, Kishore K. Performance studies on flat plate solar air heater subjected to various flow patterns. Appl Solar Energy 2014;50(2):98–102.
[27] Pfister H, Ralston T, Kim SW. A novel gridded solar air heater and an investigation of its conversion efficiency. Sol Energy 2016;136:560–70.
[28] Krishnananth SS, Kalidasa Murugavel K. Experimental study on double pass solar air heater with thermal energy storage. J King Saud Univ-Eng Sci 2013;25(2):135–40.
[29] Kumar A, Layek A. Nusselt number and friction factor correlation of solar air heater having twisted-rib roughness on absorber plate. Renewable Energy 2019;130:687–99.
[30] Baissi MT, Brima A, Aoues K, Khanniche R, Moummi N. Thermal behavior in a solar air heater channel roughened with delta-shaped vortex generators. Appl Therm Eng 2020;165:113563.
[31] Ozgen F, Esen M, Esen H. Experimental investigation of thermal performance of a double-flow solar air heater having aluminium cans. Renewable Energy 2009;34(11):2391–8.
[32] Jha RK, Choudhury C, Garg HP, Zaidi ZH. Performance prediction of a solar heated house. Energy Convers Manage 1992;33(4):263–73.
[33] Kumar R, Goel V, Kumar A. Investigation of heat transfer augmentation and friction factor in triangular duct solar air heater due to forward facing chamfered rectangular ribs: A CFD based analysis. Renewable Energy 2018;115:824–35.
[34] Jia B, Liu F, Wang Da. Experimental study on the performance of spiral solar air heater. Sol Energy 2019;182:16–21.
[35] Nidhul K, Yadav AK, Anish S, Arunachala UC. Efficient design of an artificially roughened solar air heater with semi-cylindrical side walls: CFD and exergy analysis. Sol Energy 2020;207:289–304.
[36] Esen H. Experimental energy and exergy analysis of a double-flow solar air heater having different obstacles on absorber plates. Build Environ 2008;43(6):1046–54.
[37] Bouadila S, Lazaar M, Skouri S, Kooli S, Farhat A. Energy and exergy analysis of a new solar air heater with latent storage energy. Int J Hydrogen Energy 2014;39(27):15266–74.
[38] Kurtbas İ, Durmuş A. Efficiency and exergy analysis of a new solar air heater. Renewable Energy 2004;29(9):1489–501.
[39] Abuşka M. Energy and exergy analysis of solar air heater having new design absorber plate with conical surface. Appl Therm Eng 2018;131:115–24.
[40] Acır A, Canlı ME, Ata İ, Çakıroğlu R. Parametric optimization of energy and exergy analyses of a novel solar air heater with grey relational analysis. Appl Therm Eng 2017;122:330–8.
[41] Matheswaran MM, Arjunan TV, Somasundaram D. Analytical investigation of solar air heater with jet impingement using energy and exergy analysis. Sol Energy 2018;161:25–37.
[42] Alta D, Bilgili E, Ertekin C, Yaldiz O. Experimental investigation of three different solar air heaters: Energy and exergy analyses. Appl Energy 2010;87(10):2953–73.
[43] Ghaderian J, Azwadi CN, Mohammed H. Modelling of energy and exergy analysis for a double-pass solar air heater system. J Adv Res Fluid Mech Therm Sci 2015;16:15–32.
[44] Koroneos C, Nanaki E, Xydis G. Solar air conditioning systems and their applicability—an exergy approach. Resour Conserv Recycl 2010;55(1):74–82.
[45] Kumar A, Saini RP, Saini JS. A review of thermohydraulic performance of artificially roughened solar air heaters. Renew Sustain Energy Rev 2014;1(37):100–22.
[46] Prakash O, Kumar A, editors. Solar drying technology: concept, design, testing, modeling, economics, and environment. Springer 2017;29.
[47] Menni Y, Azzi A, Chamkha A. The solar air channels: comparative analysis, introduction of arc-shaped fins to improve the thermal transfer. J Appl Comput Mech 2019;5(4):616–26.
[48] Öztürk HH. Experimental evaluation of energy and exergy efficiency of a seasonal latent heat storage system for greenhouse heating. Energy Convers Manage 2005;46(9-10):1523–42.
[49] Minh Phu N, Thanh Luan N. A review of energy and exergy analyses of a roughened solar air heater. J Adv Res Fluid Mech Therm Sci 2020;77(2):160–75.

[50] Salih MMM, Alomar OR, Ali FA, Abd HM. An experimental investigation of a double pass solar air heater performance: A comparison between natural and forced air circulation processes. Sol Energy 2019;193:184–94.
[51] Abo-Elfadl S, Hassan H, El-Dosoky MF. Study of the performance of double pass solar air heater of a new designed absorber: An experimental work. Sol Energy 2020;198:479–89.
[52] Singh S. Experimental and numerical investigations of a single and double pass porous serpentine wavy wiremesh packed bed solar air heater. Renewable Energy 2020;145:1361–87.
[53] Bashria A, Yousef A, Adam NM, Sopian K, Zaharim A, Alghoul M. Analysis of single and double passes v-grooves solar collector with and without porous media. Int J Energy Environ 2007;2:109–14.
[54] Güler HÖzge, Sözen A, Tuncer ADoğuş, Afshari F, Khanlari A, Şirin C, et al. Experimental and CFD survey of indirect solar dryer modified with low-cost iron mesh. Sol Energy 2020;197:371–84.
[55] Kesavan S, Arjunan TV. Experimental study on triple pass solar air heater with thermal energy storage for drying mint leaves. Energy Technol Policy 2018;14(1):34–48.
[56] Tuncer ADoğuş, Sözen A, Khanlari A, Amini A, Şirin C. Thermal performance analysis of a quadruple-pass solar air collector assisted pilot-scale greenhouse dryer. Sol Energy 2020;203:304–16.
[57] Orbegoso EM, Saavedra R, Madrid MD, RL.. Numerical characterization of one-step and three-step solar air heating collectors used for cocoa bean solar drying. Environ Manage 2017;203:1080–94.
[58] Khanlari A, Sözen A, Afshari F, Şirin C, Tuncer ADoğuş, Gungor A. Drying municipal sewage sludge with v-groove triple-pass and quadruple-pass solar air heaters along with testing of a solar absorber drying chamber. Sci Total Environ 2020;709:136198.
[59] Afshari F, Sözen A, Khanlari A, Tuncer ADoğuş, Şirin C. Effect of turbulator modifications on the thermal performance of cost-effective alternative solar air heater. Renewable Energy 2020;158:297–310.
[60] Sözen A, Şirin C, Khanlari A, Tuncer AD, Gürbüz EY. Thermal performance enhancement of tube-type alternative indirect solar dryer with iron mesh modification. Sol Energy 2020;207:1269–81.
[61] Khanlari A, Güler HÖzge, Tuncer ADoğuş, Şirin C, Bilge YaşarC, Yılmaz Y, et al. Experimental and numerical study of the effect of integrating plus-shaped perforated baffles to solar air collector in drying application. Renewable Energy 2020;145:1677–92.

EXPLORING THE RELATIONSHIP BETWEEN ETHICS AND KNOWLEDGE CULTURE: A CONCEPTUAL FRAMEWORK FOR SUCCESSFUL ORGANIZATIONS

Nishant Gaur, Delhi Technological University
Vikas Gupta, Delhi Technological University

ABSTRACT

The paper studies the impact of ethics on the knowledge culture of an organization. The objective of the paper is to propose the linkage between ethics and knowledge culture. It is postulated in this paper that ethical norms pertaining in organization influence individual behaviour for creation, sharing, storage and application of knowledge. The firm can integrate ethical values in knowledge culture of an organization and can have a long-term impact. Future studies can focus on empirical research and testing of the proposed conceptual model. The main limitation arises while generalizing conclusion on linkage between ethics and knowledge culture studied through literature. The paper offers an original conceptual framework integrating ethics and knowledge culture. This model proposes a roadmap for managers to implement and practice ethical conduct while creating, sharing, storing and applying knowledge.

Keywords: Ethics, Knowledge Management, Knowledge Culture, Organizational Culture, Knowledge Creation, Knowledge Sharing, Knowledge Storage, Knowledge Application

INTRODUCTION

The technical advancement facilitated by modern science lead to a progression from an industrial economy to a knowledge-based economy. The knowledge-based economy has been labeled as a network society because of information availability and accessibility from global networks (Manuel Castells, 2004). The knowledge-based economy is the economy of knowledge sharing where knowledge sharing affirms to be a vital component contributing towards effectiveness of knowledge culture (Styhre, 2002). Ethics is the key driver to achieve a successful organization (Peyman Akhavan, 2013). The knowledge culture accelerates knowledge exchange and knowledge creation within the organizational framework. Knowledge is owned and maintained by the organization and is considered as a public good (Wasko & Faraj, 2000). Knowledge culture aggregates individual knowledge into collective knowledge through open discussion, interpersonal interaction, collaboration and knowledge exchange. Some organizations focus on organizational knowledge neglecting the concern of individual's self-interest. The individualistic as well as collectivistic approach for knowledge acquisition and retention may lead to ethical conflicts in the organization. The organizations should emphasize on creating an effective balance between individualistic and collectivistic approach towards knowledge acquisition and retention. Knowledge culture should support and leverage organizational effectiveness but not contribute to unethical practice by retrenching the members who offer their ideas and innovations regularly.

The research gap observed during literature review recognizes that there is no research which studies integrated framework of ethics that facilitates all KM processes which includes knowledge creation, knowledge sharing, knowledge storage and knowledge application. There is deliberation on impact of ethics on knowledge creation in literature but there is no research on ethical components impacting all KM processes. This paper is an attempt to identify the various

Fractional Oscillator - Harmonic Oscillator with Memory Effects

Vishwamittar[a], Yashika Taneja[b] and Nipun Ahuja[b]

[a] *Department of Physics, Panjab University, Chandigarh – 160014.*
[b] *Department of Mechanical Engineering, Delhi Technological University, Delhi – 110042.*
email addresses: vm@pu,ac.in, taneja.ysh@gmail.com, and ahuja.nipun@gmail.com.

Abstract

The importance of fractional time-derivative to take care of memory effects has been brought out by considering the example of a simple oscillator.

1. Introduction

The calculus, to which students are exposed to at school level, deals with derivatives of integer order, expressed as $\frac{d^n y(x)}{dx^n}$. Natural question that can come to one's mind is: Will the idea of derivative be valid if it is extended to non-integer value of n? In fact, such a query formed part of correspondence between L'Hospital, one of the main developers of calculus, and Leibniz (one of the inventors of this branch of mathematics), nearly 325 years back in 1695. In one letter, the former sought latter's opinion about the meaning of $\frac{d^n y(x)}{dx^n}$ if n had fractional value $1/2$. It was on Sept 30, 1695, that Leibniz sent the answer adding that 'this was an apparent paradox', and made a prophetic statement that 'one day, useful consequences would be drawn from this'. Initially, only occasional reference was made to fractional derivatives by some mathematicians and serious efforts to develop the subject in a logical manner began in the first half of the 19th century, and this got contributions from many prominent mathematicians till the middle of the 20th century. Interestingly, many of them introduced their own formulae for fractional (and even any arbitrary real non-integer or complex order) derivative and integral, with strange properties, following different approaches.

A noteworthy development from the point-of-view of future applications of the subject, particularly in physical and engineering problems, occurred in 1967 when Caputo put forward his definition of the fractional time-derivative (FD) of order α as

$$ {}_aD^{\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \left[\int_a^t (t-t')^{(n-\alpha)-1} \frac{d^n f(t')}{dt'^n} dt' \right] ; \qquad n \in \mathbb{N} \quad \text{and} \quad n-1 < \alpha \leq n. \qquad (1) $$

Here, $\Gamma(z)$ is Euler's gamma function defined as $\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$, where z is a complex number such that its real part is positive. This satisfies the property $\Gamma(z+1) = z\Gamma(z)$, and in particular, $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(n+1) = n!$. It is important to note that FD involves integral of $\frac{d^n f(t')}{dt'^n}$ getting contribution with weight $(t-t')^{n-\alpha-1}$from initial time a to the instant of

observation t. In other words, the FD at time t is collection of weighted values in the earlier times so that it includes the influence of the past behaviour of the function as if it had some memory to remember the states through which it had passed and, therefore, it is nonlocal in nature. This is in contrast with conventional integer-order derivatives which are evaluated only for the value of t at the time of interest making it local. Thus, fractional order derivatives are useful in describing time-evolution of physical systems with memory effects of all types. It may be pointed out that the value of the order of derivative provides a measure of the memory of the system. However, it must be emphasized that the (physical) system described by such a function does not have any hidden intelligence but only delayed effects of collisions, interactions, causative forces, etc.

The concepts and methods of fractional calculus (FC) have been extensively used in the development of classical mechanics (Newtonian, and the latter sophisticated Lagrangian and Hamiltonian formulations), electromagnetics, nonrelativistic as well as relativistic quantum mechanics, quantum field theory, and statistical mechanics leading to their wide-ranging applications in atomic, molecular, nuclear, particle, and metal cluster physics. As a mathematical tool, FC also finds numerous applications in the description of anomalous diffusion in porous media, dense polymer solutions, composite heterogeneous films, etc; in understanding behaviour of viscoelastic materials, colloidal systems, magneto-rheological fluids, and amorphous semiconductors; in wave propagation in media with long-range interaction; in statistical mechanics of systems with long-range power-law interactions; in signal and image processing; in analysis and synthesis of different types of control systems; in the description of physics of biological structures and living organisms (e.g., DNA dynamics, protein folding, modeling for neuron activity, etc); in developing models for description of environment; and in mathematical modelling of economic processes and population growth with memory effects.
At the elementary physics level, many workers have used FC to discuss topics like Newton's second law of motion and equations of motion, a body falling under gravity in a viscous medium, projectile motion, simple oscillator, Fourier law of thermal conduction, Newton's law of cooling, Ohm's law and RC, RL circuits, and physics of decay / growth / relaxation processes. In this article, we describe harmonic oscillator in the framework of FC and bring out the memory effects in it.

2. Fractional Oscillator

A particle or an object executing repeated back and forth motion after being displaced from its equilibrium position in such a way that the restoring force is always directed towards the point of equilibrium and its magnitude is proportional to the displacement, is called a mechanical harmonic oscillator (HO). In one dimension, for a body of mass m having instantaneous displacement $x(t)$ with equilibrium or centre point at $x = 0$, such a motion (known as simple harmonic motion) is described by the differential equation,

$$m\frac{d^2x(t)}{dt^2} = -kx(t)\ . \qquad (2)$$

Here, the negative sign takes care of the fact that the directions of force and the displacement are opposite to each other. The parameter k, the force per unit displacement, is called force constant. In case, the restoring force depends on higher powers of displacement, then the oscillator is said to be anharmonic. If the oscillating system experiences a friction-like force (which is generally always present in real systems) that causes continuous dissipation of energy, then it is referred to as a damped oscillator. Furthermore, an oscillator subjected to a time dependent external force is known as a forced or driven oscillator. The basic equation (2) is accordingly amended to accommodate relevant additional terms.

In general, any system, which need not necessarily be a material object and may be something like electric or magnetic field, that can be described by an expression analogous to that in Eq. (2) or its modified version, is said to be oscillatory. Besides everyday life examples of swings, cradles, vehicle shock absorbers, musical instruments, process of hearing, etc., considered in its general sense, an oscillator (classical or quantum one- or three- dimensional and their further modifications) is ubiquitous in physics and finds wide-range applications in developing theoretical models for different phenomena in almost all branches of physics. Some typical examples are simple, compound and torsional pendulums; spring-mass system; vibrating tuning forks, strings and gas columns; electric LC and LCR circuits; electronic oscillators; vibration of atoms in molecules; vibration of lattice atoms and molecules in solids leading to creation of phonons; modelling nuclear collective motion; basics of quantum field theory; and so on.

Replacing the ordinary second-order derivative by a Caputo FD of arbitrary order α $(1 < \alpha \leq 2)$, Eq (2) can be written as

$$m_F D^{\alpha} x(t) = -kx(t), \tag{3}$$

This is fractional or generalized form of the equation of motion of a HO. To keep the meaning of $x(t)$ and k same as in Eq. (2), m_F must have dimension $MT^{\alpha-2}$. Mass parameter m_F becomes mass m for index value $\alpha = 2$. We solve Eq. (3), using Laplace transform method, wherein a function $f(t)$ is converted into a function $F(s)$ of complex frequency s through

$$F(s) \equiv \mathcal{L}\{f(t)\} = \int_0^{\infty} e^{-st} f(t) dt. \tag{4}$$

Taking LT of both sides of Eq. (3), using the fact that

$$\mathcal{L}\{D^{\alpha} f(t)\} = s^{\alpha} F(s) - s^{\alpha-1} f(0) - s^{\alpha-2} f^{(1)}(0), \tag{5}$$

with $f^{(1)}(0)$ as the first derivative of $f(t)$ at $t = 0$; we get

$$X(s) = \frac{x(0) s^{\alpha-1}}{\left(s^{\alpha} + \frac{k}{m_F}\right)} + \frac{v(0) s^{\alpha-2}}{(s^{\alpha} + \frac{k}{m_F})}. \tag{6}$$

Inverse LT of Eq. (6) gives analytical expression for instantaneous displacement as

$$x(t) = x(0) E_{\alpha,1}\left(-\frac{k}{m_F} t^{\alpha}\right) + v(0) t E_{\alpha,2}\left(-\frac{k}{m_F} t^{\alpha}\right). \tag{7}$$

where $E_{\alpha,1}$ and $E_{\alpha,2}$ are two-parameter Mittag-Leffler functions, in general, defined by

$$E_{\rho_1,\rho_2}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(n\rho_1 + \rho_2)}, \qquad (\rho_1, \rho_2 > 0), \tag{8}$$

In particular,

$$E_{2,1}(-z^2) = \sum_{n=0}^{\infty} \frac{(-z^2)^n}{\Gamma(2n+1)} = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} = \cos\,(z), \tag{9}$$

and

$$E_{2,2}(-z^2) = \sum_{n=0}^{\infty} \frac{(-z^2)^n}{\Gamma(2n+2)} = \frac{1}{z}\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1} = \frac{\sin\,(z)}{z}. \tag{10}$$

It may be mentioned that

$$\mathcal{L}\left\{t^{\rho_1 l+\rho_2-1} E^{(l)}_{\rho_1,\rho_2}(\pm a t^{\rho_1})\right\} = \frac{l!\, s^{\rho_1-\rho_2}}{(s^{\rho_1} \mp a)^{l+1}}, \qquad Re(s) > |a|^{1/\rho_1} \tag{11}$$

where $E^{(l)}_{\rho_1,\rho_2}(y) = \frac{d^l}{dy^l} E_{\rho_1,\rho_2}(y)$.

Coming to the expression for $x(t)$, we get only one of the two terms if either $v(0) = 0$ or $x(0) = 0$. Note that the condition $v(0) = 0$, $x(0) \neq 0$ implies that the oscillator is at the turning point at $t = 0$ and has amplitude $x(0)$. In contrast, the initial condition $x(0) = 0$, $v(0) \neq 0$ corresponds to the situation that the oscillator starts its motion from the equilibrium point at the origin with velocity $v(0)$. Assuming the first condition, we get

$$x(t) = x(0) E_{\alpha,1}\left(-\frac{k}{m_F} t^{\alpha}\right). \tag{12}$$

It is worth mentioning that in the framework of FC, linear momentum at any time t is expressed in terms of position at that time by

$$p(t) = m_F \frac{d^{\alpha/2} x(t)}{dt^{\alpha/2}}, \qquad 1 < \alpha \leq 2; \tag{13}$$

Note that for $\alpha = 2$, this reduces to the conventional expression for instantaneous momentum. Using Eq. (12), we have

$$p(t) = m_F \frac{d^{\frac{\alpha}{2}}\left[x(0) E_{\alpha,1}\left(-\frac{k}{m_F} t^{\alpha}\right)\right]}{dt^{\frac{\alpha}{2}}} = m_F x(0) \sum_{n=0}^{\infty} \frac{\left(-\frac{k}{m_F}\right)^n}{\Gamma(n\beta+1)} D^{\frac{\alpha}{2}} t^{\alpha n}$$

$$= m_F x(0) \sum_{n=0}^{\infty} (-\frac{k}{m_F})^n \frac{t^{\alpha(n-1/2)}}{\Gamma(n\alpha+1-\alpha/2)} = -m_F x(0) \frac{k}{m_F} t^{\alpha/2} E_{\alpha,1+\alpha/2}\left(-\frac{k}{m_F} t^{\alpha}\right). \tag{14}$$

In view of this, the total mechanical energy of the fractional oscillator at time t, viz. $\mathbb{E}(t) = (p^2(t)/2m_F) + (kx^2(t)/2)$, is given by

$$\mathbb{E}(t) = \frac{1}{2} m_F x^2(0) (\frac{k}{m_F})^2 t^{\alpha} [E_{\alpha,1+\alpha/2}\left(-\frac{k}{m_F} t^{\alpha}\right)]^2 + \frac{1}{2} k x^2(0) [E_{\alpha,1}\left(-\frac{k}{m_F} t^{\alpha}\right)]^2. \tag{15}$$

Obviously, $\mathbb{E}(t)$ varies with time implying that the total mechanical energy of a fractional oscillator is not a constant or is not conserved.

Furthermore, for maximum allowed value of α, i.e., $\alpha = 2$, using Eq. (9), we get from Eq. (12),

$$x(t) = x(0) \cos(\sqrt{\frac{k}{m}}\, t). \tag{16}$$

which gives displacement of a conventional HO with natural angular frequency $\omega = \sqrt{\frac{k}{m}}$. Next, substituting $\alpha = 2$ into Eqs. (14) and (15), we get using Eqs. (10) and (9),

$$p(t) = -m\, x(0) \frac{k}{m} t E_{2,2}\left(-\frac{k}{m_F} t^2\right) = -m\, x(0) \sqrt{\frac{k}{m}} \sin\,(\sqrt{\frac{k}{m}}\, t), \tag{17}$$

And

$$\mathbb{E}(t) = \frac{1}{2}mx^2(0)(\frac{k}{m})^2t^2[E_{2,2}\left(-\frac{k}{m}t^2\right)]^2 + \frac{1}{2}kx^2(0)[E_{2,1}\left(-\frac{k}{m}t^2\right)]^2 = \frac{1}{2}kx^2(0). \quad (18)$$

Note that expression in Eq. (17) is the same as is obtained by taking $p(t) = m\frac{dx(t)}{dt}$. Also, Eq. (18) brings out conservation of mechanical energy for the conventional HO, which is in contrast with the finding in equation (15) for a fractional oscillator.

The time dependence of $x(t)$ and $\mathbb{E}(t)$ for an oscillator with $\sqrt{\frac{k}{m_F}}$ = 1.0 rad $s^{-\alpha/2}$, as given by Eqs. (12) and (15), respectively, for different α values are projected in *Figures* 1 and 2. The plots of $\mathbb{E}(t)$ for $\alpha = 1.2$ and 1.5 are quite close to each other so that only thr former is shown in the figure. A look at *Figure* 1 reveals that for $\alpha = 2$, the plot represents a cosine variation so that displacement (of the standard oscillator) is periodic and has same amplitude. However, for $1 < \alpha < 2$, the system makes a finite number of oscillations with decreasing amplitudes and finally decays to $x(t)$ = 0 position. This attenuation of oscillations is similar to that observed in a damped HO. Moreover, this effect becomes more prominent as α decreases; the number of zeros in the $x(t)$ plots is quite small for α close to 1 (implying quite rapidly damped oscillations). In the case of total mechanical energy (*Figure* 2), the graph for $\alpha = 2$ corresponds to constancy of $\mathbb{E}(t)$ and, hence, conservation of energy, while the $\alpha < 2$ curves exhibit decrease in energy with time as for the damped or dissipative motion. Once again, the decrease in the curve is faster for lower values of α. Thus, a fractional oscillator ($1 < \alpha < 2$) behaves as a damped oscillator even though there is no resistive medium and this damping is absent for $\alpha = 2$, which corresponds to a simple HO. In other words, the past-history of the oscillator with $1 < \alpha < 2$ influences its motion by producing dampening effect as if it were interacting with itself – **memory effect**. Accordingly, it is usually referred to as fractional-order intrinsic damping. Since the effect becomes more pronounced with decrease in α, the order of fractional derivative provides a measure of memory effects. With a view to look at this feature more minutely, we recall that damping and hence dissipation in a standard HO is obtained when a velocity-dependent term ($-\gamma\frac{dx(t)}{dt}$) is added to the restoring force term on the right side of Eq. (2) so that this is external in nature. It is indeed interesting to note that though we started with zeroth order (i.e., no) derivative in the restoring force in Eq. (3), we have also got an effect associated with first-order derivative of $x(t)$. This means that the fractional derivative on the left side of this equation has led to a behaviour which is a mixture of zeroth order and first order derivatives. The inclusion of (ubiquitous) dissipation effect in a natural way in the description makes it closer to reality, and, thus, brings out the importance of FC as a logical and comprehensive tool for modeling real systems by incorporating memory effects.

It is worth pointing out that Rekhviashvili et al [6] studied vibrations of a free piezoelectric plate under standard laboratory conditions and found that their experimental data was best accounted for with α = 1.998 (reasonably small memory effect).

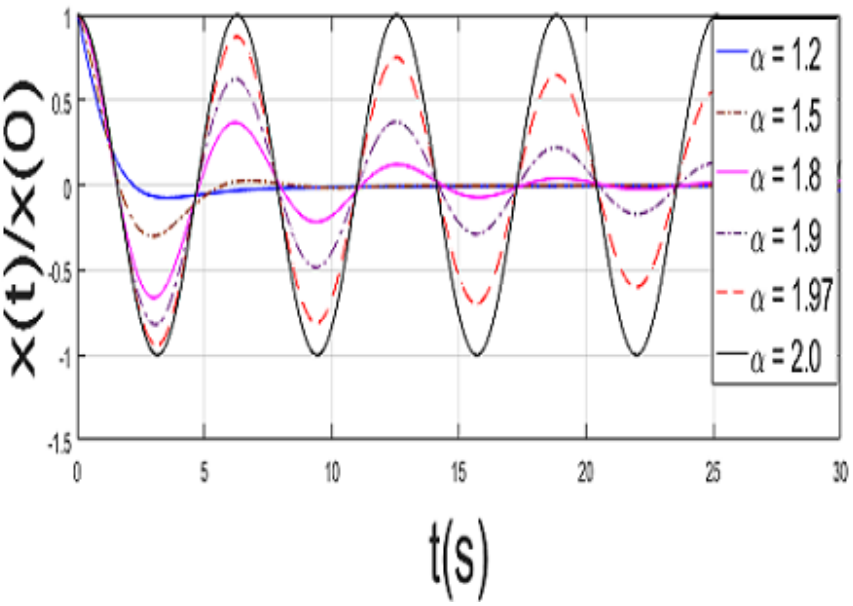


Figure 1. Graphical depiction of position of oscillator ($\sqrt{\frac{k}{m_F}}$ = 1.0 rad $s^{-\alpha/2}$) as function of time for different values of α.

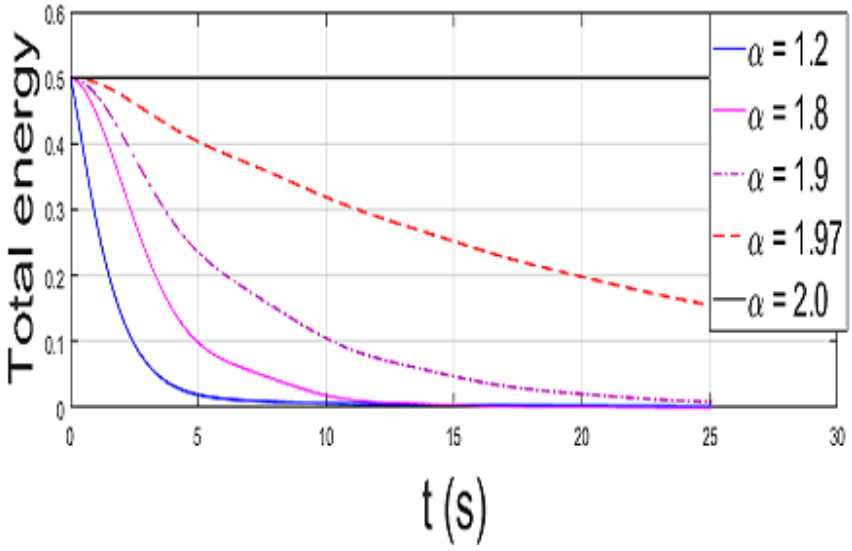


Figure 2. Dependence of total mechanical energy of oscillator ($\sqrt{\frac{k}{m_F}}$ = 1.0 rad $s^{-\alpha/2}$) on time for different values of α.

Acknowledgement

The article is dedicated with immense affection to dear Vibhuti, Shreya and Shruti.

Suggested Reading

[1] I Podlubny, *Fractional Differential Equations*, Academic, San Diego, 1999.

[2] R Herrmann, *Fractional Calculus – An Introduction for Physicists*, World Scientific, Singapore, 2014.

[3] N Laskin. *Fractional Quantum Mechanics*, World Scientific, Singapore, 2018.

[4] H Sun, Y Zhang, D Baleanu, W Chen, and Y Chen, A New Collection of Real World Applications of Fractional Calculus in Science and Engineering, *Commun. Nonlinear Sci. Numer. Simul.*, Vol. 64, pp. 213-231 (2018).

[5] B N Narahari Achar, J W Hanneken, T Enck, and T Clarke, Dynamics of the Fractional Oscillator, *Physica A*, Vol. 297, pp. 361–367 (2001).

[6] S Rekhviashvilli, A Pskhu, P Agarwal, and S Jain, Application of the Fractional Oscillator Model to Describe Damped Vibrations, Turk. J. Phys. Vol. 43, pp. 236 – 242 (2019).

Neural Computing and Applications
https://doi.org/10.1007/s00521-021-06677-1

ORIGINAL ARTICLE

Fuzzy k-plane clustering method with local spatial information for segmentation of human brain MRI image

Puneet Kumar[1] · Dhirendra Kumar[2] · Ramesh Kumar Agrawal[1]

Received: 4 May 2021 / Accepted: 27 October 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract
Human brain MRI images are complex, and matters present in the brain exhibit non-spherical shape. There exits uncertainty in the overlapping structure of brain tissue, i.e. a lack of distinctness in the class definition. Soft clustering methods can efficiently handle the uncertainty, and plane-based clustering methods are found to be more efficient for non-spherical shape data. Fuzzy k-plane clustering (FkPC) method is a soft plane-based clustering algorithms that can handle the uncertainty in medical images, but its performance degraded in the presence of noise. In this research work, we incorporated local spatial information in the FkPC clustering method to handle the noise present in the image. This spatial regularization term included in the proposed FkPC_S method refines the membership value of noisy pixel with the help of immediate neighbour pixels information. To show the effectiveness of the proposed FkPC_S method, extensive experiments are performed on one synthetic image and two publicly available human brain MRI datasets. The performance of the proposed method is compared with 10 related methods in terms of average segmentation accuracy and dice score. The experiments result shows that the proposed FkPC_S method is superior in comparison with 10 related methods in the presence of noise. Statistically significance difference and superior performance of the proposed method in comparison with other methods are also found using Friedman test.

Keywords Fuzzy k-plane clustering · Spatial constraint · Image segmentation · Magnetic resonance imaging

1 Introduction

Clustering is a well-established unsupervised learning technique in the field of machine learning. In the last few decades, various clustering techniques have been developed to represent intrinsic structures of a given dataset in terms of disjoint partitions or clusters [27]. Based on the representation of clusters, clustering can be classified into two categories, namely (i) hard clustering and (ii) soft clustering [7]. In the hard clustering approach, each data point of a given dataset is assigned to only one cluster, and mathematically, each cluster is represented in terms of crisp sets. In many real-world scenarios, hard clustering could not perform well as a given data-point may belong to more than one cluster. The soft clustering approach has been suggested to address this issue. The fuzzy clustering is a soft clustering approach where clusters are represented in terms of fuzzy sets, where membership function is used to represent degree of belongingness of a given data sample to a given partition. The fuzzy clustering retains more information than the hard clustering method as it enhances the cluster representation for better interpretation of the clustering outcome [33].

There are many ways to represent these cluster prototypes to assign an unknown data point to cluster directly. Based on the cluster prototype's representation, two types of clustering approaches are discussed in the literature. The first one is a point-based cluster prototype [1, 6, 9, 10, 17, 19–22, 31, 32, 36]. The performance of the

✉ Dhirendra Kumar
dhirendrakumar@dtu.ac.in

Puneet Kumar
puneet48_scs@jnu.ac.in

Ramesh Kumar Agrawal
rkajnu@gmail.com

[1] School of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India

[2] Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India

Published online: 12 November 2021

point-based cluster prototype is good when data samples are distributed around cluster centroid. It does not produce a satisfactory result where the data samples are not distributed around cluster centroid [5]. The second approach, a plane is determined to represent cluster prototype [8, 24, 26, 34], which is quite useful to represents both spherical and non-spherical cluster structures present in real-world data.

In the last few years, a massive amount of imaging data are generated by the healthcare industry [12]. The timely analysis of these medical imaging data is a challenging and crucial task for early disease diagnosis [2]. Diagnosis of brain diseases such as schizophrenia, Alzheimer, Parkinson, hemorrhage, vascular anomalies, and tumor requires special consideration [4]. Thus, accurate segmentation of brain tissue is a significant task in early disease detection. Computerized tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), single photon emission computed tomography (SPECT), diffusion tensor imaging (DTI), diffuse optical tomography (DOT), etc., are widely used imaging techniques to visualize sophisticated brain structure. According to the literature [1, 3, 9] MRI provides essential information about the human brain soft tissues. Hence, it is popular imaging modality for analysis of brain tissues and biomedical research. However, manual segmentation and analysis of MRI data is subjective and time consuming. To overcome these problems, computer-based analysis is becoming an invaluable tool in medicine and medical image analysis and drawn extensive attention these days [12, 28].

Segmentation using clustering is found to be a promising area in medical image analysis [14], which aims to partition an image into several homogeneous regions, which are disjoint and non-overlapping regions. It is very important to quantify different tissues of a given brain to diagnose a disease, which degenerate these tissues. The human brain mainly consists of three types of tissue, namely grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF), which are degenerated over a period of time causing the severe diseases. Another important observation regarding brain tissues is that the CSF tissues account a small proportion compared to other brain tissue, and the structure of brain tissues is non-spherical [3]. Thus, it becomes an obstacle for all traditional point-based prototype clustering methods to segment MRI brain images and provides poor segmentation.

In the literature, many methods are discussed [8, 24, 34] to represent non-spherical cluster structure, by representing the plane as a cluster prototype. The k-plane clustering (kPC) is the first plane-based clustering method proposed by Bradley et al. [8]. The kPC clustering method generates a cluster plane by the eigenvector corresponding to the smallest eigenvalue. However, the kPC clustering method considers only within-cluster information and the plane obtained is not well approximated. To overcome the kPC clustering method's limitation, Liu et al. [24] proposed an extended version of the kPC clustering method termed as k-proximal plane clustering (kPPC) method. The objective function of kPPC clustering method tries to approximate the plane prototype by minimizing the within clusters variance along with maximizing the between cluster variance. The kPPC clustering method utilizes eigenvalue problem similar to the kPC clustering method to obtain the cluster plane. However, both the kPC and the kPPC methods produce the plane prototype which are unbounded and, in some cases due to this both the methods give poor clustering results [34]. To handle this issue, Yang et al. [34] proposed local k-proximal plane clustering (LkPPC) method by incorporating the objective function of k-means clustering in the optimization problem of the kPPC clustering method. The term corresponding to the k-means objective function in the optimization problem of LkPPC method forces the plane prototype to be localized and bounded.

In the plane-based clustering approach, each data point of a given dataset is assigned to only one cluster. It is hard to obtain proper clusters in the case when the two sets of data samples are having overlapping distribution. Also, in many real-world applications, a given data sample may belong to more than one cluster. In order to handle such situation, Zadeh [35] suggested the fuzzy set theory, where a given data point may belongs to different sets using a membership value. Motivated from the drawback of the plane-based clustering, Zhu et al. suggested the fuzzy k-plane clustering (FkPC) [26] by utilizing the fuzzy set theory to represent the partition matrix in terms of k fuzzy sets in contrast to plane-based clustering. The research work [23] presented a fuzzy mix-prototype clustering (FMP) method to represent cluster prototype where the data distribution contains both spherical and non-spherical geometric structures. The objective function of the FMP clustering method is formulated by a combination of FCM and FkPC objective terms with trade-off parameter. The trade-off parameter controls the cluster prototype according to the distribution of data. Another research work is suggested by Wang et al. [30] termed as the fuzzy style k-plane clustering (FSkPC) method. The FSkPC method is an extended version of the FkPC clustering method by encoding a style information matrix as a regularization term.

Incorporation of fuzzy set theory in plane-based clustering is advantageous as the representation capabilities of the cluster are enhanced by fuzzy representation of the partition matrix. This has an important application in the case of MRI image analysis where a given voxel may belong to more than one tissue class specifically on the

boundary of two types of tissues. Also, MRI images are affected by artifacts such as low spatial resolution, inhomogeneity, noise, low contrast, and partial volume effects [15]. In the presence of these artifacts, segmentation of MRI image becomes more cumbersome and challenging. Although the fuzzy-based kPC methods address the problem of kPC clustering method, the performance of these methods for image segmentation is not good in the presence of the imaging artifacts.

Many research works based on point-based cluster prototype have been suggested in the literature to handle these artifacts. However, these methods are not suitable where the data samples are not distributed around cluster centroid [5]. In the literature, less attention is given to plane-based clustering method in comparison with point-based clustering method to handle noise. Motivated by this, we propose modification in FkPC to handle noise for image segmentation. For this, the objective function of the proposed fuzzy-based k-plane clustering utilizes a spatial regularization term to handle the noise in the segmentation process. The proposed approach is termed as FkPC_S throughout the manuscript.

The contribution of the proposed approach can be summarized as:

- To handle noise in image segmentation process, a local spatial information term is used.
- The fuzzy set theory is used to formulate the optimization problem of the proposed plane-based clustering method.
- The proposed fuzzy plane-based clustering method has been investigated on MRI image segmentation.

To check the effectiveness of the proposed FkPC_S method for image segmentation, experiments have been carried out using two publicly available brain MRI dataset and one synthetic image dataset. We have compared the performance of proposed FkPC_S method with several related methods such as kPC [8], kPPC [24], LkPPC [34], FCM [6], FkPC [26], EnFCM [31], FCM_S [1], FCM_S1 [10], FCM_S2 [10] and FGFCM [9] in terms of the average segmentation accuracy and dice score. The experimental results indicate superior performance of the proposed FkPC_S method is better in comparison with the state-of-the-art methods for MRI image segmentation task.

The rest part of the article is organized as follows. In Sect. 2, we briefly discuss the kPC, the FkPC and the FCM_S method. In Sect. 3, the formulation of the optimization problem of the proposed FkPC_S method is presented along with the solution. In Sect. 4, the details of dataset and experimental results are included. Finally, Sect. 5 includes the conclusion.

2 Background and related work

Consider $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_N}\} \in \mathbf{R^{d*N}}$ to represent an image with N numbers of d-dimensional pixels. Our task is to segment the image using the clustering of the pixel intensities in k regions such that pixels intensity are close to the respective cluster plane. The parameters of j^{th} cluster plane are represented by $\{\mathbf{w_j}, b_j\}$, where $\mathbf{w_j} \in R^d$ and $b_j \in R$, $\forall j \in \{1, 2, \ldots k\}$. The element of the fuzzy matrix over the set $\mathbf{X}$ is denoted by μ_{ij} representing the membership value of i^{th} element of $\mathbf{X}$ to the j^{th} cluster.

2.1 kPC

Bradley et al. [8] proposed first plane-based clustering algorithm which is referred to as k-plane clustering (kPC). The objective function of the kPC method is formulated which minimizes the overall sum of the squares of distances of i^{th} datapoints from j^{th} cluster plane prototype. The optimization problem of the kPC method is expressed as :

$$\begin{aligned} &\min_{\mathbf{w_j}, b_j} \sum_{i=1}^{N} \sum_{j=1}^{k} \|x_i^T \mathbf{w_j} + b_j\|^2 \\ &\text{s.t. } \|\mathbf{w_j}\|^2 = 1, \ 1 \leq j \leq k \end{aligned} \tag{1}$$

where $\mathbf{w_j} \in R^d$ represents normal vector to the plane and $b_j \in R$ represents bias. A plane discriminant function for each cluster is given as:

$$x^T \mathbf{w_j} + b_j = 0, \ \ j = 1, ..., k \tag{2}$$

The above optimization problem (1) is reduced to eigenvalue problems using the Lagrange's method of undetermined multiplier. After solving the Lagrangian function corresponding to (1), we obtain plane parameters by a $d \times d$ characteristic matrix denoted by $\mathbf{D_j}$. The characteristic matrix is represented by:

$$\mathbf{D_j} = \mathbf{A_j}^T \left(\mathbf{I} - \frac{\mathbf{e}\mathbf{e}^T}{N_j} \right) \mathbf{A_j} \tag{3}$$

where d is the dimension of data, N_j number of datapoints belonging to j^{th} cluster, $\mathbf{A_j}$ is a set of N_j number of datapoints belonging to j^{th} cluster, and $\mathbf{e}$ is the column vector of ones of size N_j. The plane parameter $\mathbf{w_j}$ is an eigenvector corresponding to the smallest eigenvalue of a $d \times d$ characteristic matrix $\mathbf{D_j}$. Using the values of $\mathbf{w_j}$, we can compute b_j as follows:

$$b_j = \frac{-\mathbf{e}^T \mathbf{A_j} \mathbf{w_j}}{N_j} \tag{4}$$

Further, the cluster assignment is done by the following equation.

$$\mathbf{Y} = argmin_j\{|\mathbf{X}\mathbf{w_j} + b_j|, j = 1, ..., k\} \tag{5}$$

where $|\cdot|$ denotes the absolute value. The kPC method alternatively computes the cluster plane prototype for each cluster and assigns data points to the nearest cluster plane prototype until no changes detected in cluster assignment. The outline of the kPC method is summarized in Algorithm 1.

Algorithm 1 The kPC algorithm

Initialisation:
The number of clusters k, number of iteration T
Set ϵ small value threshold as a stopping criteria.
Set iteration count t=0.
1.Randomly initialise cluster label Y_t
2.**Repeat**
3.for each $j = 1, ..., k$
3.1 Update the plane parameters w_j using (3)
3.2 Update the plane parameters b_j using (4)
4.Update the cluster assignment using(5)
5. t $\leftarrow$ t+1;
6. Stop Until $||Y_t - Y_{t-1}|| < \epsilon$ or T no of maximum iteration is achieved Then, set $Y_t = Y_{t-1}$
Return: $\{w_j, b_j\}$, for j=1,...,k

2.2 FkPC

In the research work [26], the authors exploits the fuzzy set theory in the conventional kPC method to fuzzy k-plane clustering (FkPC). In this research work, the partition matrix is represented using k fuzzy sets similar to fuzzy k-means clustering in contrast to kPC method where the partition matrix is crisp in nature. In this way, the FkPC method increases the representational capability of the clusters. The optimization problem of the FkPC method is given as:

$$\min_{\{u_{ij}, \mathbf{w_j}, b_j\}} \sum_{i=1}^{N}\sum_{j=1}^{k} \mu_{ij}^m (x_i^T \mathbf{w_j} + b_j)^2$$
$$\text{s.t. } \|\mathbf{w_j}\|^2 = 1, \textstyle\sum_{j=1}^{k} \mu_{ij} = 1,\ 1 \le i \le N,\ 1 \le j \le k \tag{6}$$

where x_i denotes a datapoints in $\mathbf{X}$, $\{\mathbf{w_j}, b_j\}$ are the parameters of the cluster prototype, and $\mathbf{U} = [\mu_{ij}]_{N\times k}$ represents the fuzzy partition matrix corresponding to the membership value for i^{th} datapoint to j^{th} cluster.

The above optimization problem (6) is reduced to eigenvalue problems using the Lagrange's method of undetermined multiplier. After solving the Lagrangian function corresponding to (6), for each j^{th} cluster, we obtain plane parameters by a $d \times d$ characteristic matrix denoted by $\mathbf{D_j}$. The characteristic matrix is represented in the following equation:

$$\mathbf{D_j}\mathbf{w_j} = \xi_\mathbf{j}\mathbf{w_j} \tag{7}$$

where $\mathbf{w_j}$, $\xi_\mathbf{j}$ is eigenvector and eigenvalue of jth cluster.

The expression of $\mathbf{D_j}$ characteristic matrix for j^{th} cluster can be expressed as:

$$\mathbf{D_j} = \frac{\left(\sum_{i=1}^{N} \mu_{ij}^m x_i\right)\left(\sum_{i=1}^{N} \mu_{ij}^m x_i^T\right)}{\sum_{i=1}^{N} \mu_{ij}^m} - \sum_{i=1}^{N} \mu_{ij}^m x_i x_i^T \tag{8}$$

The plane parameter $\mathbf{w_j}$ is an eigenvector corresponding to the smallest eigenvalue. Using the values of $\mathbf{w_j}$, we can compute b_j as follows:

$$b_j = \frac{-\sum_{i=1}^{N} \mu_{ij}^m x_i^T \mathbf{w_j}}{\sum_{i=1}^{N} \mu_{ij}^m} \tag{9}$$

Using the values of $\mathbf{w_j}$ & b_j, we can compute the fuzzy partition membership matrix μ_{ij}computed as follows:

$$\mu_{ij} = \left[\frac{(x_i^T \mathbf{w_j} + b_j)^2}{\sum_{p=1}^{k} (x_i^T \mathbf{w_p} + b_p)^2}\right]^{\frac{-1}{m-1}} \tag{10}$$

The FkPC method alternatively computes the cluster plane prototype and fuzzy partition matrix until terminating condition is satisfied. The outline of FkPC algorithm is summarized in Algorithm 2

Algorithm 2 FkPC

Initialisation:
The given dataset X, the fuzziness index m,
the number of cluster k, the number of iteration T.
Set ϵ small value threshold as a stopping criteria.
Set iteration count t=0.

1.Randomly initialise fuzzy membership matrix U_t
2.**Repeat**
3for each $j = 1, ..., k$
3.1 Update the plane parameters w_j using (7)
3.2 Update the plane parameters b_j using (9)
5.Update the fuzzy membership matrix using (10)
6.t $\leftarrow$ t+1;
7.Stop **Until** $||U_t - U_{t-1}|| < \epsilon$ or the T no of maximum iteration is achieved and set $U_t = U_{t-1}$
Return : $U = \{\mu_{ij}\}_{k\times N}$, $\{w_j, b_j\}$, for j=1,...,k

2.3 FCM_S

In the research work [1], Ahmed et.al. proposed a modified FCM method with a spatial term referred to as FCM_S. The FCM_S clustering method obtain better segmentation performance in the presence of noise. Basically, the FCM_S clustering method smooths a pixel intensity value within a specified window of neighbourhood. The optimization problem of the FCM_S method is given as:

$$\min_{\{U,V\}} \sum_{i=1}^{N}\sum_{j=1}^{k} \mu_{ij}^{m}\left\|x_i - v_j\right\|^2 + \frac{\alpha}{N_R}\sum_{i=1}^{N}\sum_{j=1}^{k}\mu_{ij}^{m}\sum_{r\in N_i}\left\|x_i - v_j\right\|^2 \quad \text{s.t.} \sum_{p=1}^{k}\mu_{ip} = 1,\ 1 \le i \le N \tag{11}$$

where $\mathbf{U} = \{\mu_{ij}\}_{N\times k}$ is fuzzy partition matrix of i^{th} datapoint to the j^{th} cluster, and $\mathbf{V} = \{v_j\}_{k\times 1}$ represents cluster center of j^{th} cluster. The expressions for fuzzy partition matrix and cluster center can be obtained after solving optimization problem (11) using Lagrange method of undetermined multiplier and are given as:

$$v_j = \frac{\sum_{i=1}^{N}\mu_{ij}^{m}(x_i + \frac{\alpha}{N_R}\sum_{r\in N_i}x_r)}{(1+\alpha)\sum_{i=1}^{N}\mu_{ij}^{m}},\ 1 \le j \le k \tag{12}$$

$$\mu_{ij} = \left(\frac{(\|x_i - v_j\|^2 + \frac{\alpha}{N_R}\sum_{r\in N_i}\|x_r - v_j\|^2)}{\sum_{p=1}^{k}(\|x_i - v_p\|^2 + \frac{\alpha}{N_R}\sum_{r\in N_i}\|x_r - v_p\|^2)}\right)^{\frac{-1}{(m-1)}} \tag{13}$$

The outline of FCM_S algorithm is given in Algorithm 3

Algorithm 3 FCM_S

Initialisation:
The given dataset X, the fuzziness index m, the number of cluster k, α is regularisation parameter and the T number of iteration.
Set ϵ small value threshold as a stopping criteria.
Set iteration count t=0.
1.Randomly initialise fuzzy membership matrix U_t
3.**Repeat**
4.for each $j = 1, ..., k$
4.1 Update the cluster centres V_t using (12)
5.Calculate the membership matrix U_t using (13)
6.t $\leftarrow$ t+1;
7.Stop **Until** $||U_t - U_{t-1}|| < \epsilon$ or the T no of maximum iteration is achieved and set $U_t = U_{t-1}$
Return : $U = \{\mu_{ij}\}_{k\times N}$, $\{w_j, b_j\}$, for j=1,...,k

3 The proposed fuzzy k-plane clustering method with local spatial information

In this section, we present the proposed modified fuzzy k-plane clustering by introducing local spatial information term in the fuzzy k-plane clustering method to handle noise for MRI image segmentation.

3.1 Motivation of proposed approach

The performance of the FkPC method for image segmentation problem is not good in the presence of noise, which leads to inaccurate segmentation results. It is mainly due to the non-robust Euclidean distance measure such as L_2-norm and ignorance of the local spatial information in segmentation process. In the literature, less attention is given to plane-based clustering method in comparison with point-based clustering method to handle noise in non-spherical shape clusters. This motivated us to develop clustering method for image segmentation by incorporating local spatial information term in plane-based clustering method to handle noise.

3.2 Description of proposed method

The optimization problem of the proposed fuzzy k-plane clustering with local spatial information (FkPC_S) method is comprised of two terms. The first term minimizes within-class scatteredness by assigning a high membership value to pixel intensity close to the cluster plane and assigning a low membership value to pixels far from the cluster plane. The second term includes a spatial regularization term which refines the membership value corresponding to noisy pixel using the immediate neighbourhood pixels information for correct cluster assignment. The optimization problem of the proposed method is given as:

$$\min_{\{\mu_{ij},\mathbf{w_j},b_j\}} \sum_{i=1}^{N}\sum_{j=1}^{k}\mu_{ij}^{m}(x_i^T\mathbf{w_j} + b_j)^2 + \frac{\alpha}{N_R}\sum_{i=1}^{N}\sum_{j=1}^{k}\mu_{ij}^{m}\sum_{r\in N_i}(x_r^T\mathbf{w_j} + b_j)^2 \quad \text{s.t.} \sum_{j=1}^{k}\mu_{ij} = 1, \left\|\mathbf{w_j}\right\|_2^2 = 1,\ 1 \le i \le N,\ 1 \le j \le k \tag{14}$$

where $x_i \in \mathbf{X}$ is the i^{th} datapoint and $\mathbf{X}$ is the set of d-dimensional pixels of size N to be clustered in k groups. $m \in (1, \text{inf})$ is fuzziness constant, μ_{ij} is an element of fuzzy partition matrix over the set $\mathbf{X}$ representing the membership value of i^{th} element of $\mathbf{X}$ to the j^{th} cluster, respectively. The α controls the regularization parameter of local spatial information. The neighbourhood window size (NWin)in an image is $w \times w$. $\mathrm{N_i}$ stands for the set of neighbours that exist in a window around $\mathrm{x_i}$ and $\mathrm{N_R}$ is the cardinality of $\mathrm{N_i}$.

3.3 Derivation for the cluster prototype and membership value

The optimization problem of the proposed method can be solved for cluster prototype and membership value. The expressions for fuzzy partition matrix $\mathbf{U} = \{\mu_{ij}\}_{c\times N}$ and plane parameter $\{\mathbf{w_j}, b_j\}$ is obtained by solving Lagrangian function $\mathbf{L}\{\mathbf{w_j}, b_j, \mu_{ij}, \lambda_i, \xi_\mathbf{j}\}$ of proposed method which is represented as:

$$\mathbf{L}\{\mathbf{w_j}, b_j, \mu_{ij}, \lambda_i, \xi_\mathbf{j}\} = \sum_{i=1}^{N}\sum_{j=1}^{k}\mu_{ij}^m(x_i^T\mathbf{w_j} + b_j)^2 + \frac{\alpha}{N_R}\sum_{i=1}^{N}\sum_{j=1}^{k}\mu_{ij}^m\sum_{r\in N_i}(x_r^T\mathbf{w_j} + b_j)^2 + \sum_{i=1}^{N}\lambda_i\left(1 - \sum_{j=1}^{k}\mu_{ij}\right) + \sum_{j=1}^{k}\xi_\mathbf{j}\left(\|\mathbf{w_j}\|_2^2 - 1\right) \tag{15}$$

where λ_i and $\xi_\mathbf{j}$ are the Lagrange multipliers. For finding the optimal solution, equating the partial derivatives of the Lagrangian function $\mathbf{L}\{\mathbf{w_j}, b_j, \mu_{ij}, \lambda_i, \xi_j\}$ w.r.t $\mathbf{w_j}, b_j, \mu_{ij}, \lambda_i, \xi_\mathbf{j}$ to zero leads to the reduction of the optimization problem to eigenvalue problem.

Taking partial differentiation of $\mathbf{L}\{\mathbf{w_j}, b_j, \mu_{ij}, \lambda_i, \xi_\mathbf{j}\}$ w.r.t $\mathbf{w_j}$ gives:

$$\frac{\partial L}{\partial \mathbf{w_j}} = \sum_{i=1}^{N}\mu_{ij}^m(x_i^T\mathbf{w_j} + b_j)x_i^T + \frac{\alpha}{N_R}\sum_{i=1}^{N}\mu_{ij}^m\sum_{r\in N_i}(x_r^T\mathbf{w_j} + b_j)x_r^T + \xi_\mathbf{j}\mathbf{w_j} \tag{16}$$

Taking partial differentiation of $\mathbf{L}\{\mathbf{w_j}, b_j, \mu_{ij}, \lambda_i, \xi_\mathbf{j}\}$ w.r.t b_j gives:

$$\frac{\partial L}{\partial b_j} = \sum_{i=1}^{N}\mu_{ij}^m(x_i^T\mathbf{w_j} + b_j) + \frac{\alpha}{N_R}\sum_{i=1}^{N}\mu_{ij}^m\sum_{r\in N_i}(x_r^T\mathbf{w_j} + b_j) \tag{17}$$

Taking partial differentiation of $\mathbf{L}\{\mathbf{w_j}, b_j, \mu_{ij}, \lambda_i, \xi_\mathbf{j}\}$ w.r.t μ_{ij} gives:

$$\frac{\partial L}{\partial \mu_{ij}} = m\mu_{ij}^{m-1}(x_i^T\mathbf{w_j} + b_j)^2 + \frac{\alpha}{N_R}m\mu_{ij}^{m-1}\sum_{r\in N_i}(x_r^T\mathbf{w_j} + b_j)^2 - \lambda_i \tag{18}$$

Taking partial differentiation of $\mathbf{L}\{\mathbf{w_j}, b_j, \mu_{ij}, \lambda_i, \xi_\mathbf{j}\}$ w.r.t λ_i gives:

$$\frac{\partial L}{\partial \lambda_i} = 1 - \sum_{i=1}^{k}\mu_{ij} \tag{19}$$

Taking partial differentiation of $\mathbf{L}\{\mathbf{w_j}, b_j, \mu_{ij}, \lambda_i, \xi_\mathbf{j}\}$ w.r.t $\xi_\mathbf{j}$ gives:

$$\frac{\partial L}{\partial \xi_j} = ||w_j||_2^2 - 1 \tag{20}$$

3.3.1 Derivation of plane prototype

To obtain the cluster plane parameter, we have to equate partial derivative of Lagrangian function w.r.t. b_j i.e. (17) to zero.

$$\sum_{i=1}^{N}\mu_{ij}^m(x_i^T\mathbf{w_j} + b_j) + \frac{\alpha}{N_R}\sum_{i=1}^{N}\mu_{ij}^m\sum_{r\in N_i}(x_r^T\mathbf{w_j} + b_j) = 0 \tag{21}$$

$$\sum_{i=1}^{N}\mu_{ij}^m b_j + \frac{\alpha}{N_R}\sum_{i=1}^{N}\mu_{ij}^m\sum_{r\in N_i}b_j + \sum_{i=1}^{N}\mu_{ij}^m x_i^T\mathbf{w_j} + \frac{\alpha}{N_R}\sum_{i=1}^{N}\mu_{ij}^m\sum_{r\in N_i}x_r^T\mathbf{w_j} = 0 \tag{22}$$

$$b_j\left(\sum_{i=1}^{N}\mu_{ij}^m + \frac{\alpha}{N_R}\sum_{i=1}^{N}\mu_{ij}^m * N_R\right) = -\left(\sum_{i=1}^{N}\mu_{ij}^m x_i^T + \frac{\alpha}{N_R}\sum_{i=1}^{N}\mu_{ij}^m\sum_{r\in N_i}x_r^T\right)\mathbf{w_j} \tag{23}$$

The expression of b_j can be represented as:

$$b_j = \frac{-\left(\sum_{i=1}^{N}\mu_{ij}^m x_i^T + \frac{\alpha}{N_R}\sum_{i=1}^{N}\mu_{ij}^m\sum_{r\in N_i}x_r^T\right)\mathbf{w_j}}{(1+\alpha)\sum_{i=1}^{N}\mu_{ij}^m} \tag{24}$$

In short, b_j can rewritten in the form as:

$$b_j = -\mathbf{T_j}\mathbf{w_j} \tag{25}$$

where

$$\mathbf{T_j} = \frac{\left(\sum_{i=1}^{N}\mu_{ij}^m x_i^T + \frac{\alpha}{N_R}\sum_{i=1}^{N}\mu_{ij}^m\sum_{r\in N_i}x_r^T\right)}{(1+\alpha)\sum_{i=1}^{N}\mu_{ij}^m} \tag{26}$$

On substituting b_j from equation (24) to equation(16) gives the solution for $\mathbf{w_j}$.

$$\sum_{i=1}^{N}\mu_{ij}^m(x_i^T\mathbf{w_j} - \mathbf{T_j}\mathbf{w_j})x_i^T + \frac{\alpha}{N_R}\sum_{i=1}^{N}\mu_{ij}^m\sum_{r\in N_i}(x_r^T w_j - \mathbf{T_j}\mathbf{w_j})x_r^T + \xi_\mathbf{j}\mathbf{w_j} = 0 \tag{27}$$

$$\sum_{i=1}^{N}\mu_{ij}^m x_i(x_i^T w_j - \mathbf{T_j}\mathbf{w_j}) + \frac{\alpha}{N_R}\sum_{r\in N_i}\sum_{i=1}^{N}\mu_{ij}^m x_r(x_r^T\mathbf{w_j} - \mathbf{T_j}\mathbf{w_j}) + \xi_\mathbf{j}\mathbf{w_j} = 0 \tag{28}$$

The above equation can be rewritten in the form

$$[(\sum_{i=1}^{N}\mu_{ij}^{m}x_i+\frac{\alpha}{N_R}\sum_{r\in N_i}\sum_{i=1}^{N}\mu_{ij}^{m}x_r)\mathbf{T_j} -(\sum_{i=1}^{N}\mu_{ij}^{m}x_ix_i^T+\frac{\alpha}{N_R}\sum_{r\in N_i}\sum_{i=1}^{N}\mu_{ij}^{m}x_rx_r^T)]\mathbf{w_j}=\xi_{\mathbf{j}}\mathbf{w_j} \tag{29}$$

Further, the above equation can be summarized in the form of eigenvalue problem and can be rewritten as:

$$\mathbf{D_j}\mathbf{w_j}=\xi_{\mathbf{j}}\mathbf{w_j} \tag{30}$$

where

$$\mathbf{D_j}=\left(\sum_{i=1}^{N}\mu_{ij}^{m}x_i+\frac{\alpha}{N_R}\sum_{r\in N_i}\sum_{i=1}^{N}\mu_{ij}^{m}x_r\right)\mathbf{T_j}- \left(\sum_{i=1}^{N}\mu_{ij}^{m}x_ix_i^T+\frac{\alpha}{N_R}\sum_{r\in N_i}\sum_{i=1}^{N}\mu_{ij}^{m}x_rx_r^T\right) \tag{31}$$

where $\mathbf{D_j}$, $\xi_{\mathbf{j}}$ and $\mathbf{w_j}$ are the characteristic matrices, the eigenvalue of $\mathbf{D_j}$ and eigenvector of $\mathbf{D_j}$, respectively. Once $\mathbf{w_j}$ is determined, then b_j can be determined by equation (24).

3.3.2 Derivation of membership function

To obtain iterative formula for membership function, we have to equate partial derivation equation (18) to zero, i.e.$\frac{\partial L}{\partial \mu_{ij}}=0$

$$m\mu_{ij}^{m-1}(x_i^T\mathbf{w_j}+b_j)^2+\frac{\alpha}{N_R}m\mu_{ij}^{m-1}\sum_{r\in N_i}(x_r^T\mathbf{w_j}+b_j)^2-\lambda_i=0 \tag{32}$$

$$\mu_{ij}=\left(\frac{\lambda_i}{m((x_i^T\mathbf{w_j}+b_j)^2+\frac{\alpha}{N_R}\sum_{r\in N_i}(x_r^T\mathbf{w_j}+b_j)^2)}\right)^{\frac{1}{m-1}} \tag{33}$$

from the equation(19) $\sum_{p=1}^{k}\mu_{ip}=1$, we can write the expression as

$$\sum_{p=1}^{k}\left(\frac{\lambda_i}{m((x_i^T\mathbf{w_p}+b_p)^2+\frac{\alpha}{N_R}\sum_{r\in N_i}(x_r^T\mathbf{w_p}+b_p)^2)}\right)^{\frac{1}{m-1}}=1 \tag{34}$$

or

$$\lambda_i=\frac{1}{\left(\frac{1}{\sum_{p=1}^{k}m((x_i^T\mathbf{w_p}+b_p)^2+\frac{\alpha}{N_R}\sum_{r\in N_i}(x_r^T\mathbf{w_p}+b_p)^2)}\right)} \tag{35}$$

Now, eliminating λ_i from equation (33) by substituting above equation (35), Hence, iterative formula for membership matrix function is expressed as

$$\mu_{ij}=\left[\frac{(x_i^T\mathbf{w_j}+b_j)^2+\frac{\alpha}{N_R}\sum_{r\in N_i}(x_r^T\mathbf{w_j}+b_j)^2}{\sum_{p=1}^{k}((x_i^T\mathbf{w_p}+b_p)^2+\frac{\alpha}{N_R}\sum_{r\in N_i}(x_r^T\mathbf{w_p}+b_p)^2)}\right]^{\frac{-1}{m-1}} \tag{36}$$

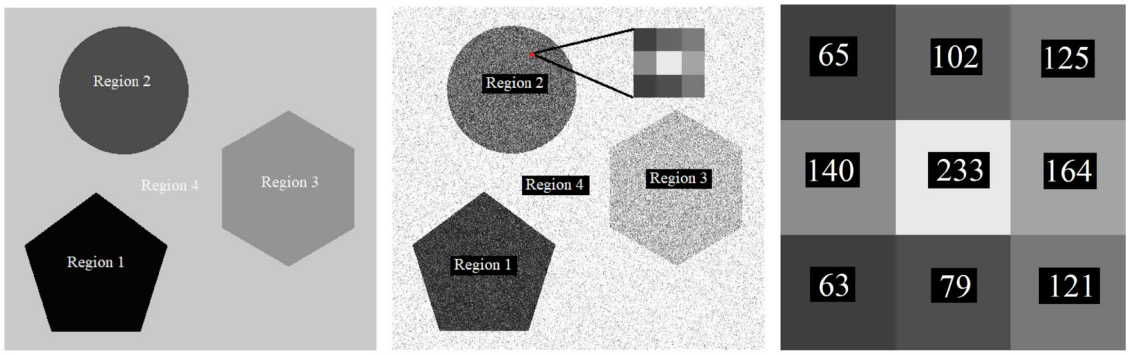
The outline of proposed method is summarized in algorithm 4.

Algorithm 4 The proposed FkPC_S algorithm

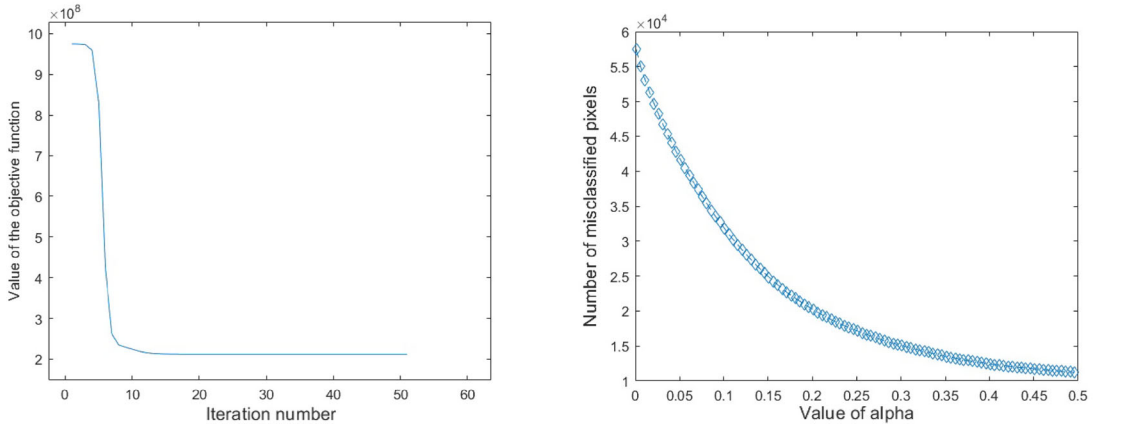
Initialisation:
The given dataset X, the number of cluster k, the number of iteration T, and m is the fuzziness index constant
Set ϵ small value threshold as a stopping criteria
Set iteration count t=0
1.Randomly initialise Fuzzy membership matrix U_t
2.**Repeat**
3.for each $j=1,...,k$
3.1 Update the plane parameters $\mathbf{w_j}$ using (30)
3.2 Update the plane parameters b_j using (24)
4.Update the Fuzzy membership matrix using (36)
5.t $\leftarrow$ t+1;
6. Stop **Until** $||U_t-U_{t-1}||<\epsilon$ or the T no of maximum iteration is achieved and set $U_t=U_{t-1}$
7. **Return** : $U=\{\mu_{ij}\}_{k\times N}$ is fuzzy partition matrix, $\{\mathbf{w_j},b_j\}$, for j=1,...,k .

3.4 Demonstration of the proposed approach in the presence of noise

We demonstrate the working of the proposed method on a synthetic image data (shown in Fig. 1a) corrupted with a Rician noise with standard deviation 20 as noise level (NL) in Fig. 1b. We show the working of proposed method on 3×3 image patch, which is extracted from region 2 of the noisy corrupted synthetic image (shown in Fig 1c). The proposed method in each iteration refines the membership of center pixel of noisy image patch with the help of neighbourhood information (shown in Table 1). It can be noted from Fig.1d that the value of the objective function of the proposed method is minimized and converges in finite number of iterations. We also show the variation of number of miss-classified pixels with regularization parameter on noise corrupted image (shown in Fig. 1e. It can be noted that the number of misclassified pixels decreases substantially initially followed by a marginal decrease with an increase in the value of regularization parameter. We also show the membership values of each region for different iterations in Table 1 corresponding to the 3×3 image patch. It can be observed from Table 1 that in the initial iteration the membership values are uniformly distributed. The membership value of central pixel of noisy image patch is given in boldface in each region. In the 5th iteration, the membership values for noisy central pixel are gradually modified and cross its value above 0.5 for region 2 and below 0.5 for other regions. As the number of iterations increases further due to correction in membership



(a) Synthetic image **(b)** Synthetic image corrupted with Rician Noise (NL(σ) = 20) **(c)** Noisy 3 × 3 image patch



(d) Value of the objective function Vs iteration **(e)** Variation of number of miss-classified pixels with regularisation parameter (α)

Fig. 1 Demonstration of the proposed method on 3 × 3 image patch of synthetic image corrupted with Rician noise (NL(σ) = 20)

values of the noisy pixels to the given region the cluster plane prototype value is also eventually modified. The effect can be seen in Table 1 where the membership matrix converges after 40th iterations. After the defuzzification, these noisy data points are assigned to the desired cluster. This shows that the proposed method successfully reduces the effect of noise and correctly classifies the noisy data points. In the next section, we present experiments on MRI images to show the effectiveness and applicability of the proposed FkPC_S method on MRI datasets for segmentation.

4 Experimental setup and results

To check the effectiveness of the proposed FkPC_S method, we compare the performance of the proposed method with 10 related methods, which includes: kPC [8], KPPC [24], LkPPC [34], FCM [6], FkPC [26], EnFCM[31], FCM_S [1], FCM_S1 [10], FCM_S2 [10], and FGFCM [9]. To check the segmentation performance, we used average segmentation accuracy (ASA) and dice score (DS), which are mathematically defined as:

$$ASA = \sum_{i=1}^{k} \frac{|X_i \cap Y_i|}{\sum_{j=1}^{k} |X_j|}$$

$$DS = \frac{2|X_i \cap Y_i|}{|X_i| + |Y_i|}$$

where X_i denotes the set of pixels belonging to the i^{th} class of ground truth image, Y_i denotes the set of pixels belonging to the i^{th} cluster of segmented image, $|X_i|$ denotes the cardinality of pixels in X_i corresponding to i^{th} region of ground truth image, and similarly $|Y_i|$ denotes the cardinality of pixels in Y_i corresponding to i^{th} region of segmented image. The high value of DS and ASA for a given method signifies the better performance of the method. The best performance achieved by a method compared to other methods is shown in boldface in all the tables.

Table 1 The membership values obtained using the proposed method for four segmented regions corresponding to 3×3 window in a synthetic image with Rician noise ($\sigma = 20$)

Iteration/membership value	Region 1			Region 2			Region 3			Region 4		
1	0.2509	0.2509	0.2509	0.2519	0.2519	0.2519	0.2480	0.2480	0.2480	0.2492	0.2492	0.2492
	0.2509	**0.2509**	0.2513	0.2519	**0.2519**	0.2528	0.2480	**0.2480**	0.2470	0.2492	**0.2492**	0.2488
	0.2507	0.2509	0.2503	0.2515	0.2518	0.2505	0.2484	0.2481	0.2494	0.2494	0.2492	0.2498
5	0.2853	0.2832	0.2813	0.5504	0.5537	0.5549	0.0607	0.0603	0.0608	0.1036	0.1028	0.1029
	0.2861	**0.2916**	0.2086	0.5460	**0.5401**	0.7217	0.0623	**0.0619**	0.0234	0.1056	**0.1064**	0.0463
	0.2969	0.2878	0.2957	0.4850	0.5376	0.3176	0.0850	0.0652	0.1664	0.1331	0.1093	0.2202
10	0.1232	0.0839	0.0155	0.8639	0.9072	0.9828	0.0035	0.0024	0.0005	0.0094	0.0065	0.0012
	0.1015	**0.2243**	0.5618	0.8874	**0.7525**	0.4146	0.0030	**0.0063**	0.0059	0.0080	**0.0170**	0.0176
	0.0259	0.0698	0.4217	0.9698	0.9221	0.4082	0.0012	0.0022	0.0493	0.0031	0.0059	0.1208
15	0.0689	0.0442	0.0044	0.9218	0.9498	0.9950	0.0031	0.0020	0.0002	0.0062	0.0040	0.0004
	0.0567	**0.1381**	0.3289	0.9354	**0.8438**	0.6475	0.0026	**0.0060**	0.0076	0.0052	**0.0121**	0.0161
	0.0240	0.0394	0.4101	0.9714	0.9549	0.4459	0.0016	0.0019	0.0494	0.0031	0.0038	0.0946
20	0.0549	0.0345	0.0022	0.9364	0.9600	0.9975	0.0030	0.0019	0.0001	0.0057	0.0036	0.0002
	0.0453	**0.1133**	0.2587	0.9474	**0.8693**	0.7176	0.0025	**0.0059**	0.0078	0.0048	**0.0115**	0.0160
	0.0238	0.0318	0.3978	0.9709	0.9628	0.4590	0.0018	0.0018	0.0500	0.0034	0.0035	0.0932
25	0.0527	0.0329	0.0019	0.9388	0.9616	0.9978	0.0029	0.0019	0.0001	0.0056	0.0036	0.0002
	0.0435	**0.1093**	0.2472	0.9493	**0.8735**	0.7291	0.0025	**0.0058**	0.0078	0.0048	**0.0114**	0.0160
	0.0238	0.0306	0.3952	0.9708	0.9640	0.4615	0.0019	0.0018	0.0502	0.0035	0.0035	0.0931
30	0.0523	0.0327	0.0018	0.9391	0.9619	0.9979	0.0029	0.0018	0.0001	0.0056	0.0036	0.0002
	0.0432	**0.1086**	0.2454	0.9496	**0.8741**	0.7309	0.0025	**0.0058**	0.0078	0.0048	**0.0114**	0.0160
	0.0238	0.0305	0.3948	0.9708	0.9642	0.4619	0.0019	0.0018	0.0502	0.0035	0.0035	0.0931
35	0.0523	0.0327	0.0018	0.9392	0.9619	0.9979	0.0029	0.0018	0.0001	0.0056	0.0036	0.0002
	0.0431	**0.1086**	0.2451	0.9496	**0.8742**	0.7311	0.0025	**0.0058**	0.0078	0.0048	**0.0114**	0.0160
	0.0238	0.0304	0.3947	0.9708	0.9642	0.4619	0.0019	0.0018	0.0502	0.0035	0.0035	0.0931
40	0.0523	0.0327	0.0018	0.9392	0.9619	0.9979	0.0029	0.0018	0.0001	0.0056	0.0036	0.0002
	0.0431	**0.1085**	0.2451	0.9496	**0.8743**	0.7312	0.0025	**0.0058**	0.0078	0.0048	**0.0114**	0.0160
	0.0238	0.0304	0.3947	0.9708	0.9642	0.4619	0.0019	0.0018	0.0502	0.0035	0.0035	0.0931
45	0.0523	0.0327	0.0018	0.9392	0.9619	0.9979	0.0029	0.0018	0.0001	0.0056	0.0036	0.0002
	0.0431	**0.1085**	0.2451	0.9496	**0.8743**	0.7312	0.0025	**0.0058**	0.0078	0.0048	**0.0114**	0.0160
	0.0238	0.0304	0.3947	0.9708	0.9642	0.4620	0.0019	0.0018	0.0502	0.0035	0.0035	0.0931
50	0.0523	0.0327	0.0018	0.9392	0.9619	0.9979	0.0029	0.0018	0.0001	0.0056	0.0036	0.0002
	0.0431	**0.1085**	0.2451	0.9496	**0.8743**	0.7312	0.0025	**0.0058**	0.0078	0.0048	**0.0114**	0.0160
	0.0238	0.0304	0.3947	0.9708	0.9642	0.4620	0.0019	0.0018	0.0502	0.0035	0.0035	0.0931

4.1 Dataset description

To validate and compare the performance of the proposed FkPC_S method with state of-the-art methods, we have utilized three image datasets with given ground truth. The first image dataset is a synthetic gray image which is having four homogeneous regions with intensities values 4, 76, 148 and 201 for each, respectively (shown in Fig.1a). We corrupted the gray synthetic image with Rician noise and Gaussian white noise with different levels of noise. The second one is publicly available simulated brain MRI database dataset [11] with given ground truth. The BrainWeb dataset is a simulated T1-weighted MRI image which are corrupted with Rician noise (3%, 5%, 7% and 9%). In this dataset, each MRI is a 3-D volume of the size $181 \times 217 \times 181$ with $1 \times 1 \times 1$ mm^3 resolution. Another publicly available real brain MRI dataset is acquired from Internet Brain Segmentation Repository (IBSR) with the given ground truth. In the IBSR dataset, each MRI image is 3-D volume of size $256 \times 256 \times 56$ with $1 \times 1 \times 1$ mm^3 resolution. The IBSR is corrupted with Rician noise ($\sigma = 5$ and $\sigma = 10$) to show the effectiveness of the incorporated spatial term in the proposed method in comparison with the state-of-the-art methods. We used the brain extraction tool

[29] for scull striping for each MRI dataset. All clustering methods such as kPC, kPPC, LkPPC, FCM, FkPC, EnFCM, FCM_S, FCM_S1, FCM_S2, FGFCM and the proposed FkPC_S method were implemented using MATLAB 2021a [25] running on a computer with Intel i7 2.60 GHz with 16 GB of RAM. Table 2 shows the values of the parameter used in proposed FkPC_S method and existing methods.

4.2 Results and discussion on synthetic images

We performed the experiments on the synthetic images corrupted with Gaussian white noise ($\sigma = 5$, $\sigma = 10$ and $\sigma = 15$) and Rician noise ($\sigma = 10$, $\sigma = 15$ and $\sigma = 20$). Tables 3 and 5 show the ASA obtained for the image corrupted with Gaussian noise and Rician noise, respectively. Tables 4 and 6 show the DS obtained for the image corrupted with Gaussian noise and Rician noise, respectively. It can be noted from Tables 3, 5,4, 6 that the performance of the proposed FkPC_S method in terms of various quantitative measures is better in comparison with other methods for different levels and type of noise.

The qualitative results corresponding to the proposed FkPC_S method and other related are presented in Fig. 2 for synthetic images corrupted with Rician noise ($\sigma = 15$). It can be noted from Fig.2 that the performance of the proposed FkPC_S method is better than other related methods.

4.3 Results and discussion on simulated brain MRI images

To show the efficacy of the proposed FkPC_S method for MRI image segmentation in presence of noise, we performed the experiments on the simulated brain images corrupted with different levels of noise. Table 7 and Table 8 show the comparison in terms the ASA and DS for the Brainweb MRI images. The quantitative results are also presented in terms of ASA with the help of bar chart corresponding to different level of Rician noise in Fig. 3. The following observation can be drawn from Tables 7, 8 and Fig. 3:

- The proposed FkPC_S method performs better in comparison with other related method in terms of ASA for all levels of noise except 3% noise.
- The proposed FkPC_S method performs better in comparison with other related method in terms of DS for all levels of noise except 3% noise.
- The value of the DS with the proposed method for both GM and WM brain tissues is higher in comparison with other related methods and marginally less for CSF tissue in comparison with other related methods.
- The segmentation performance of all methods decreases with increase in the level of noise. However, the reduction in the performance of the proposed FkPC_S method is less compared to other related methods (shown in Fig. 3) with the increase in the noise level.

To show the effectiveness of segmentation performance of our proposed FkPC_S method in comparison with other related methods in terms of the Dice score for GM, WM and CSF is summarized in Table 8. The value of the DS for both GM & WM brain tissues is higher in comparison with other related methods and in the case of CSF tissue the value is comparative in comparison with other related methods. This shows that the proposed FkPC_S method can accurately identify the brain regions in the Brainweb MRI image dataset. We present qualitative results of brain tissues, i.e. GM, WM, and CSF highlighted with green, blue, and red color, respectively, on a brain web simulated

Table 2 List of parameter values corresponding to different methods

Methods	Synthetic image dataset	BrainWeb dataset	IBSR clinical dataset
kPC	k = 4	k = 4	k = 4
kPPC	k = 4	k = 4	k = 4
LkPPC	k = 4	k = 4	k = 4
FCM	k = 4, m = 2	k = 4, m = 2	k = 4, m = 2
FkPC	k = 4, m = 2	k = 4, m = 2	k = 4, m = 2
EnFCM	k = 4, m = 2, NWin= 3×3	k = 4, m = 2, NWin = 3×3	k=4, m=2, NWin= 3×3
FCM_S	k = 4, m = 2, NWin = 3×3	k = 4, m = 2, NWin = 3×3	k = 4, m = 2, NWin= 3×3
FCM_S1	k = 4, m = 2, NWin= 3×3	k = 4, m = 2, NWin = 3×3	k = 4, m = 2, NWin = 3×3
FCM_S2	k = 4, m = 2, NWin = 3×3	k = 4, m = 2, NWin = 3×3	k = 4, m = 2, NWin = 3×3
FGFCM	k = 4, m = 2, NWin = 3×3	k = 4, m = 2, NWin = 3×3	k = 4, m = 2, NWin = 3×3
FkPC_S	k = 4, m = 2, NWin = 3×3	k = 4, m = 2, NWin = 3×3	k= 4, m = 2, NWin = 3×3

Table 3 Quantitative segmentation performance of proposed method and related methods on synthetic image dataset corrupted with Gaussian noise in terms of ASA

NL(σ)	kPC	kPPC	LkPPC	FCM	FkPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
5	0.6906	0.6956	0.7083	0.6889	0.6889	0.7469	0.7466	0.7445	0.7453	0.7420	**0.9328**
10	0.6163	0.6168	0.6261	0.6169	0.6169	0.6529	0.6403	0.6410	0.6503	0.6491	**0.7360**
15	0.5815	0.5841	0.5889	0.5799	0.5799	0.6010	0.5873	0.5885	0.6062	0.6036	**0.6351**
Avg. Performance	0.6295	0.6322	0.6411	0.6286	0.6286	0.6669	0.6581	0.6580	0.6673	0.6649	**0.7680**

Table 4 Quantitative segmentation performance of proposed method and related methods on synthetic image dataset corrupted with Gaussian noise in terms of Dice score

NL(σ)	Cluster region	kPC	kPPC	LkPPC	FCM	FkPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
5	Region 1	0.8180	0.8212	0.8305	0.8148	0.8148	0.8436	0.8392	0.8391	0.8414	0.8413	**0.9550**
	Region 2	0.3176	0.3252	0.3480	0.3203	0.3203	0.4105	0.4152	0.4099	0.4037	0.4018	**0.7917**
	Region 3	0.4830	0.4989	0.5414	0.4973	0.4973	0.6568	0.6774	0.6674	0.6629	0.6453	**0.9508**
	Region 4	0.7972	0.7992	0.7950	0.7992	0.7992	0.8559	0.8671	0.8635	0.8648	0.8529	**0.9738**
10	Region 1	0.7717	0.7717	0.7798	0.7717	0.7717	0.7874	0.7707	0.7733	0.7826	0.7853	**0.7975**
	Region 2	0.2414	0.2427	0.2545	0.2427	0.2427	0.2977	0.2932	0.2918	0.2872	0.2915	**0.4477**
	Region 3	0.3227	0.3281	0.3498	0.3303	0.3303	0.4345	0.4365	0.4284	0.4431	0.4269	**0.8144**
	Region 4	0.6963	0.6974	0.6984	0.6978	0.6978	0.7717	0.7865	0.7818	0.7850	0.7713	**0.9221**
15	Region 1	0.7488	0.7510	**0.7553**	0.7464	0.7464	0.7522	0.7312	0.7347	0.7530	**0.7553**	0.7197
	Region 2	0.2106	0.2127	0.2203	0.2133	0.2133	0.2557	0.2573	0.2556	0.2503	0.2533	**0.3256**
	Region 3	0.2582	0.2623	0.2767	0.2642	0.2642	0.3310	0.3373	0.3298	0.3498	0.3338	**0.6543**
	Region 4	0.6209	0.6221	0.6251	0.6228	0.6228	0.7112	0.7302	0.7252	0.7303	0.7151	**0.8701**
Avg. performance		0.5239	0.5277	0.5396	0.5267	0.5267	0.5923	0.5952	0.5917	0.5962	0.5895	**0.7686**

Table 5 Quantitative segmentation performance of proposed method and related methods on synthetic image dataset corrupted with Rician noise in terms of ASA

NL(σ)	kPC	kPPC	LkPPC	FCM	FkPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
10	0.9011	0.9036	0.9081	0.8946	0.8946	0.9606	0.9719	0.9697	0.9700	0.9620	**0.9911**
15	0.7703	0.7730	0.7975	0.7699	0.7699	0.8473	0.8595	0.8550	0.8506	0.8322	**0.9844**
20	0.6972	0.7026	0.7171	0.6992	0.6992	0.7540	0.7529	0.7509	0.7491	0.7461	**0.9416**
Avg. Performance	0.7895	0.7930	0.8076	0.7879	0.7879	0.8540	0.8614	0.8585	0.8566	0.8468	**0.9724**

MRI image corrupted with (7% & 9% noise level) shown in Figs. 4 and 5, respectively. It can be noted from Figs. 4 and 5 that the proposed FkPC_S method provide superior segmentation results in comparison with other related methods in the presence of high level of noise.

4.4 Results and discussion on real clinical MRI brain images

We perform the experiments on the real brain MRI images corrupted with Rician noise (σ = 5 and σ = 10). The quantitative performance is obtained on these images in terms of average segmentation accuracy (ASA) and Dice score (DS) for comparing the proposed FkPC_S method results with other related methods. The comparison of the proposed method with other existing methods in terms of

Table 6 Quantitative segmentation performance of proposed method and related methods on synthetic image dataset corrupted with Rician noise in terms of Dice score

NL(σ)	Cluster region	kPC	kPPC	LkPPC	FCM	FkPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
10	Region 1	0.9437	0.9437	0.9466	0.9375	0.9375	0.9769	0.9832	0.9817	0.9819	0.9779	**0.9961**
	Region 2	0.7143	0.7189	0.7338	0.6975	0.6975	0.8763	0.9088	0.9016	0.9012	0.8799	**0.9749**
	Region 3	0.8670	0.8803	0.8903	0.8729	0.8729	0.9539	0.9709	0.9695	0.9700	0.9548	**0.9819**
	Region 4	0.9280	0.9338	0.9303	0.9323	0.9323	0.9703	0.9784	0.9776	0.9792	0.9708	**0.9879**
15	Region 1	0.8749	0.8749	0.8923	0.8710	0.8710	0.9134	0.9179	0.9156	0.9155	0.9037	**0.9925**
	Region 2	0.4454	0.4525	0.5097	0.4497	0.4497	0.6064	0.6353	0.6247	0.6122	0.5802	**0.9560**
	Region 3	0.6069	0.6254	0.6859	0.6352	0.6352	0.7773	0.8123	0.8042	0.7831	0.7549	**0.9727**
	Region 4	0.8189	0.8217	0.7941	0.8236	0.8236	0.8782	0.8923	0.8891	0.8836	0.8713	**0.9792**
20	Region 1	0.8365	0.8400	0.8555	0.8365	0.8365	0.8636	0.8586	0.8582	0.8603	0.8601	**0.9679**
	Region 2	0.3303	0.3400	0.3821	0.3394	0.3394	0.4365	0.4427	0.4372	0.4255	0.4197	**0.8411**
	Region 3	0.4484	0.4637	0.5201	0.4711	0.4711	0.5906	0.6081	0.6014	0.5845	0.5722	**0.9173**
	Region 4	0.7380	0.7365	0.6653	0.7339	0.7339	0.7940	0.8069	0.8041	0.7997	0.7901	**0.9354**
Avg. performance		0.7127	0.7193	0.7338	0.7167	0.7167	0.8031	0.8180	0.8137	0.8081	0.7946	**0.9586**

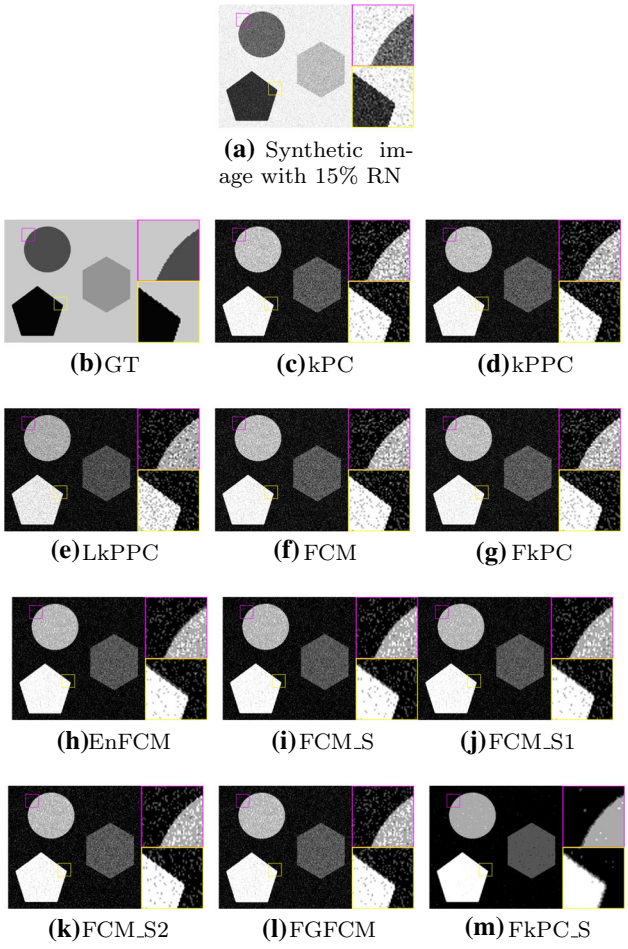


Fig. 2 Qualitative segmentation performance of proposed method and related methods on synthetic image dataset corrupted with 15% Rician noise

average ASA and average DS over all real brain MRI images is shown in Fig. 6 corresponding for different levels of Rician noise. The following observation can be drawn from Tables 9, 10, 11, 12, 13, 14, 15, 16 and Fig. 6 for the real brain MRI image:

- The proposed FkPC_S method performs better in comparison with other related method in terms of average ASA over all images for both 5% and 10% levels of noise.
- The proposed FkPC_S method performs better in comparison with other related method in terms of average DS of GM over all images for both 5% and 10% levels of noise.
- The proposed FkPC_S method performs better in comparison with other related method in terms of average DS of WM over all images for 10% level of noise.
- The proposed FkPC_S method performs better in comparison with other related method in terms of average DS of CSF over all images for both 5% and 10% levels of noise.

To show the effectiveness of the proposed (FkPC_S) method related to conventional FkPC method, we have calculated percentage increment in ASA as follows:

$$Percentage\ increment = \frac{ASA(FkPC_S) - ASA(FkPC)}{ASA(FkPC)} * 100 \tag{37}$$

The percentage increment in the ASA for the images corrupted with noise is shown in Fig. 7. We can observe from

Table 7 Quantitative segmentation performance of proposed method and related methods on synthetic Brainweb MRI image dataset in terms of ASA.

Image/methods	kPC	kPPC	LkPPC	FCM	FkPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
3% Noise	0.9358	0.9360	0.9360	0.9358	0.9358	0.9007	**0.9361**	0.9358	0.9360	0.8892	0.9359
5% Noise	0.9152	0.9157	0.9158	0.9160	0.9160	0.8932	0.9219	0.9219	0.9209	0.9062	**0.9224**
7% Noise	0.8787	0.8787	0.8787	0.8742	0.8742	0.8766	0.9029	0.9032	0.9024	0.8898	**0.9104**
9% Noise	0.8312	0.8312	0.8312	0.8207	0.8207	0.8641	0.8921	0.8922	0.8847	0.8729	**0.9081**
Avg. performance	0.8902	0.8904	0.8904	0.8867	0.8867	0.8836	0.9133	0.9133	0.9110	0.8895	**0.9192**

Table 8 Quantitative segmentation performance of proposed method and related methods on synthetic Brainweb MRI image dataset in terms of Dice score

Image/methods		kPC	kPPC	LkPPC	FCM	FkPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
GM	3% Noise	0.9392	0.9395	0.9395	0.9393	0.9393	0.8938	**0.9400**	0.9394	0.9397	0.8814	0.9397
	5% Noise	0.9116	0.9125	0.9125	0.9130	0.9130	0.8847	0.9205	0.9205	0.9191	0.9013	**0.9216**
	7% Noise	0.8663	0.8663	0.8663	0.8619	0.8619	0.8636	0.8971	0.8977	0.8967	0.8807	**0.9067**
	9% Noise	0.8055	0.8055	0.8055	0.7950	0.7950	0.8429	0.8778	0.8780	0.8692	0.8541	**0.8988**
WM	3% Noise	0.9774	0.9776	0.9776	0.9772	0.9772	0.9600	**0.9784**	0.9772	0.9773	0.9522	0.9777
	5% Noise	0.9615	0.9620	0.9621	0.9621	0.9621	0.9544	0.9682	0.9682	0.9666	0.9610	**0.9688**
	7% Noise	0.9328	0.9328	0.9328	0.9300	0.9300	0.9435	0.9559	0.9563	0.9532	0.9505	**0.9627**
	9% Noise	0.8936	0.8936	0.8936	0.8860	0.8860	0.9285	0.9437	0.9442	0.9357	0.9321	**0.9581**
CSF	3% Noise	0.8531	0.8534	0.8534	0.8532	0.8532	0.7837	0.8509	0.8534	**0.8540**	0.7622	0.8515
	5% Noise	0.8303	0.8307	0.8306	0.8313	0.8313	0.7747	0.8324	0.8324	**0.8333**	0.8006	0.8293
	7% Noise	0.7966	0.7966	0.7965	0.7911	0.7911	0.7548	**0.8074**	0.8072	0.8120	0.7789	0.8050
	9% Noise	0.7566	0.7566	0.7566	0.7458	0.7458	0.7592	0.8082	0.8071	0.8044	0.7792	**0.8098**
Avg. performance		0.8771	0.8773	0.8772	0.8738	0.8738	0.8620	0.8984	0.8985	0.8968	0.8695	**0.9025**

Fig. 7a for simulated Brain web images corrupted with Rician noise (σ = 3, σ = 5, σ = 7 and σ = 9) that the percentage increment in the ASA increases as the level of Rician noise increases. Similar observation can be noted from Fig. 7b for IBSR real MRI images corrupted with Rician noise (σ = 5 and σ = 10). This effectiveness is due to the incorporation of spatial regularization term in conventional FkPC method that reduces the effect of noise and provides better segmentation. This suggests that the proposed method will be effective when acquired MRI images are embedded with Rician noise.

4.5 Computational complexity

Let I be the maximum number of iterations, N be the number of pixels of the image, d be the number of channels, and k be the number of possible regions in the image. The computational burden of the proposed method involves the calculation of the plane prototype and the partition matrix. The computational complexity for calculating the plane prototype $\mathbf{w_j}$ involves solving the eigenvalue problem given by Eq. (30) which take $O(d^3)$ time. The calculation of b_j takes $O(dN)$ time which is easily inferred from Eq. (24). Thus, the time complexity of each plane parameter is $O(k(dN + d^3))$ for generating all k clustering planes. The calculation of membership value μ_{ij} calculation using Eq. (36) takes $O(kdN)$. Since I iterations are required to obtain the converged values of cluster prototype and membership value, the overall computational complexity of the proposed method requires $O(I(k(dN + d^3 + dN))$. We present comparison of the proposed method in comparison with 10 existing methods in terms of average computation time of 100 runs on a simulated brain MRI of size 181 $\times$ 217 in Table 17. We can observe from Table 17 that the proposed FkPC_S Method takes less computation time in comparison with five methods (LkPPC, FCM, FCM_S, FCM_S1 and FCM_S2) and more compared to five methods (kPC, KPPC, FkPC, EnFCM and FGFCM).

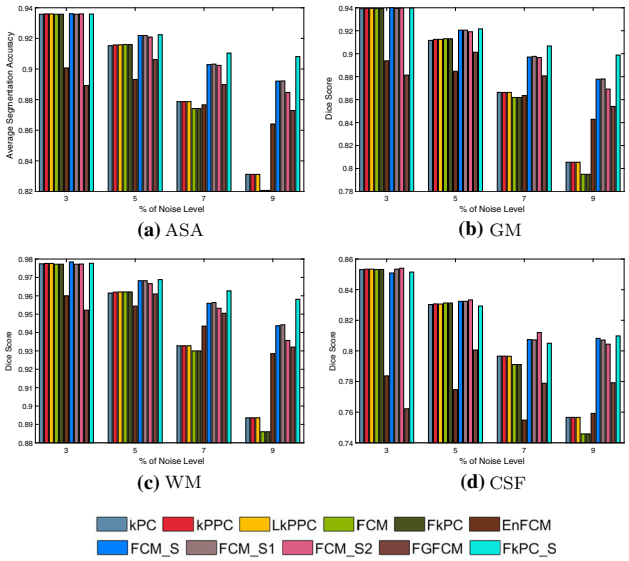


Fig. 3 Qualitative segmentation performance of proposed method and related methods on simulated brain image dataset with 3%, 5%, 7% and 9% noise level in terms of a) ASA b) DS for GM c) DS for WM d) DS for CSF

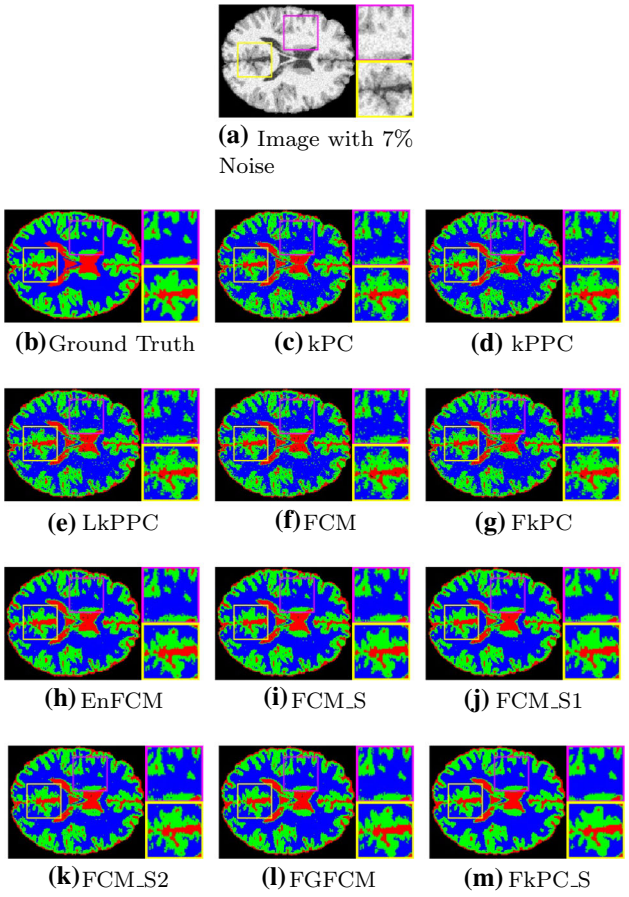


Fig. 4 Qualitative segmentation performance of proposed method and related methods on simulated brain web image dataset corrupted with 7% Rician noise

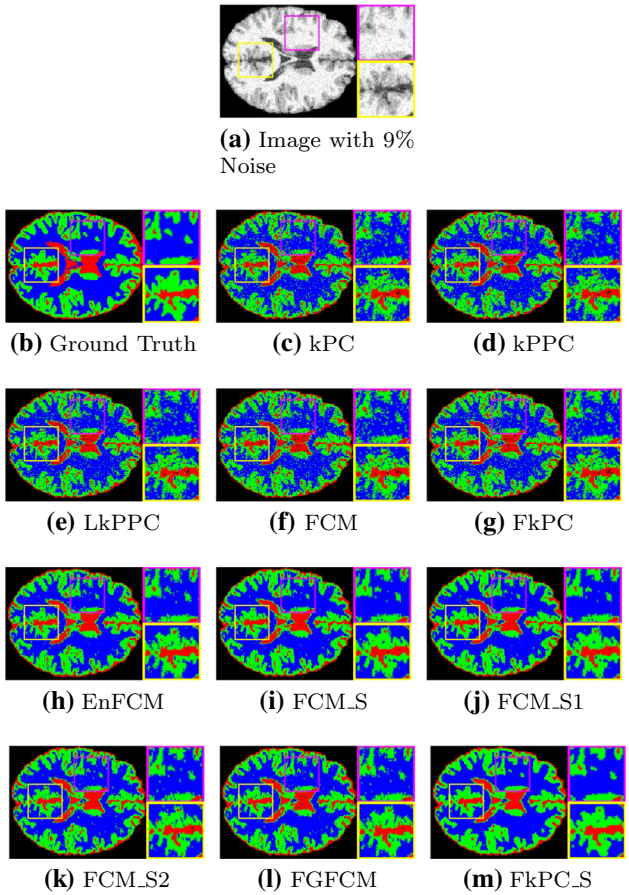


Fig. 5 Qualitative segmentation performance of proposed method and related methods on simulated brain web image dataset corrupted with 9% Rician noise

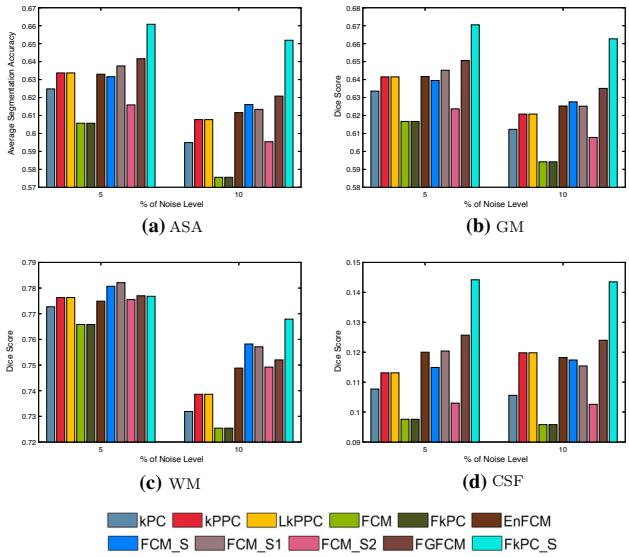


Fig. 6 Qualitative segmentation performance of proposed method and related methods on real brain IBSR image dataset with 5% & 10% Rician noise in terms of **a** ASA **b** DS for GM **c** DS for WM **d** DS for CSF

Table 9 Quantitative segmentation performance of proposed method and related methods on real IBSR brain MRI image dataset with Rician noise (NL = 5) in terms of ASA

Images/methods	kPC	kPPC	LkPPC	FCM	F_kPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
111_2	0.6618	0.6622	0.6622	0.6526	0.6526	0.6607	0.6621	0.6630	0.6577	**0.6694**	0.6683
112_2	0.6555	0.6640	0.6640	0.6354	0.6354	0.6682	0.6579	0.6634	0.6389	0.6699	**0.7081**
11_3	0.6154	0.6222	0.6222	0.6040	0.6040	0.6180	0.6132	0.6158	0.6033	0.6210	**0.6262**
12_3	0.6802	0.6860	0.6860	0.6559	0.6560	0.6900	0.6899	0.6988	0.6678	0.7009	**0.7459**
13_3	0.6488	0.6552	0.6552	0.6333	0.6333	0.6653	0.6635	0.6693	0.6391	0.6695	**0.6829**
16_3	0.6450	0.6644	0.6644	0.6220	0.6220	0.6513	0.6581	0.6654	0.6487	0.6599	**0.6737**
17_3	0.6751	0.6769	0.6769	0.6721	0.6721	0.6870	0.6789	0.6824	0.6759	**0.6953**	**0.6953**
205_3	0.6395	0.6434	0.6434	0.6138	0.6138	0.6270	0.6319	0.6369	0.6203	0.6436	**0.6873**
4_8	0.5351	0.5366	0.5366	0.5270	0.5270	0.5434	0.5461	0.5499	0.5362	0.5515	**0.5573**
5_8	0.5299	0.5301	0.5301	0.4965	0.4965	0.5393	0.5452	0.5559	0.5173	0.5490	**0.6012**
6_10	0.5193	**0.5709**	**0.5709**	0.4896	0.4896	0.5362	0.5391	0.5468	0.5142	0.5351	0.5695
8_4	0.6918	0.6922	0.6922	0.6665	0.6665	0.7093	0.6935	0.7032	0.6715	**0.7340**	0.7135
Avg. Performance	0.6248	0.6337	0.6337	0.6057	0.6057	0.6330	0.6316	0.6376	0.6159	0.6416	**0.6608**

Table 10 Quantitative segmentation performance of proposed method and related methods on real IBSR brain MRI image dataset with Rician noise (NL = 10) in terms of ASA

Images/methods	kPC	kPPC	LkPPC	FCM	F_kPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
111_2	0.6440	0.6507	0.6507	0.6268	0.6268	0.6530	0.6556	0.6520	0.6427	0.6548	**0.6632**
112_2	0.6501	0.6534	0.6534	0.6192	0.6192	0.6614	0.6598	0.6570	0.6302	0.6636	**0.7155**
11_3	0.6040	0.6088	0.6088	0.5854	0.5854	0.6100	0.6104	0.6090	0.5961	0.6082	**0.6279**
12_3	0.6400	0.6430	0.6430	0.6203	0.6203	0.6699	0.6733	0.6688	0.6445	0.6718	**0.7390**
13_3	0.6208	0.6233	0.6233	0.6024	0.6024	0.6452	0.6553	0.6511	0.6258	0.6585	**0.6761**
16_3	0.5504	0.5890	0.5890	0.5349	0.5349	0.5909	0.6040	0.5979	0.5776	0.6044	**0.6536**
17_3	0.6632	0.6638	0.6638	0.6502	0.6502	0.6737	0.6845	0.6843	0.6663	0.6890	**0.6979**
205_3	0.6100	0.6119	0.6119	0.5848	0.5848	0.6118	0.6234	0.6222	0.6014	0.6210	**0.6728**
4_8	0.5346	**0.5995**	**0.5995**	0.5170	0.5170	0.5444	0.5479	0.5462	0.5348	0.5511	0.5660
5_8	0.4805	0.4835	0.4835	0.4596	0.4596	0.4915	0.4966	0.4939	0.4780	0.5037	**0.5407**
6_10	0.4673	0.4890	0.4890	0.4502	0.4502	0.4881	0.4905	0.4884	0.4750	0.4994	**0.5601**
8_4	0.6741	0.6766	0.6766	0.6559	0.6559	0.6993	0.6921	0.6900	0.6719	**0.7242**	0.7104
Avg. Performance	0.5949	0.6077	0.6077	0.5756	0.5756	0.6116	0.6161	0.6134	0.5954	0.6208	**0.6519**

4.6 Statistical test

To check the statistical significant difference between the proposed method and other related method, we employ a non-parametric Friedman test [16]. For this, we formulate null hypothesis denoted by (H_0) which states that there is no statistical significant difference between the performance of the proposed method and other related methods. On the other hand, (H_1) is the alternative hypothesis, which states that there is statistical significant difference between the performance of the proposed and other related methods. For a given performance measure, Null (H_0), alternated hypothesis (H_1) can be expressed as:

$$\begin{aligned} H_0 &: \mu_{FkPC_S} = \mu_{FGFCM} = \mu_{FCM_S2} \\ &= \mu_{FCM_S1} = \mu_{FCM_S} = \mu_{EnFCM} = \mu_{FkPC} \\ &= \mu_{FCM} = \mu_{LkPPC} = \mu_{kPPC} = \mu_{kPC} \\ H_1 &: \mu_{FkPC_S} \neq \mu_{FGFCM} \neq \mu_{FCM_S2} \\ &\neq \mu_{FCM_S1} \neq \mu_{FCM_S} \neq \mu_{EnFCM} \neq \mu_{FkPC} \\ &\neq \mu_{FCM} \neq \mu_{LkPPC} \neq \mu_{kPPC} \neq \mu_{kPC} \end{aligned}$$

To rank the proposed method and other related methods, we employ ASA performance obtained on several MRI images. The average ranking is computed by:

Table 11 Quantitative segmentation performance of proposed method and related methods on real IBSR brain MRI image dataset with Rician noise (NL = 5) in terms of Dice score for GM

Images/methods	kPC	kPPC	LkPPC	FCM	F_kPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
111_2	0.6621	0.6626	0.6626	0.6541	0.6541	0.6618	0.6626	0.6631	0.6589	**0.6720**	0.6706
112_2	0.6708	0.6790	0.6790	0.6533	0.6533	0.6849	0.6766	0.6816	0.6565	0.6865	**0.7295**
11_3	0.6367	0.6434	0.6434	0.6276	0.6276	0.6410	0.6357	0.6380	0.6260	0.6449	**0.6503**
12_3	0.7251	0.7312	0.7312	0.7001	0.7002	0.7372	0.7361	0.7457	0.7132	0.7491	**0.7911**
13_3	0.6467	0.6531	0.6531	0.6321	0.6321	0.6684	0.6647	0.6714	0.6375	0.6730	**0.6899**
16_3	0.6419	0.6605	0.6605	0.6228	0.6228	0.6513	0.6531	0.6606	0.6431	0.6594	**0.6712**
17_3	0.6694	0.6704	0.6704	0.6696	0.6696	0.6826	0.6747	0.6776	0.6717	0.6910	**0.6942**
205_3	0.6219	0.6267	0.6267	0.5944	0.5944	0.6094	0.6159	0.6221	0.6001	0.6321	**0.6838**
4_8	0.5589	0.5569	0.5569	0.5557	0.5557	0.5530	0.5585	**0.5597**	0.5559	0.5575	0.5585
5_8	0.5495	0.5497	0.5497	0.5226	0.5226	0.5580	0.5580	0.5677	0.5305	0.5660	**0.6144**
6_10	0.5298	0.5738	0.5738	0.5037	0.5037	0.5433	0.5457	0.5517	0.5216	0.5417	**0.5785**
8_4	0.6908	0.6906	0.6906	0.6649	0.6649	0.7088	0.6925	0.7028	0.6699	**0.7342**	0.7144
Avg. performance	0.6336	0.6415	0.6415	0.6167	0.6167	0.6417	0.6395	0.6452	0.6237	0.6506	**0.6705**

Table 12 Quantitative segmentation performance of proposed method and related methods on real IBSR brain MRI image dataset with Rician noise (NL = 10) in terms of Dice score for GM

Images/methods	kPC	kPPC	LkPPC	FCM	F_kPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
111_2	0.6471	0.6539	0.6539	0.6305	0.6305	0.6549	0.6578	0.6538	0.6444	0.6564	**0.6652**
112_2	0.6665	0.6692	0.6692	0.6360	0.6360	0.6807	0.6772	0.6743	0.6464	0.6818	**0.7356**
11_3	0.6319	0.6358	0.6358	0.6138	0.6139	0.6363	0.6371	0.6358	0.6229	0.6344	**0.6525**
12_3	0.6840	0.6874	0.6874	0.6646	0.6646	0.7148	0.7185	0.7136	0.6880	0.7158	**0.7852**
13_3	0.6211	0.6237	0.6237	0.6025	0.6025	0.6467	0.6577	0.6538	0.6256	0.6635	**0.6816**
16_3	0.5752	0.6063	0.6063	0.5589	0.5589	0.6019	0.6059	0.6007	0.5855	0.6148	**0.6529**
17_3	0.6614	0.6621	0.6621	0.6523	0.6523	0.6715	0.6823	0.6820	0.6627	0.6915	**0.6976**
205_3	0.5980	0.6003	0.6003	0.5727	0.5727	0.5960	0.6102	0.6089	0.5842	0.6086	**0.6657**
4_8	0.5660	**0.5926**	**0.5926**	0.5535	0.5535	0.5669	0.5656	0.5651	0.5606	0.5731	0.5727
5_8	0.5213	0.5236	0.5236	0.5030	0.5030	0.5266	0.5273	0.5249	0.5129	0.5387	**0.5662**
6_10	0.5012	0.5175	0.5175	0.4854	0.4854	0.5055	0.5006	0.5002	0.4910	0.5171	**0.5648**
8_4	0.6740	0.6770	0.6770	0.6572	0.6572	0.7017	0.6913	0.6892	0.6700	**0.7252**	0.7121
Avg. performance	0.6123	0.6208	0.6208	0.5942	0.5942	0.6253	0.6276	0.6252	0.6078	0.6351	**0.6627**

$$R_j = \frac{1}{M}\sum_{i=1}^{M} r_i^j \tag{38}$$

where R_j symbol represents the average rank of j^{th} method and M is total number of images taken into consideration and $(r_i^j, 1 \le i \le M, 1 \le j \le t)$ denotes the rank of j^{th} method on i^{th} image in terms of the performance measure, ASA. We consider 6 synthetic images corrupted with noise, 4 Brain web images with (3%,5%,7% & 9%) Rician noise, and 12 Real brain IBSR images with 10% Rician noise for ranking. The average Friedman ranking of the proposed FkPC_S method and other related methods is reported in Table 18. The low numerical value in Table 18 represents better algorithm. It can be observed from Table 18 that the proposed FkPC_S method performs better in comparison with other related methods.

The statistic measure F_{ID}, suggested by Iman and Davenport [18], is computed:

$$F_{ID} = \frac{(M-1)\chi_F^2}{M(t-1)-\chi_F^2} \tag{39}$$

which is distributed according to F-distribution with $t-1$ and $(t-1)(M-1)$ degrees of freedom and χ_F^2 is the Friedman's statistic defined as

Table 13 Quantitative segmentation performance of proposed method and related methods on real IBSR brain MRI image dataset with Rician noise (NL = 5) in terms of Dice score for WM

Images/methods	kPC	kPPC	LkPPC	FCM	F_kPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
111_2	0.8430	0.8430	0.8430	0.8457	0.8457	0.8382	0.8460	0.8443	**0.8471**	0.8406	0.8298
112_2	**0.8276**	0.8269	0.8269	0.8194	0.8194	0.8262	0.8216	0.8230	0.8200	0.8273	0.8216
11_3	0.7635	0.7642	0.7642	0.7663	0.7663	0.7666	0.7689	0.7688	0.7670	**0.7694**	0.7594
12_3	0.8162	0.8177	0.8177	0.8128	0.8128	0.8150	0.8220	**0.8232**	0.8173	0.8189	0.8137
13_3	0.8539	0.8572	0.8572	0.8444	0.8444	0.8499	0.8592	**0.8604**	0.8494	0.8528	0.8511
16_3	0.7918	0.7920	0.7920	0.7853	0.7853	0.7922	0.8041	**0.8061**	0.8017	0.7922	0.7966
17_3	0.7977	0.7965	0.7965	0.8001	0.8001	0.7949	0.7967	0.7941	**0.8029**	0.7945	0.7792
205_3	0.7855	**0.7861**	**0.7861**	0.7791	0.7791	0.7778	0.7836	0.7848	0.7828	0.7812	0.7820
4_8	0.6598	0.6616	0.6616	0.6553	0.6553	0.6720	0.6753	0.6777	0.6661	0.6763	**0.6681**
5_8	0.6753	0.6756	0.6756	0.6455	0.6455	0.6883	0.7047	0.7107	0.6820	0.6964	**0.7277**
6_10	0.6602	**0.6981**	**0.6981**	0.6367	0.6367	0.6783	0.6839	0.6891	0.6667	0.6772	0.6958
8_4	0.7982	0.7971	0.7971	0.7986	0.7986	0.7994	0.8018	**0.8033**	0.8025	0.7974	0.7968
Avg. performance	0.7727	0.7763	0.7763	0.7658	0.7658	0.7749	0.7807	0.7821	0.7755	0.7770	**0.7768**

Table 14 Quantitative segmentation performance of proposed method and related methods on real IBSR brain MRI image dataset with Rician noise (NL = 10) in terms of Dice score for WM

Images/methods	kPC	kPPC	LkPPC	FCM	F_kPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
111_2	0.8165	0.8184	0.8184	0.8141	0.8141	0.8275	**0.8343**	0.8331	0.8321	0.8291	0.8279
112_2	0.8107	0.8116	0.8116	0.8009	0.8009	0.8139	0.8226	**0.8227**	0.8138	0.8176	0.8215
11_3	0.7454	0.7451	0.7451	0.7439	0.7439	0.7606	**0.7613**	0.7608	0.7571	0.7581	0.7576
12_3	0.7783	0.7786	0.7786	0.7728	0.7728	0.7957	0.8026	0.8015	0.7955	0.8004	**0.8092**
13_3	0.8170	0.8175	0.8175	0.8077	0.8077	0.8357	0.8433	0.8410	0.8312	0.8382	**0.8477**
16_3	0.6835	0.7087	0.7087	0.6813	0.6813	0.7277	0.7551	0.7504	0.7360	0.7355	**0.7812**
17_3	0.7815	0.7816	0.7816	0.7773	0.7773	0.7840	0.7934	**0.7948**	0.7925	0.7864	0.7833
205_3	0.7629	0.7631	0.7631	0.7543	0.7543	0.7654	0.7735	0.7734	0.7675	0.7684	**0.7790**
4_8	0.6395	**0.6678**	**0.6678**	0.6319	0.6319	0.6515	0.6630	0.6617	0.6505	0.6557	0.6666
5_8	0.5931	0.5949	0.5949	0.5759	0.5759	0.6143	0.6257	0.6240	0.6085	0.6220	**0.6637**
6_10	0.5722	0.5927	0.5927	0.5617	0.5617	0.6151	0.6288	0.6265	0.6137	0.6205	**0.6826**
8_4	0.7828	0.7826	0.7826	0.7834	0.7834	0.7938	0.7944	**0.7953**	0.7919	0.7926	0.7940
Avg. Performance	0.7319	0.7386	0.7386	0.7254	0.7254	0.7488	0.7582	0.7571	0.7492	0.7520	**0.7679**

$$\frac{12M}{t(t+1)}\left[\sum_j R_j^2 - \frac{t(t+1)^2}{4}\right] \tag{40}$$

where t is total number of methods considered in comparison and M number is total number of images taken into consideration for calculating the average rank. In our experiments, $t = 11$ and $M = 22$. The p-value obtained by Iman and davenport statistic measure is 3.13E-64 which endorses the rejection of null hypothesis H_0 at the significance level of 0.05.

However, the p-value obtained is not reasonable for comparison with the control method, i.e. the one that emerges with the lowest rank. So adjusted p-value is calculated which take into account the accumulated error and provide the correct correlation. These adjusted p-values [13] are calculated for different related methods with respect to the proposed FkPC_S method (lowest rank). For this, a set of post-hoc procedures are defined and adjusted p-value is computed using a popular Holm procedure is reported in Table 19. It can be noted that there is a statistical difference in terms of ASA between the proposed FkPC_S method and other related methods.

Table 15 Quantitative segmentation performance of proposed method and related methods on real IBSR brain MRI image dataset with Rician noise (NL = 5) in terms of Dice score for CSF

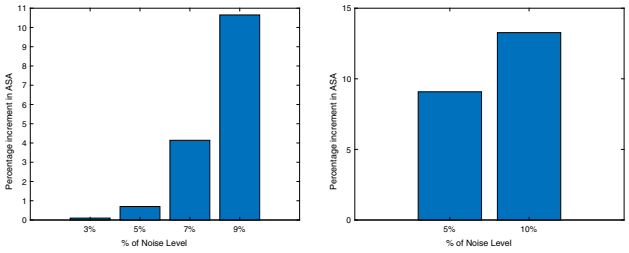
Images/methods	kPC	kPPC	LkPPC	FCM	F_kPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
111_2	0.0972	0.0974	0.0974	0.0911	0.0911	0.1067	0.1025	0.1031	0.0952	0.1096	**0.1216**
112_2	0.0848	0.0888	0.0888	0.0802	0.0802	0.1022	0.0913	0.0943	0.0821	0.1038	**0.1123**
11_3	0.1229	0.1271	0.1271	0.1118	0.1118	0.1304	0.1199	0.1251	0.1109	0.1267	**0.1453**
12_3	0.0722	0.0740	0.0740	0.0599	0.0599	0.0801	0.0772	0.0831	0.0633	0.0851	**0.1428**
13_3	0.0477	0.0487	0.0487	0.0464	0.0464	0.0543	0.0523	0.0539	0.0487	0.0571	**0.0586**
16_3	0.1841	0.2007	0.2007	0.1671	0.1671	0.1833	0.2029	**0.2055**	0.1891	0.1870	0.1924
17_3	0.1612	0.1652	0.1652	0.1570	0.1570	0.1909	0.1704	0.1843	0.1616	0.1958	**0.1900**
205_3	0.1113	0.1120	0.1120	0.0967	0.0967	0.1192	0.1110	0.1151	0.1017	0.1278	**0.1803**
4_8	0.1133	0.1162	0.1162	0.1022	0.1022	0.1317	0.1324	0.1403	0.1142	0.1418	**0.1781**
5_8	0.0655	0.0655	0.0655	0.0604	0.0604	0.0719	0.0712	0.0756	0.0629	0.0672	**0.1111**
6_10	0.0950	0.1225	0.1225	0.0833	0.0833	0.1049	0.1046	0.1115	0.0896	0.1045	**0.1339**
8_4	0.1372	0.1390	0.1390	0.1147	0.1147	0.1646	0.1432	0.1527	0.1160	**0.2019**	0.1645
Avg. performance	0.1077	0.1131	0.1131	0.0976	0.0976	0.1200	0.1149	0.1204	0.1030	0.1257	**0.1442**

Table 16 Quantitative segmentation performance of proposed method and related methods on real IBSR brain MRI image dataset with Rician noise (NL = 10) in terms of Dice score for CSF

Images/methods	kPC	kPPC	LkPPC	FCM	F_kPC	EnFCM	FCM_S	FCM_S1	FCM_S2	FGFCM	FkPC_S
111_2	0.0977	0.1003	0.1003	0.0898	0.0898	0.1046	0.1027	0.1003	0.0932	0.1067	**0.1217**
112_2	0.0873	0.0873	0.0873	0.0770	0.0770	0.1042	0.0946	0.0924	0.0776	0.1031	**0.1161**
11_3	0.1163	0.1212	0.1212	0.1073	0.1074	0.1208	0.1199	0.1198	0.1076	0.1219	**0.1472**
12_3	0.0732	0.0742	0.0742	0.0644	0.0644	0.0910	0.0867	0.0819	0.0698	0.0901	**0.1473**
13_3	0.0475	0.0479	0.0479	0.0455	0.0455	0.0519	0.0531	0.0525	0.0496	0.0553	**0.0578**
16_3	0.1392	0.1655	0.1655	0.1334	0.1334	0.1657	0.1794	0.1762	0.1587	0.1725	**0.1849**
17_3	0.1559	0.1564	0.1564	0.1457	0.1457	0.1794	0.1772	0.1767	0.1536	0.1823	**0.1835**
205_3	0.1027	0.1029	0.1029	0.0912	0.0912	0.1155	0.1109	0.1109	0.0984	0.1188	**0.1691**
4_8	0.1386	**0.2662**	**0.2662**	0.1176	0.1176	0.1456	0.1503	0.1465	0.1323	0.1548	0.1893
5_8	0.0688	0.0699	0.0699	0.0645	0.0645	0.0722	0.0807	0.0784	0.0679	0.0739	**0.1016**
6_10	0.1031	0.1112	0.1112	0.0940	0.0940	0.1077	0.1073	0.1041	0.0936	0.1144	**0.1395**
8_4	0.1367	0.1349	0.1349	0.1190	0.1190	0.1597	0.1460	0.1448	0.1290	**0.1941**	0.1644
Avg. performance	0.1056	0.1198	0.1198	0.0958	0.0958	0.1182	0.1174	0.1154	0.1026	0.1240	**0.1435**

5 Conclusion and future direction

In this research work, a fuzzy k-plane with local spatial information (FkPC_S) clustering method is proposed to handle noise for MRI image segmentation. The FkPC_S method is formulated by modifying the FkPC method objective function with a spatial regularization term to address noise artefacts in the segmentation procedure. In the presence of noise, the performance of image segmentation degrades. The proposed FkPC_S method takes advantage of the spatial regularization term to retain information in a noisy environment and produce better segmentation performance. Experimental results are presented in terms of both quantitative and qualitative measures on the synthetic image with different noise levels and two publicly available MRI datasets. Experimental results demonstrate the effectiveness of the proposed FKPC_S method compared to 10 existing related algorithms in terms of average segmentation accuracy and Dice score. The computation time of the proposed FkPC_S method is also compared with existing methods, which is better than 5 existing methods and inferior than 5 existing methods. A statistical test is also performed to investigate the statistically significant difference between the performance of the proposed method and other 10 related methods. The statistical test endorses the proposed FkPC_S



(a) Brainweb dataset **(b)** IBSR dataset

Fig. 7 Percentage increment in segmentation performance in terms of ASA of the proposed method (FkPC_S) with respect to the conventional method FkPC method for **a** simulated Brainweb MRI dataset with 3%, 5%, 7% and 9% noise level, **b** IBSR real MRI dataset with 5% and 10% noise level on overall MRI dataset on average

Table 17 Comparison of computation time

Algorithms	Time (in sec.)
kPC	0.0607
kPPC	0.0743
LkPPC	0.5107
FCM	0.2578
FkPC	0.1457
EnFCM	0.0746
FCM_S	0.2827
FCM_S1	0.2631
FCM_S2	0.2709
FGFCM	0.2247
FkPC_S	0.2399

Table 18 Average Ranking of the algorithms (Friedman)

Algorithm	Ranking
kPC	7.84
kPPC	5.99
LkPPC	5.60
FCM	10.10
FkPC	10.10
EnFCM	6.00
FCMS	3.56
FCMS1	4.54
FCMS2	6.20
FGFCM	4.56
FkPC_S	1.51

method, which shows superior performance of the proposed method over other related methods.

The limitation of the proposed FkPC_S method is the manual tuning of spatial regularization parameter (α), which is essential to obtain an accurate segmentation result. Due to this additional term, the computation time of our proposed method (FkPC_S) is more than the FkPC method. Our proposed FkPC_S method works for imaging data only, while the FkPC method work for all kinds of data.

In future, it will be interesting to extend the proposed method in intuitionistic fuzzy metric space. There is a need to find a way to bound the plane prototype to avoid the infinite extending problem of the Fuzzy k-plane clustering method. Further, we can extend the proposed method for multi-modal imaging data.

Table 19 Adjusted p-values (Friedman)

Algortihm	Unadjusted P value	Pholm value
kPC	1.44E-15	1.15E-14
kPPC	1.70E-08	9.19E-08
LkPPC	2.56E-07	1.02E-06
FCM	2.50E-27	2.50E-26
FkPC	2.50E-27	2.50E-26
EnFCM	1.53E-08	9.19E-08
FCMS	9.98E-03	9.98E-03
FCMS1	1.33E-04	3.72E-04
FCMS2	3.42E-09	2.39E-08
FGFCM	1.24E-04	3.72E-04

Author Contributions All authors contributed to the study conception and design. Material preparation and analysis were performed by Puneet Kumar, Dhirendra Kumar, and R. K. Agrawal. The first draft of the manuscript was written by Puneet Kumar, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The first author has received UGC funding for this research.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Ahmed MN, Yamany SM, Mohamed N, Farag AA, Moriarty T (2002) A modified fuzzy c-means algorithm for bias field estimation and segmentation of mri data. IEEE Tran Med Imaging 21(3):193–199
2. Aiello M, Cavaliere C, D'Albore A, Salvatore M (2019) The challenges of diagnostic imaging in the era of big data. J Clin Med 8(3):316

3. Bai X, Zhang Y, Liu H, Chen Z (2018) Similarity measure-based possibilistic fcm with label information for brain mri segmentation. IEEE Trans cybern 49(7):2618–2630
4. Bangiyev S, Chang D, Urbanek D E, Dickey RC, Reti R, Findlay D (2021) Neurological disorders. In: Oral board review for oral and maxillofacial surgery, pp. 429–441. Springer
5. Bezdek JC, Coray C, Gunderson R, Watson J (1981) Detection and characterization of cluster substructure i. linear structure: fuzzy c-lines. SIAM J Appl Math 40(2):339–357
6. Bezdek JC, Ehrlich R, Full W (1984) Fcm: the fuzzy c-means clustering algorithm. Comput Geosci 10(2–3):191–203
7. Bora DJ, Gupta D, Kumar A (2014) A comparative study between fuzzy clustering algorithm and hard clustering algorithm. arXiv preprint arXiv:1404.6059
8. Bradley PS, Mangasarian OL (2000) K-plane clustering. J Global Opt 16(1):23–32
9. Cai W, Chen S, Zhang D (2007) Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. Pattern Recognit 40(3):825–838
10. Chen S, Zhang D (2004) Robust image segmentation using fcm with spatial constraints based on new kernel-induced distance measure. IEEE Trans Syst Man Cybern Part B Cybern 34(4):1907–1916
11. Cocosco CA, Kollokian V, Kwan RKS, Pike GB, Evans AC (1997) Brainweb: Online interface to a 3d mri simulated brain database. In: NeuroImage. Citeseer
12. Dash S, Shakyawar SK, Sharma M, Kaushik S (2019) Big data in healthcare: management, analysis and future prospects. J Big Data 6(1):1–25
13. Derrac J, García S, Molina D, Herrera F (2011) A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evolut Comput 1(1):3–18
14. Dhanachandra N, Chanu YJ (2017) A survey on image segmentation methods using clustering techniques. Eur J Eng Tech Res 2(1):15–20
15. Doi K (2007) Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Gr 31(4–5):198–211
16. Friedman M (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Ass 32(200):675–701
17. Hartigan JA, Wong MA (1979) Ak-means clustering algorithm. J R Stat Soc Ser C Appl Stat 28(1):100–108
18. Iman RL, Davenport JM (1980) Approximations of the critical region of the fbietkan statistic. Commun Stat Theory Methods 9(6):571–595
19. Kumar D, Agrawal R, Verma H (2020) Kernel intuitionistic fuzzy entropy clustering for mri image segmentation. Soft Comput 24(6):4003–4026
20. Kumar D, Agrawal RK, Kirar JS (2019) Intuitionistic fuzzy clustering method with spatial information for mri image segmentation. In: 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–7. IEEE
21. Kumar D, Agrawal RK, Kumar P (2020) Bias-corrected intuitionistic fuzzy c-means with spatial neighborhood information approach for human brain mri image segmentation. IEEE Transactions on Fuzzy Systems
22. Kumar D, Verma H, Mehra A, Agrawal R (2019) A modified intuitionistic fuzzy c-means clustering approach to segment human brain mri image. Multimed Tools Appl 78(10):12663–12687
23. Liu J, Pham TD, Yan H, Liang Z (2018) Fuzzy mixed-prototype clustering algorithm for microarray data analysis. Neurocomputing 276:42–54
24. Liu LM, Guo YR, Wang Z, Yang ZM, Shao YH (2017) k-proximal plane clustering. Int J Mach Learn Cybern 8(5):1537–1554
25. The Mathworks, Inc., Natick, Massachusetts: MATLAB version 9.10.0.1613233 (R2021a) (2021)
26. Pan ZLWSt, Bin YhH (2008) Improved fuzzy partitions for k-plane clustering algorithm and its robustness research. J Electr Inf Tech. Vol. 8
27. Saxena A, Prasad M, Gupta A, Bharill N, Patel OP, Tiwari A, Er MJ, Ding W, Lin CT (2017) A review of clustering techniques and developments. Neurocomputing 267:664–681
28. Shehab N, Badawy M, Arafat H (2021) Big data analytics and preprocessing. In: Machine learning and big data analytics paradigms: analysis, applications and challenges, pp. 25–43. Springer
29. Smith S M (2000) Bet: brain extraction tool. FMRIB TR00SMS2b, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain), Department of Clinical Neurology, Oxford University, John Radcliffe Hospital, Headington, UK
30. Suhang G, Nojima Y, Ishibuchi H, Wang S (2020) Fuzzy style k-plane clustering. IEEE Transactions on Fuzzy Systems
31. Szilagyi L, Benyo Z, Szilágyi SM, Adam H (2003) Mr brain image segmentation using an enhanced fuzzy c-means algorithm. In: Proceedings of the 25th annual international conference of the IEEE engineering in medicine and biology society (IEEE Cat. No. 03CH37439), vol. 1, pp. 724–726. IEEE
32. Verma H, Gupta A, Kumar D (2019) A modified intuitionistic fuzzy c-means algorithm incorporating hesitation degree. Pattern Recogn Lett 122:45–52
33. Xie XL, Beni G (1991) A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 13(8):841–847
34. Yang ZM, Guo YR, Li CN, Shao YH (2015) Local k-proximal plane clustering. Neural Comput Appl 26(1):199–211
35. Zadeh LA (1996) Fuzzy sets. In: Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh, pp. 394–432. World Scientific
36. Zhang Y, Bai X, Fan R, Wang Z (2018) Deviation-sparse fuzzy c-means with neighbor information constraint. IEEE Trans Fuzzy Syst 27(1):185–199

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2021 6th International Conference on Signal Processing, Computing and Control (ISPCC) | 978-1-6654-2554-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISPCC53510.2021.9609429

Highly Compact Triple Ring Slotted Circular Polarized Implantable Antenna for Bio-medical Applications

Rajiv Kumar Nehra
Electronics and Communication Engineering(GGSIPU)
Bharatividyapeeth's College of Engineering,New Delhi (GGSIPU)
New Delhi, India
rajiv.nehra@bharatividyapeeth.edu

N S Raghava
Electronics and Communication Engineering
Delhi Technological University (DTU)
New Delhi, India
nsraghava@dtu.ac.in

***Abstract*— In this paper, A biomedical implantable antenna is submitted with a highly compact size of 9.77 mm3 and a facile design that exhibits ISM band characteristics (2.39 GHz). The proposed three circular rings implantable antenna has a minimum volume of 9.77mm3 which is the smallest volume of all implantable antennas yet. The recommended antenna has an axial ratio below 3dB from 2.38 GHz to 2.67 GHz (ISM Band) which results in circular polarisation, shows good impedance bandwidth from 2.32 GHz to 2.47 GHz (6.4%). The antenna has a peak gain value of -27.04 dB at 2.39 GHz and achieved a SAR value of 944.2 W/KG for 1 gram of tissue which is acceptable as per IEEE standard. Simulation of the proposed antenna is carried out at a skin depth of 3 mm. The antenna is simulated on the HFSS tool and shows good performance in all aspects of implantable medical devices.**

Keywords*— Industrial Scientific and Medical (ISM), Specific Absorption Rate (SAR), Axial Ratio, Impedance Bandwidth.

I. INTRODUCTION

Human health in the world is a great concern for all of us. Thousands of people are dying every day from various diseases. A lot of Deaths are happened due to unawareness of internal organ check-ups of patients. So, there is a technology called biomedical device (BMDs) which is implanted inside the human body and suppose to diagnose the internal organs of patients. In biomedical device, we must use an antenna which helps to establish the wireless communication between BMDs and external receiving station. Due to this reason, we call such an antenna an implantable biomedical antenna. Since this type of antenna is implanted inside the human body so the size of these antennas is of great concern. Hence implantable antenna must be having a highly compact size (miniaturized). During the telemetry sessions, patients should enjoy mobility instead of being in a rigid direction. So, the antenna must be designed with circular polarisation having an axial ratio below 3db. Specific absorption rate (SAR) is also an important concern parameter which may cause serious problem to patients. As per IEEE, there are two standards adopted one for 1gram of tissue and the other for 10 grams of tissue [1-2]. Since an Implantable antenna is nothing but conducting material so antenna must be designed with a minimum SAR value which can avoid any reaction or infection of antenna material with human body organs. For all the BMDs there are two available band

- Medical implantable communication System (402.5 MHz to 405.6 MHz)
- ISM Bands (900 MHz,2.45GHz)

So, all implantable antennas must be designed in these given bands. The simulation is carried out in a High-frequency structure simulator (HFSS). In simulation skin model is assigned which has similar specifications to the real skin of a human. By taking the suitable value of dielectric constant and bulk conductivity we can develop a box that resembles the actual skin of a human. The property of the skin is highly dependent upon the operating frequency of the antenna as well [3].

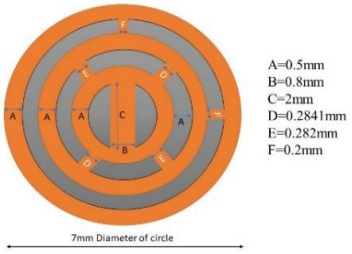
There are various techniques to design antennas with good gain, compact size, acceptable SAR, and available bands. PIFA antennas are highly recommended for implantable medical devices. In a PIFA antenna, a sorting pin is inserted between radiating surface and the ground of the antenna. Due to this, the current path is highly increased and increases the equivalent inductance which causes to shift higher resonating frequency to a lower value. This is how miniaturization is achieved using a PIFA antenna [4]. But there is some drawback in PIFA antenna that inserting a sorting pin in the substrate is a very challenging task. This makes the fabrication of said antenna very difficult and leads to having a very complex fabrication. In this research work, the proposed antenna is designed with the absence of a sorting pin that makes the fabrication of the antenna so easy. The performance of implantable antennas is also improved by introducing slots on the ground plane. Slots on the ground cause change in current distribution which influence the impedance of the antenna. When the impedance of the antenna match with feed impedance then maximum power is transferred and get enhanced parameters of the antenna. Likewise recommended research work also introduces the slots on the ground plane and gets circular polarisation.

II. ANTENNA DESIGN

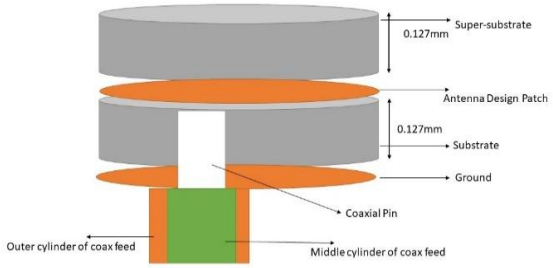
The proposed implantable antenna configuration resides of three rings on the radiating surface which are having 0.5 mm width individually. The outermost radius of the radiating surface is 3.5 mm. Each ring on the antenna patch is separated by a 0.5 mm distance. The smallest radius ring is

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 01,2021 at 05:28:17 UTC from IEEE Xplore. Restrictions apply.

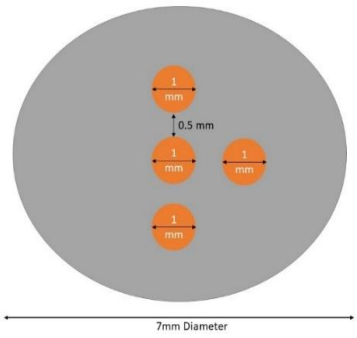
sorted with the middle ring by 4 strips with different dimensions given in figure 1. The smallest radius Ring is also sorted with a strip of 2mm x 0.8mm dimensions in the centre. The outermost ring and middle ring are also sorted by two rectangles with 0.2 mm wide and 0.5 mm long dimensions. These three rings structure sorted to each other of the radiating surface, is situated on Rogers RT/Duroid substrate with a standard thickness of 0.127 mm and radius of 3.5 mm. The given substrate has a high value of dielectric constant 10.2, relative permeability 1, and loss tangent of 0.0023. The high value of the dielectric constant leads to miniaturize the size of the antenna, reduces the effective wavelength, and improves the radiation properties of the antenna and low value of loss tangent shows low ohmic losses in the antenna. Same material Rogers RT/Duroid is used as super substrate with the same thickness which is used to avoid any reaction with human body since Antenna is a conducting material, decouples absorption from the antenna and also stabilize the effective permittivity of the substrate. Both substrate and super substrate materials are selected to have minimum reflection and biocompatibility. Total size of proposed antenna is 3.14*(3.5)2*(0.254) mm3 [9.77 mm3]. The proposed antenna is one of the smallest sizes among recent research papers. Recommended Antenna is used coaxial feed for 50-ohm impedance matching. To design the coax feed with 50 ohms, the specific diameter of the pin, middle and outer cylinder of coax feed is taken into consideration with specific dimensions (0.28mm, 0.4mm, 0.935mm) respectively.



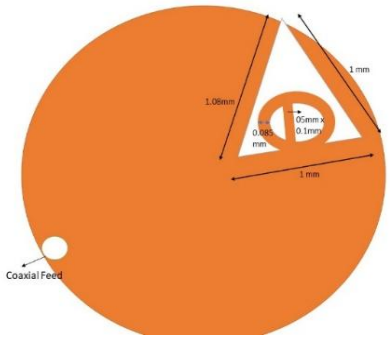
(a)



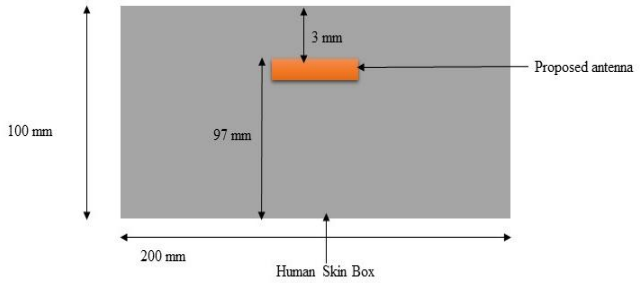
(b)



(c)



(d)



(e)

Fig 1. Proposed antenna geometry (a) Top view of patch surface (b) Side view of the proposed antenna. (c) Metamaterial structure on a super substrate of the proposed antenna. (d) Bottom view of proposed antenna. (e) proposed antenna placement inside skin box with 3 mm skin depth.

The proposed antenna is simulated on ANSYS high-frequency structure simulator. A human skin modal is assigned by taking the suitable value of dielectric constant and bulk conductivity in the ISM band as given in table 1. The proposed antenna is placed inside of skin box with a 3 mm depth of penetration.

TABLE I. Electrical Properties of Skin in ISM Band [5]

Tissue type	**Skin**
Dielectric Constant e_r	42.8
Bulk Conductivity	1.593 (S/m)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 01,2021 at 05:28:17 UTC from IEEE Xplore. Restrictions apply.

III. Results & Discussion

A. Return Loss

To avoid any antenna from reflection loss, we must design an antenna having a return loss of less than -10 dB. With this criterion, more than 90% of the incident power is transmitted through the antenna and reduces impedance mismatching.

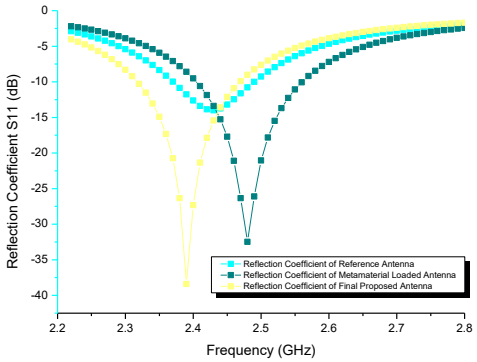


Fig 2. Return Loss Characteristics of Proposed Antennas.

Reference antenna with three rings is optimized by taking the specific position of coax feed at point 0, -3.2, 0. The reference antenna exhibits a resonating frequency of 2.43 GHz and having impedance bandwidth from 2.37 GHz to 2.49 GHz. On the super-substrate of reference triple ring implantable antenna, four circles of radius 0.5 mm are designed with a gap of 0.5 mm. The specific radius and position of the four circles are optimized on the super-substrate which is shown in figure 1(c). Due to these four circles on super-substrate, the resonating frequency is shifted from 2.43 GHz to 2.48 GHz which remains in the ISM band. The S11 parameter is enhanced from a minimum value of -14 dB to -32.5 dB as shown in figure 2. After these four rings on the super substrate, a small triangle slot is introduced as shown in fig 1(d) on the ground surface with the ring inside it. Due to this slot, the results of the antenna have been enhanced up to a large extent. The resonating frequency is shifted from 2.48 GHz to 2.39 GHz which results in the miniaturization of the antenna. S11 parameter is highly enhanced from -32.5 dB to -38.4 dB.

B. Axial Ratio.

The reference antenna exhibits linear polarisation with an axial ratio of more than 3 Db. But with the introduction of four circles on super-substrate and slot on the ground plane, the axial ratio reduced from a higher value to a lower value. But the axial ratio of metamaterial loaded antenna is still above 3Db in the frequency range from 2.2 GHz to 2.8 GHz. The proposed antenna with ground slots exhibits an axial ration less than 3 dB which makes this proposed antenna a circularly polarised antenna. This axial ratio of less than 3 dB results in the antenna into a circular polarised antenna which has a great advantage. The proposed implantable antenna exhibits circular polarisation from 2.38 GHz to 2.67 GHz frequency band which is the ISM band. During telemetry sessions, all patients can enjoy full mobility just because of the circular polarisation of implantable antenna. The graph of axial ratio for all 3 proposed antennas is shown below in figure 3.

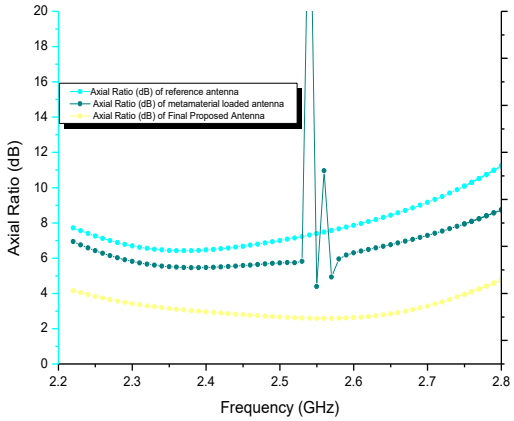
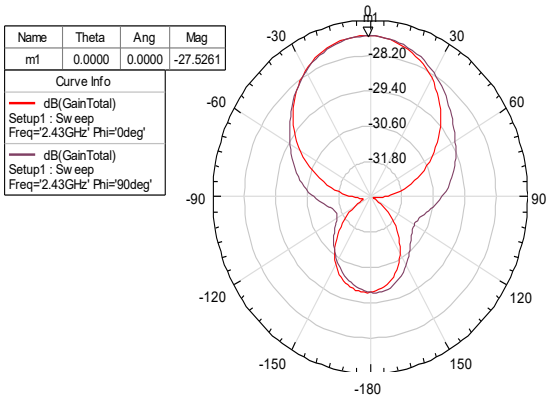


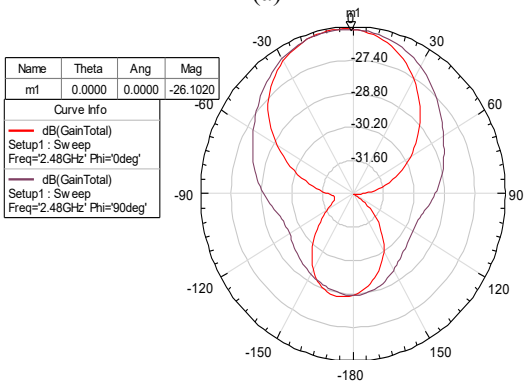
Fig 3. Axial Ratio characteristics of Proposed antennas.

C. Radiation Pattern

Directive gain of the proposed antenna shows minute enhancement over reference antenna design. Since an antenna is implanted inside the human body so antenna does face a lot of interference. Due to the extremely high lossy medium, implantable antenna gain usually has a low value [9]. The radiation pattern of the proposed antenna is presented below in figure 4. Reference antenna exhibits a gain of -27.5 dB at resonating frequency 2.43 GHz. After loading of metamaterial on the super substrate, the gain is enhanced a little value from -27.5 dB to -26.1 dB at resonating frequency 2.48 GHz. The final proposed antenna with ground slots exhibits a gain of -27.04 dB with the improvement of 0.5 dB at resonating frequency 2.39GHz.

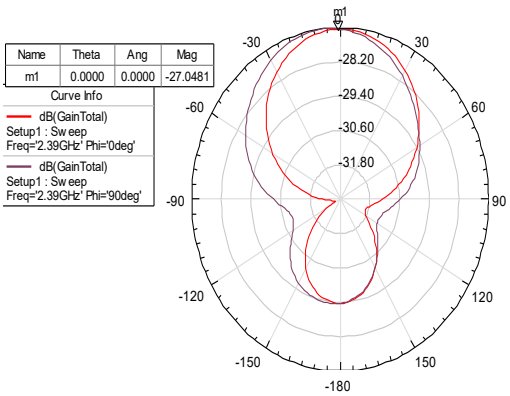


(a)



(b)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 01,2021 at 05:28:17 UTC from IEEE Xplore. Restrictions apply.



(c)

Fig 4. Radiation Pattern characteristics (a) Radiation Pattern of Reference Antenna (b) Radiation Pattern of Metamaterial loaded antenna (c) Radiation Pattern of the final proposed antenna.

D. Specific Absorption Rate

As per IEEE standard safety guidelines, the human tissues should not absorb1.6 w/kg or above for 1 gram of human body tissue and 2 w/kg or above for 10 grams of human body tissue. Designing the implantable antenna within the standard value of SAR is our major concern otherwise antenna will show a prejudicial outcome on the human body. The absorbed power by the human body and electric field is explained below.

$$SAR=\int_{sample} \frac{\sigma(r)\,|\vec{E}(r)|^2}{\rho(r)}\,dr \qquad (1)$$

$$\vec{E}(r)=\frac{1}{jw\epsilon\mu}[\kappa^2\vec{A}+\nabla(\nabla\cdot\vec{A})] \qquad (2)$$

Where $\sigma(r)$ a current density of antenna

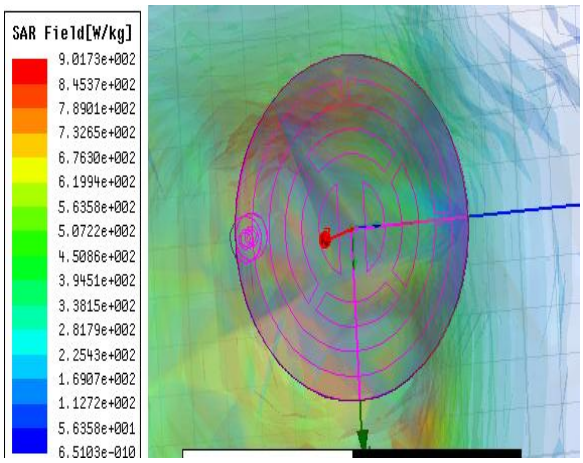
$\vec{A}$ is the magnetic vector potential.

$\vec{E}(r)$ is the electric field with the function of distance r from an antenna.

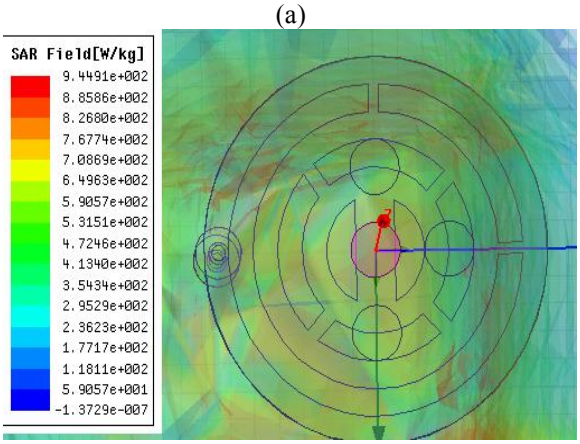
$\kappa^2\vec{A}$ is an electric field in the far-field region.

$\nabla(\nabla\cdot\vec{A})$ is an electric field in the near field region.

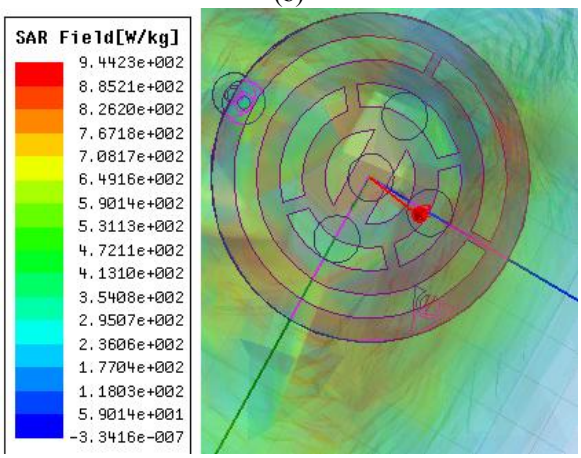
SAR will be increased if an electric field is high, but this value can be reduced by having uniform current distribution along the antenna surfaces and electric field in the near region from antenna dominates and absorbs power. This parameter is the most important reflection for patient safety. All proposed antennas are having an acceptable range of SAR which is shown below in figure 5.



(a)



(b)



(c)

Fig 5. (a) SAR of Reference Antenna at 2.43 GHz (b) SAR of Metamaterial Loaded Antenna at Frequency 2.48 GHz (c) SAR of Final Proposed Antenna at Frequency 2.39 GHz.

Table II. SAR and input power of all three proposed antennas.

Antenna	**Max SAR for 1 gram**	**Max input power (mW) C95.1-1999(1g-avg)**
Reference antenna	901.7	1.77
Metamaterial Loaded Antenna	944.9	1.69
Final proposed antenna at 2.43GHz	944.2	1.69

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 01,2021 at 05:28:17 UTC from IEEE Xplore. Restrictions apply.

E. *Performance Comparison of Proposed Antenna with Recently published work*

Table III. Performance Comparison of Proposed Antenna with Recently published work

Ref Paper	Volume (mm^3)	Operating frequency GHz	(%) Bandwidth	Gain (dBi)	SAR for 1 gram	CP
6	816	2.45	6.2	-9	….	No
7	20.06	0.402	35	-23.7	219	No
8	153.67	0.915	3.8	-29	…..	Yes
Final Proposed Antenna	**9.77**	**2.39**	**6.46**	**-27**	**944**	**Yes**

From this comparison table, it is demonstrated that the proposed antenna shows good performance as compared to recent research work. We have designed an antenna that has a very compact size, low return loss, accepted value of SAR, and circularly polarised in the ISM band. No use of a sorting pin makes this proposed work very effective, effortless, and facile for fabrication.

IV. CONCLUSION

All implantable antennas are very difficult to design. The size and fabrication of these antennas is a great concern. In this research, we have designed such an antenna which has the smallest size as compare with recent research work related to the implantable antenna. The volume of the proposed antenna is having only 9.77 mm3. Most of the implantable device's researchers are designing PIFA antenna in which there is a sorting pin between the ground and radiating surface. Due to this reason, fabrication becomes a very challenging task. But in our research work proposed antenna is having a very simple structure without any sorting pin. The proposed antenna shows very good results in the ISM band with resonating frequency 2.39 GHz and -38.6 dB low value of return loss. Moreover, this implantable antenna is circularly polarised in each range of frequencies from 2.38 GHz to 2.67 GHz with less than a 3 dB axial ratio. SAR (944.4 W/KG) of the proposed antenna is still acceptable according to IEEE standards. Any implantable antenna shows degradation in its performance just because of the surrounding human environment. But our proposed antenna demonstrates the best result regarding return loss, Axial ratio, SAR, and radiation pattern.

ACKNOWLEDGMENT

For research support, I acknowledge Prof. N S Raghava, Head of Department, Electronics and Communication Engineering, Delhi Technological University, New Delhi for his inevitable support and wisdom for this research work. His immense support and great knowledge in this field, have helped me in acquiring a greater sense of understanding towards this work. His selfless contribution has eventually helped in widening my mind to enhance my learning by facilitating it in the desired field.

REFERENCES

[1] C. L. Tsai, K. W. Chen, and C. L. Yang, "Implantable wideband low-specific-absorption rate antenna on a thin flexible substrate," IEEE Antennas and Wireless Propagation Letters, vol. 15, pp. 1048-1052, 2016.

[2] Yilmaz T, Karacolak T, Topsakal E. "Characterization and testing of skin mimicking material for implantable antennas operating at ISM band (2.4 GHz-2.48 GHz)" IEEE Antennas and Wireless Propagation Letters. 2008, 7: 418-420. DOI:10.1109/LAWP.2008.2001736.

[3] Z-J. Yang, and S. Xiao, "A wideband implantable antenna for 2.4 GHz ISM band biomedical application," In Proc. International Workshop on Antenna Technology (iWAT), China, 2018, pp. 1-3.

[4] F. Gozasht, Md. Delwar Hossain, Ananda Sanagavarapu Mohan "Miniaturized E-Shaped PIFA Antenna for Wideband Implantable Biomedical Applications" 2013 International Conference on Electromagnetics in Advanced Applications (ICEAA) https://doi.org/10.1109/ICEAA.2013.6632355.

[5] Dielectric Properties of Body Tissues (IFAC), Available: http://niremf.ifac.cnr.it/tissprop/.

[6] S. Alamri, A. AlAmoudi, and R. Langley, "Gain enhancement of implanted antenna using lens and parasitic ring," Electron. Lett., vol.52, no. 10, pp. 800–801, May 2016.

[7] R. Lesnik, N. Verhovski, I. Mizrachi, B. Milgrom, and M. Haridim, "Gain enhancement of a compact implantable dipole for biomedical applications," IEEE Antennas Wirel. Propag. Lett., vol. 17, no. 10, pp.1778–1782, Oct. 2018.

[8] C. Liu, Y. Zhang, and X. Liu, "Circularly polarized implantable antenna for 915 MHz ISM-Band far-field wireless power transmission," IEEE Antennas and Wireless Propagation Letters, vol. 17, no. 3, pp. 373–376, March 2018.

[9] S. Das and D. Mitra, "A compact wideband flexible implantable slot antenna design with enhanced gain," IEEE Trans. Antennas Propag., vol. 66, no. 8, pp. 4309–4314, Aug. 2018.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 01,2021 at 05:28:17 UTC from IEEE Xplore. Restrictions apply.

Physics of Plasmas

Influence of strongly coupled plasma environment on photoionization of H-like O^{7+} ion

Cite as: Phys. Plasmas **28**, 112706 (2021); https://doi.org/10.1063/5.0055265
Submitted: 27 April 2021 • Accepted: 24 October 2021 • Published Online: 22 November 2021

Dishu Dawra, Mayank Dimri, A. K. Singh, et al.

View Online Export Citation CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Effects of electron cyclotron heating on the toroidal flow in LHD plasmas
Physics of Plasmas **28**, 102501 (2021); https://doi.org/10.1063/5.0057344

Scattering of ultrashort laser pulses in Maxwellian plasmas: Transition scattering
Physics of Plasmas **28**, 102104 (2021); https://doi.org/10.1063/5.0065862

Static properties of a warm dense uniform electron gas
Physics of Plasmas **28**, 112704 (2021); https://doi.org/10.1063/5.0062259



AIP Publishing

© 2021 Author(s).

Physics of Plasmas ARTICLE scitation.org/journal/php

Influence of strongly coupled plasma environment on photoionization of H-like O^{7+} ion

Cite as: Phys. Plasmas **28**, 112706 (2021); doi: 10.1063/5.0055265
Submitted: 27 April 2021 · Accepted: 24 October 2021 ·
Published Online: 22 November 2021

View Online Export Citation CrossMark

Dishu Dawra,[1] Mayank Dimri,[1,2,3] A. K. Singh,[1,2,a)] Alok K. S. Jha,[4] Rakesh Kumar Pandey,[3] Rinku Sharma,[5] and Man Mohan[1,2]

AFFILIATIONS

[1]Department of Physics & Astrophysics, University of Delhi, Delhi 110007, India
[2]Department of Physics, Deen Dayal Upadhyaya College, University of Delhi, Delhi 110078, India
[3]Department of Physics, Kirori Mal College, University of Delhi, Delhi 110007, India
[4]School of Physical Sciences, Jawaharlal Nehru University, Delhi 110067, India
[5]Department of Applied Physics, Delhi Technological University, Delhi 110042, India

[a)]Author to whom correspondence should be addressed: avni.physics@gmail.com

ABSTRACT

The effect of strongly coupled plasma environment on the photoionization cross sections of the H-like O^{7+} ion is studied within the framework of relativistic configuration interaction technique implemented in the flexible atomic code. The analytical b-potential of Li and Rosmej [Phys. Lett. A **384**, 126478 (2020)] and uniform electron gas model potential of Saha and Fritzsche [J. Phys. B **40**, 259 (2007)] have been employed for incorporating the plasma shielding effects. The level energies for unscreened H-like O^{7+} ion are in excellent agreement with the National Institute of Standards and Technology values. The bound and continuum state wavefunctions are determined by solving the modified Dirac equations. The photoionization cross sections evaluated in the dipole approximation are obtained from the bound-free transition matrix element. Cross section calculations are performed over an electron density range of 1.0×10^{19}–2.0×10^{23} cm^{-3} and a temperature range of 200–1000 eV. The plasma environment is found to considerably influence the photoionization cross sections. It has been observed that the ionization thresholds move to the low energy region with the increase in free electron densities. The present results will be beneficial for the interpretation of spectra observed in a variety of astrophysical and laboratory plasmas.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0055265

I. INTRODUCTION

The influence of dense plasma environment on the atomic properties has attracted considerable attention during the past few years.[1–5] This is primarily due to its practical implications in many areas of physics, such as the inertial confinement fusion research, x-ray lasers, laser produced plasmas, planetary science, and plasma spectroscopy.[6–12] In such an environment, the interaction between the bound electrons with atomic nucleus is screened by the neighboring ions and free electrons. The photoionization process is one of the fascinating fields of research in plasma physics due to its extreme sensitivity to the detailed electronic structure of atoms/ions. The photoionization cross sections are needed in the modeling and diagnostics of laboratory and astrophysical plasmas.[13] The presence of dense plasma environment alters the electronic wavefunctions and atomic structure properties of atoms/ions and leads to lowering of the ionization potential when compared with free ions. Moreover, with the recent advances in synchrotron light sources,[14–17] it has become possible to study matter under extreme conditions of high temperature and high density, and thus it is feasible to create plasma form of matter.[18,19]

Depending upon the plasma conditions, several model potentials are available for studying the influence of plasma screening induced by the surrounding ions and fast electrons. The more commonly known model potentials are Debye–Hückel potential and ion sphere potential.[20,21] The choice of specific potential depends upon the strength of the plasma coupling parameter (Γ) which is defined as the ratio of average Coulomb and kinetic energy. In the case of weakly coupled plasma ($\Gamma < 1$), the Debye–Hückel potential is the extensively utilized potential for investigating the atomic structure properties of ions. However, for strongly coupled plasma ($\Gamma \geq 1$), the plasma shielding effects can be appropriately depicted by the ion sphere potential. Strongly coupled plasmas are mostly encountered in laser produced plasmas, inertial confinement fusion, and astrophysical systems, such as Jovian planets and interior of stars.[22] In short pulse plasmas, such as the x-ray free-electron lasers (XFELs)[15–17] and higher-harmonic

Published under an exclusive license by AIP Publishing

generation (HHG),[14] a certain lattice structure of the solid material remains even when the temperature is rather high that seriously influences the ionization potential depression. As a result, the ionization potential depression may be different for the same coupling parameters but different irradiation conditions. Therefore, the situation is better described by a physical landscape of the recently developed atomic-solid-plasma model[23] rather than by a coupling parameter only.

Several studies have been performed for investigating the effect of plasma screening on the photoionization of one-electron and many-electron systems. Sahoo and Ho[24] calculated the photoionization cross sections for Li and Na alkali metal atoms in plasma environment based on the Debye–Hückel potential. Qi *et al.*[25] have investigated the photoionization cross sections for the ground and excited states ($n \leq 4$) of Li atom immersed in weakly coupled plasma environment. Also radiative recombination rate coefficients have been calculated for Li^+ ion in a broad range of plasma parameters. Furthermore, the theoretical photoionization cross sections for the 1s2s (^{1}S) and 1s2p (^{1}P) states of He atom have been presented by Sahoo and Ho[26] in a weakly coupled plasma environment using the complex-coordinate rotation method. Lin and Ho[27,28] have explored the quadrupole photoionization of hydrogen atoms by the complex-coordinate rotation method using the Debye–Hückel and modified Debye–Hückel potentials. Furthermore, the influence of screened Coulomb and exponential-cosine-screened Coulomb potentials on the photoionization of H and He^+ has been studied by Lin and Ho[29] within the framework of the complex-coordinate rotation method.

A detailed theoretical study on the photoionization of one- and two-electron atoms has been presented by Chang and Fang[30] in an external plasma environment based on the Debye–Hückel model. Das[31] studied the effect of strongly coupled plasma on the ground state photoionization cross section of H- and Li-like ions by making use of the ion sphere model potential. Xie *et al.*[32] have estimated the relativistic effects in the photoionization cross sections of H-like ions for a wide range of screening lengths and photoelectron energies. Qi and Ning[33] investigated the dynamic processes and dipole polarizibilities of sodium atom immersed in Debye plasmas. The occurrence of zero-energy resonances in atomic processes has been investigated by Clauser and Barrachina[34] within the Debye plasma environment. Later on, Qi *et al.*[35] presented the photoionization of H-like ions in dense quantum plasmas using the cosine-Debye–Hückel potential. The photoionization for He^+ ion moving in fusion plasma has been investigated by Hu[36] using the Ritz variational method. Zhao and coworkers[37] have studied the photoionization of H-like C^{5+} ion in warm dense plasmas using the static screened potential. Recently, Chen *et al.*[38,39] presented a detail investigation of the photoionization of H-like Al^{12+} and C^{5+} ions in strongly coupled plasma environment by utilizing the Stanton and Murillo potential and ion sphere model potential, respectively.

In accordance with the above discussion, a large number of efforts have been devoted in the past for the photoionization cross section calculations of highly charged ions, but most of the photoionization calculations are confined to the implementation of Debye–Hückel potential and are not relevant to systems having strong gradients of density. Therefore, the goal of the present study is to investigate the photoionization processes of the H-like O^{7+} ion embedded in a strongly coupled plasma environment. Concerning our photoionization cross section calculations, bound and continuum state wavefunctions are determined by solving the Dirac equations using the analytical plasma screening b-potential of Li and Rosmej[40] and the uniform electron gas model (UEGM) potential[7] to account for the plasma shielding effects within the relativistic configuration interaction (RCI) method. The calculations for excitation energies, ionization potentials, radiative parameters, photoionization, and radiative recombination cross sections of O^{7+} ion are carried out in the density and temperature regimes of $n_e = 1.0 \times 10^{19}$–$2.0 \times 10^{23}\,cm^{-3}$ and $kT = 200$–$1000\,eV$, respectively. The photoionization cross sections evaluated in the dipole approximation are obtained from the bound-free transition matrix element. It is worth mentioning that the lowering of energies and ionization potential results in the variation of photoionization and radiative recombination cross sections, which has been emphasized for the first time for the H-like O^{7+} ion. We firmly believe that our presently reported results will be beneficial in the analysis of spectra observed in a variety of astrophysical and laboratory plasmas as such high temperature and density conditions of plasma to highly charged oxygen ions are of considerable interest for stellar interior regions.[41,42]

This manuscript is structured as follows: In Sec. II, we have briefly discussed the underlying theory utilized for the calculation of photoionization cross sections. Results of our findings are provided and addressed in Sec. III. Ultimately, in Sec. IV, we end with a brief summary and perspective for the future work.

II. THEORETICAL METHODS

A brief overview of the adopted theoretical methods for studying the effect of plasma shielding on the H-like O^{7+} ion immersed in strongly coupled plasma has been presented by making use of the ion sphere model. In this model, one can make the assumption that the ion is depicted by a point-like nucleus having charge Z immersed at the center of a spherical cavity, including sufficient electrons to ascertain electrical neutrality of the ion sphere. This spherical cavity is known as an ion sphere or Wigner–Seitz cell.

A. The analytical b-potential

The influence of plasma shielding on the properties of atomic structure has been modeled using the analytical b-potential of Li and Rosmej,[40] given as

$$V_{IS} = -\frac{Z}{r} + \frac{N_f}{R_0}\left[1 + \frac{1}{x-1} - \frac{1}{x-1}\left(\frac{r}{R_0}\right)^{x-1}\right], \tag{1}$$

with

$$R_0 = \left[\frac{3N_f}{4\pi n_e}\right]^{\frac{1}{3}}; \quad N_f = Z - N_b, \tag{2}$$

$$x = 3 - \frac{b}{\pi}\sqrt{\frac{N_f}{R_0\,T_e}}, \tag{3}$$

where r, R_0, n_e, N_b, and N_f represent the radial coordinate of the electron, radius of the ion sphere, electron density, number of bound electrons, and the number of free electrons inside the ion sphere, respectively. T_e denotes the temperature of electron, and b is a parameter that characterizes the self-consistent electron distribution in the ion sphere, e.g., for a Maxwellian electron distribution, the multiconfiguration Dirac–Fock self-consistent finite temperature ion sphere model multiconfiguration Dirac-Fock self-consistent finite temperature ion

Published under an exclusive license by AIP Publishing

sphere model (MCDF-SCFTIS) simulations are matched with b ≈2. When x = 3 (T_e is infinite), Eq. (1) is equivalent to UEGM potential,[7] which is written as

$$V_{IS}(r, R_0) = -\frac{Z}{r} + \frac{N_f}{2R_0}\left[3 - \left(\frac{r}{R_0}\right)^2\right]. \quad (4)$$

B. The RCI method incorporating the plasma shielding effect

RCI employs a modified Dirac–Fock–Slater central-field potential[43] that involves an approximate treatment of the exchange interaction. A self-consistent-field (SCF) iterative procedure is carried out in which the orbitals are optimized and the average energy of a fictitious mean configuration is minimized. This method has also been successfully implemented in our earlier works.[44–48] Relativistic effects are considered using the Dirac Coulomb Hamiltonian. Higher order QED effects have been incorporated with the Breit interaction in the zero energy limit for the exchange photon along with hydrogenic approximations for the self-energy and vacuum polarization effects. In this method, the relativistic Hamiltonian for the *N*-electron atom/ion embedded in strongly coupled plasma environment has the general form

$$\hat{H} = \sum_{i=1}^{N} \hat{H}_D(i) + \sum_{i<j}^{N} \frac{1}{r_{ij}}, \quad (5)$$

where the single-electron Dirac–Hamiltonian $\hat{H}_D(i)$ is defined as

$$\hat{H}_D(i) = c\,\boldsymbol{\alpha}.\boldsymbol{p} + (\beta - 1)c^2 + V_{IS}(r, R_0). \quad (6)$$

Here, the first two terms signify the relativistic kinetic energy of a bound electron and the final term $V_{IS}(r, R_0)$ represents the modified potential experienced by the bound electron within the ion sphere.

C. The continuum wavefunction incorporating the plasma shielding effect

The theoretical model used for the continuum wavefunctions calculation is based upon the distorted wave approximation. The large and small components, $P_{nk}(r)$ and $Q_{nk}(r)$, of the continuum wavefunction in the central potential approximation[43] are determined by the subsequent differential equations,

$$\left(\frac{d}{dr} + \frac{k}{r}\right) P_{nk}(r) = \alpha\left(\varepsilon_{nk} - V^N(r) - V^{ee}(r) - V_{IS} + \frac{2}{\alpha^2}\right) Q_{nk}(r),$$
$$\left(\frac{d}{dr} - \frac{k}{r}\right) Q_{nk}(r) = \alpha\left(-\varepsilon_{nk} + V^N(r) + V^{ee}(r) + V_{IS}\right) P_{nk}(r), \quad (7)$$

where α denotes the fine structure constant and ε_{nk} depicts the energy eigenvalues of the radial orbitals. $V^N(r)$ and $V^{ee}(r)$ represent the attractive nuclear potential and the repulsive electron–electron interaction potential, respectively. In this method, the normalization condition satisfied by the radial components of the wavefunction is given by

$$\int_0^\infty \left[P_{nk}^2(r) + Q_{nk}^2(r)\right] dr = 1. \quad (8)$$

The bound electron wavefunctions might not be equal to zero outside the ion sphere for higher values of free electron density.

D. Photoionization and radiative recombination cross section calculations

The photoionization cross sections are calculated in the dipole approximation by making use of the above wavefunctions.[43] The relativistic photoionization cross section is given as

$$\sigma_{PI} = 2\pi\alpha \frac{1 + \alpha^2\varepsilon}{1 + 0.5\alpha^2\varepsilon} \frac{df}{dE}, \quad (9)$$

and the radiative recombination cross section is given by the Milne relation[43]

$$\sigma_{RR} = \frac{\alpha^2}{2} \frac{g_i}{g_f} \frac{\omega^2}{\varepsilon(1 + 0.5\alpha^2\varepsilon)} \sigma_{PI}, \quad (10)$$

where α is the fine structure constant, ω is the photon energy, g_i and g_f are the statistical weights of the initial and final bound states, ε is the photoelectron energy, and $\frac{df}{dE}$ represents the oscillator strength given by

$$\frac{df}{dE} = \frac{\omega}{g_i} [L]^{-1} (\alpha\omega)^{2L-2} S, \quad (11)$$

where L denotes the rank of the multipole operator, $[L]$ stands for $2L + 1$, and the generalized line strength S is given by

$$S = \sum_{kJ_T} |\langle \Psi_f,\, k; J_T || O^L || \Psi_i \rangle|^2, \quad (12)$$

where k depicts the relativistic quantum number of free electron, J_T represents the total angular momentum of the free state, and O^L is the multiple operator inducing the transition.

III. RESULTS AND DISCUSSION

In the present study, a modification in the relativistic configuration interaction method has been done by including two ion sphere potentials, namely, analytical b-potential and UEGM potential, to investigate the influence of strongly coupled plasma environment. In our RCI calculations, an atomic model consisting of six configurations 1s, 2s, 2p, 3s, 3p, and 3d has been used owing to the fact that there is absence of electron correlation effects in H-like ions. In Table I, we show the level energies for H-like O^{7+} ion subjected to a dense plasma environment at different values of kT and n_e ranging from 200 to 1000 eV and 1.0×10^{19} to 2.0×10^{23} cm^{-3}, respectively. The results for the plasma free case have also been tabulated. Inspection of Table I shows that the reported results for unscreened H-like O^{7+} ion are in excellent agreement with the values compiled from the National Institute of Standards and Technology (NIST).[49] Moreover, the results from the UEGM model have also been listed. One can infer that when the effect of dense plasma effects has been taken into consideration, the transition energies are decreasing with the increase in free electron density. For a given kT, the transition energies of O^{7+} ion change very little with free electron density n_e between 1.0×10^{19} and 1.0×10^{22} cm^{-3}, though change rapidly beyond 1.0×10^{22} cm^{-3}.

To gain more insight, Fig. 1 displays the variation of our plasma energy shifts for the $2s_{1/2}$, $2p_{3/2}$, $3s_{1/2}$, and $3d_{5/2}$ levels of O^{7+} ion at different values of kT and n_e. It is interesting to observe that the shifts of the transition energies increase with the increase in plasma shielding strengths for a fixed value of kT. For example, the plasma energy shifts

Published under an exclusive license by AIP Publishing

TABLE I. Excitation energies (in eV) of H-like O^{7+} ion at various free electron densities n_e (cm^{-3}) and temperatures kT (eV) from the analytical b-potential.[40] The energies from the uniform electron gas model (UEGM)[7] have also been given.

			n_e						
kT	States	NIST[49]	0	1.0×10^{19}	1.0×10^{20}	1.0×10^{21}	1.0×10^{22}	1.0×10^{23}	2.0×10^{23}
200 eV	$1s_{1/2}$	0.0000	0.0000	0.00	0.00	0.00	0.00	0.00	0.00
	$2p_{1/2}$	653.4937	653.4926	653.49	653.49	653.49	653.43	652.84	652.17
	$2s_{1/2}$	653.5028	653.5019	653.50	653.50	653.49	653.41	652.57	651.64
	$2p_{3/2}$	653.6799	653.6789	653.68	653.68	653.67	653.61	653.01	652.33
	$3p_{1/2}$	774.5788	774.5778	774.58	774.58	774.54	774.22	770.96	757.23
	$3s_{1/2}$	774.5816	774.5806	774.58	774.58	774.54	774.17	770.47	756.04
	$3d_{3/2}$	774.6340	774.6330	774.63	774.63	774.61	774.38	772.01	759.51
	$3p_{3/2}$	774.6340	774.6330	774.63	774.63	774.60	774.28	771.01	757.23
	$3d_{5/2}$	774.6523	774.6513	774.65	774.65	774.63	774.40	772.02	759.51
600 eV	$1s_{1/2}$			0.00	0.00	0.00	0.00	0.00	0.00
	$2p_{1/2}$			653.49	653.49	653.49	653.45	652.99	652.48
	$2s_{1/2}$			653.50	653.50	653.50	653.43	652.78	652.05
	$2p_{3/2}$			653.68	653.68	653.67	653.63	653.17	652.65
	$3p_{1/2}$			774.58	774.58	774.55	774.28	771.55	759.31
	$3s_{1/2}$			774.58	774.58	774.55	774.24	771.12	758.13
	$3d_{3/2}$			774.63	774.63	774.61	774.42	772.48	761.60
	$3p_{3/2}$			774.63	774.63	774.60	774.34	771.60	759.31
	$3d_{5/2}$			774.65	774.65	774.63	774.44	772.49	761.60
1000 eV	$1s_{1/2}$			0.00	0.00	0.00	0.00	0.00	0.00
	$2p_{1/2}$			653.49	653.49	653.49	653.45	653.03	652.56
	$2s_{1/2}$			653.50	653.50	653.50	653.44	652.83	652.16
	$2p_{3/2}$			653.68	653.68	653.67	653.63	653.21	652.73
	$3p_{1/2}$			774.58	774.58	774.55	774.30	771.71	759.90
	$3s_{1/2}$			774.58	774.58	774.55	774.26	771.29	758.71
	$3d_{3/2}$			774.63	774.63	774.61	774.44	772.60	762.19
	$3p_{3/2}$			774.63	774.63	774.61	774.35	771.76	759.90
	$3d_{5/2}$			774.65	774.65	774.63	774.45	772.62	762.19
UEGM	$1s_{1/2}$			0.00	0.00	0.00	0.00	0.00	0.00
	$2p_{1/2}$			653.49	653.49	653.49	653.46	653.15	652.79
	$2s_{1/2}$			653.50	653.50	653.50	653.45	652.99	652.47
	$2p_{3/2}$			653.68	653.68	653.67	653.64	653.32	652.96
	$3p_{1/2}$			774.58	774.58	774.56	774.35	772.19	761.73
	$3s_{1/2}$			774.58	774.58	774.55	774.31	771.82	760.54
	$3d_{3/2}$			774.63	774.63	774.62	774.47	772.97	764.03
	$3p_{3/2}$			774.63	774.63	774.61	774.40	772.23	761.73
	$3d_{5/2}$			774.65	774.65	774.64	774.49	772.99	764.03

for $3d_{5/2}$ level are about 0.26 and 15.15 eV for free electron densities 1.0×10^{22} and $2.0 \times 10^{23}\,cm^{-3}$ at kT = 200 eV, respectively. Furthermore, the plasma energy shifts decrease with the increase in kT values. The cause for this behavior is that with the increase in kT, increasingly more electrons distribute far away from the nucleus, thus decreasing the plasma screening effect between the nucleus and electron. For a given $n_e = 2.0 \times 10^{23}\,cm^{-3}$, the shifts of the $3d_{5/2}$ level are about 15.15, 13.05, and 12.46 eV at kT = 200, 600, and 1000 eV, respectively. It can also be seen from Fig. 1 that the plasma energy shifts increase with an increase in electron densities from the UEGM potential. A close examination of this figure shows that the plasma energy shift is less prominent for lower lying $2l$ states as these states are strongly bound by the attractive nuclear potential and stay relatively less influenced at lower electron densities.

Figure 2 illustrates the variation of the fine-structure levels in the shell of H-like O^{7+} of n = 3. It can be inferred from the figure that energy levels within the same shell depend strongly on the orbital angular momentum l. With the increase in electron density n_e, energy

Published under an exclusive license by AIP Publishing

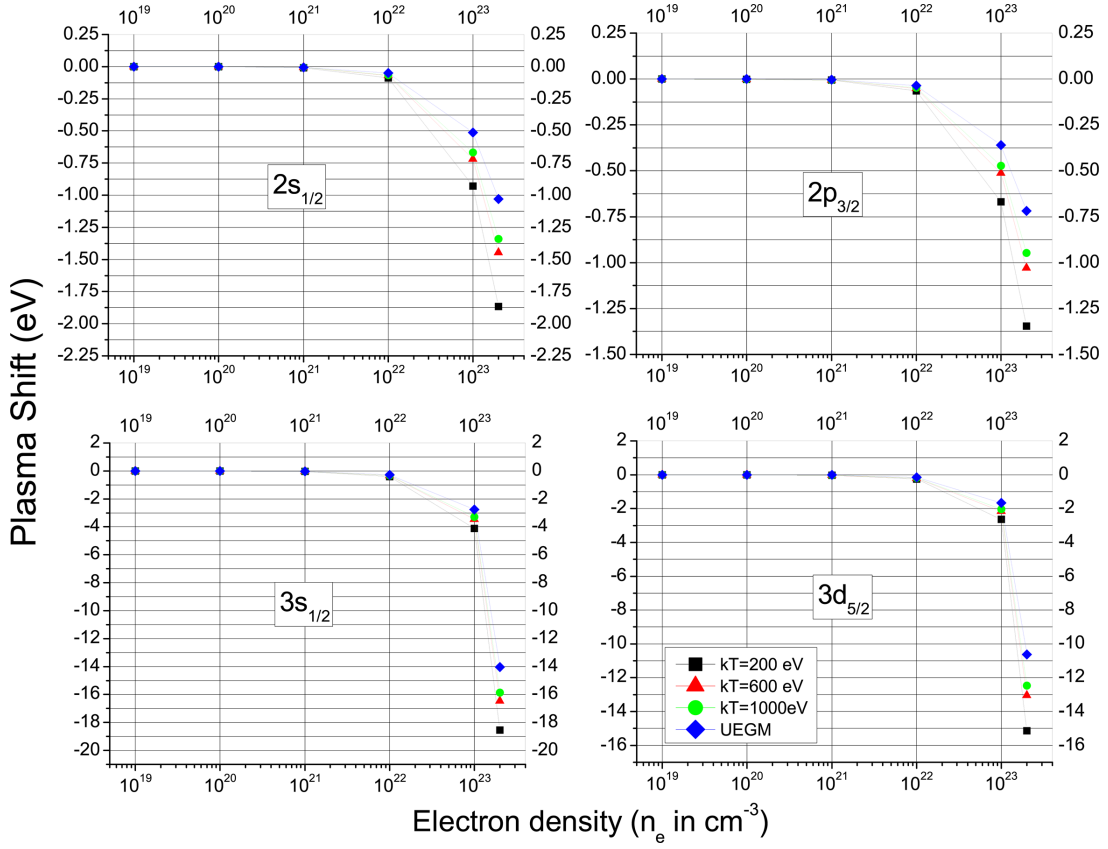


FIG. 1. Plasma energy shift for H-like O^{7+} ion immersed in a dense plasma from the RCI calculations using the analytical b-potential[40] and UEGM[7] potentials.

levels with larger l value shift toward the high-energy regime. Also the energy difference between orbits with different l becomes larger with an increase in electron density.

Figure 3 displays the influence of the plasma shielding on the excitation energies of $3p_{1/2}$ level for H-like C^{5+}, O^{7+}, and Al^{12+} ions at various electron densities. For a given electron temperature of kT = 200 eV, the excitation energies become more and more red-shifted with the increase in electron density. Moreover, the plasma shielding becomes less important with the increase in nuclear charges Z. For instance, the plasma energy shifts are about 35.45 eV and 3.65 eV for C^{5+} and Al^{12+} ions at free electron density $n_e = 2.0 \times 10^{23}$ cm^{-3}, respectively. In Table II, we compare the level energies of H-like C^{5+} ion immersed in a dense plasma environment with the results of Chen *et al.*[39] at electron densities 1.0×10^{20}, 1.0×10^{21}, and 1.0×10^{22} cm^{-3}, respectively. A good agreement is obtained between our values and those obtained by Chen *et al.*[39] The results for the unscreened case have also been provided, which are in excellent agreement with the NIST ones.

The variation in transition probabilities for $1s_{1/2} \rightarrow 2p_{3/2}$ transition of H-like O^{7+} ion at different values of kT and n_e is shown in Fig. 4. One can see that by considering plasma shielding effects, the transition probabilities show a decreasing trend with the increase in n_e

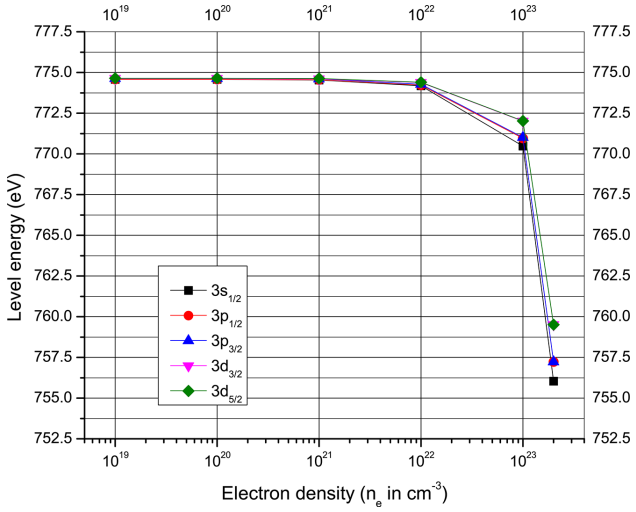


FIG. 2. Variation of the energy levels in the shell of H-like O^{7+} of n = 3 at kT = 200 eV.

Published under an exclusive license by AIP Publishing

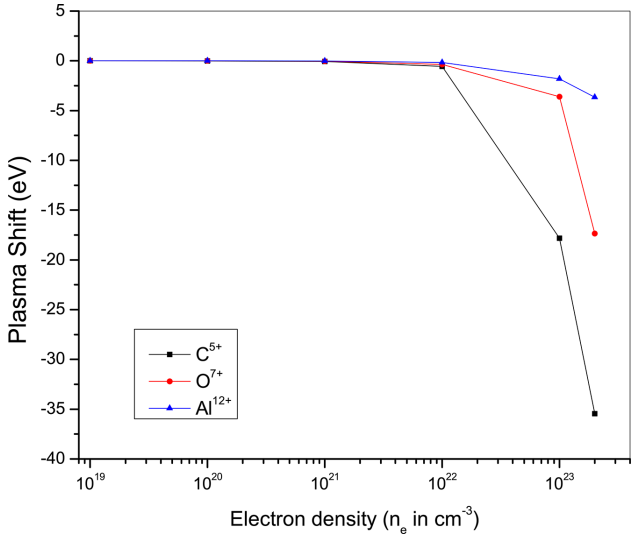


FIG. 3. Plasma energy shift of $3p_{1/2}$ level for H-like C^{5+}, O^{7+}, and Al^{12+} ions as a function of the free electron density at kT = 200 eV.

for a given kT, whereas they progressively decrease with a decrease in kT values for a given value of n_e. For example, the plasma screening effect on the transition probabilities of O^{7+} ion is about 0.07% (0.05%) and 1.44% (1.07%) for $n_e = 1.0 \times 10^{22}$ and $2.0 \times 10^{23}\,cm^{-3}$, with kT = 200 (1000 eV), respectively. Therefore, the changes are very small as the lower lying state $2p_{3/2}$ is strongly bound by the nuclear attraction and remain less affected at low densities. As expected, a similar behavior for transition probabilities is seen from the UEGM potential. Moreover, the transition probability of $1s_{1/2} \rightarrow 2p_{3/2}$ transition for the unscreened case is $25.6853 \times 10^{11}\ s^{-1}$, which shows good agreement with the NIST value of $25.6700 \times 10^{11}\ s^{-1}$.

TABLE II. Comparison of the level energies (in eV) for C^{5+} ion immersed in a plasma environment obtained from our UEGM[7] with the results of Chen *et al.*[39]

		n_e			
Levels	NIST[49]	0	1.0×10^{20}	1.0×10^{21}	1.0×10^{22}
$1s_{1/2}$	0.0000	0.0000	0.0000	0.0000	0.0000
$2p_{1/2}$	367.4740	367.4722	367.4720	367.4667	367.3846
		367.4741[a]	367.4734[a]	367.4677[a]	367.4108[a]
$2s_{1/2}$	367.4773	367.4757	367.4749	367.4669	367.4110
		367.4773[a]	367.4764[a]	367.4682[a]	367.3858[a]
$2p_{3/2}$	367.5329	367.5311	367.5304	367.5245	367.4672
		367.5330[a]	367.5323[a]	367.5266[a]	367.4696[a]
$3p_{1/2}$	435.5468	435.5450	435.5410	435.4984	435.0648
		435.5468[a]	435.5426[a]	435.5053[a]	435.1286[a]
$3s_{1/2}$	435.5477	435.5460	435.5413	435.5039	435.1280
		435.5478[a]	435.5430[a]	435.5000[a]	435.0658[a]
$3d_{3/2}$	435.5642	435.5624	435.5583	435.5210	435.1444
		435.5642[a]	435.5613[a]	435.5354[a]	435.2736[a]
$3p_{3/2}$	435.5642	435.5625	435.5597	435.5339	435.2728
		435.5642[a]	435.5601[a]	435.5227[a]	435.1458[a]
$3d_{5/2}$	435.5700	435.5682	435.5653	435.5394	435.2778
		435.5700[a]	435.5671[a]	435.5412[a]	435.2793[a]

[a]Values from Chen *et al.*[39]

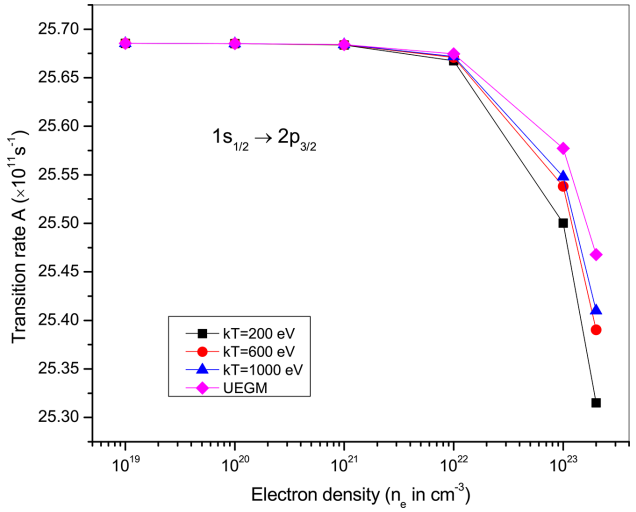


FIG. 4. Variation in transition rates for $1s_{1/2} \rightarrow 2p_{3/2}$ transition of H-like O^{7+} as a function of the free electron density.

The variation of scaled transition rates for $1s_{1/2} \rightarrow 2p_{1/2}$ transition of H-like C^{5+}, O^{7+}, and Al^{12+} ions with respect to the free electron density has been illustrated in Fig. 5. It can be seen that by considering plasma shielding effects, the scaled transition rates decrease with the increase in free electron density n_e and increase with the increase in nuclear charge for kT = 200 eV.

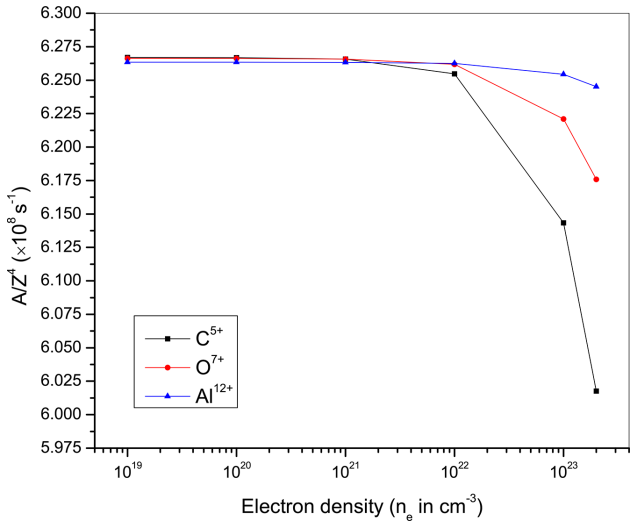


FIG. 5. Scaled transition rates (A/Z^4) for $1s_{1/2} \rightarrow 2p_{1/2}$ transition of H-like ions as a function of the free electron density at kT = 200 eV.

Published under an exclusive license by AIP Publishing

TABLE III. 1s ionization energies (in eV) of H-like O^{7+} ion at various free electron densities n_e (cm^{-3}) and temperatures kT (eV) from the present RCI calculations with analytical b-potential[40] and UEGM[7] potential.

			kT			
n_e	Plasma free	NIST[49]	200	600	1000	UEGM
0	871.4088	871.4099				
1.0×10^{19}			868.6361	868.6485	868.6522	868.6648
1.0×10^{20}			865.4050	865.4450	865.4570	865.4972
1.0×10^{21}			858.3752	858.5059	858.5447	858.6728
1.0×10^{22}			843.0036	843.4371	843.5633	843.9731
1.0×10^{23}			809.1527	810.6100	811.0230	812.3313
2.0×10^{23}			792.4632	794.5641	795.1539	797.0050

In the photoionization process, there prevails a system of (e^-+ion) with an electron in the continuum. The photoionization process corresponding to photoionization of the ground state of O VIII is given as

$$\mathrm{O\,VIII}\left(1s^2S_{1/2}\right) + h\nu \rightarrow \mathrm{O\,IX} + e^-.$$

Photoionization of the H-like ions in the ground state $1s^2S_{1/2}$ or in an excited state $nl\ ^2L_J$ leaves a bare ion (proton) with monotonically weakening interaction as a function of the energy of the ejected photoelectron.[41]

In Table III, we have tabulated the 1s ionization energies of H-like O^{7+} ion immersed in dense plasma at different values of kT and n_e. From Table III, one can see a depression of the ionization potential with the increase in free electron density n_e for a given electron temperature kT. This significant decrease in the ionization potential will apparently modify the continuum state wavefunctions and, thus, affect photoionization cross sections. The results for the unscreened case have been given. The 1s ionization energy for plasma free case is 871.4088 eV, which is in excellent agreement with the NIST value of 871.4099 eV.

Furthermore, a comparison of our calculated ionization potentials for C^{5+} ion under a dense plasma has been made with the values reported by Chen *et al.*[39] in Table IV. Our presented results are in better agreement with the previously reported results of Chen *et al.*[39] The 1s ionization energy for the plasma free case is 489.9915 eV, which agrees well with the NIST value of 489.9932 eV.

Figure 6 illustrates the variation of ionization potential depression of the ground state of O^{7+} ion with respect to free electron density at kT = 200, 600, and 1000 eV, respectively. Inspection of Fig. 6 shows that when the ion is immersed in a dense plasma environment, the ionization potential depression is significantly increased with an increase in n_e. For example, at kT = 200 eV, the ionization potential depression is about 2.77 and 78.95 eV for electron densities 1.0×10^{19} and 2.0×10^{23} cm^{-3}, respectively. Moreover, the ionization potential depression decreases with the increase in kT for a given value of n_e.

TABLE IV. Our calculated ionization potentials (in eV) for C^{5+} ion under a dense plasma obtained from UEGM[7] at different electron densities (n_e in cm^{-3}).

n_e	NIST[49]	Our results	Chen *et al.*[39]
0	489.9932	489.9915	489.9932
1.0×10^{20}		485.2676	485.2677
1.0×10^{21}		479.8148	479.8130
1.0×10^{22}		468.0720	468.0662
1.0×10^{23}		442.8225	442.8081

In Fig. 7, we show the scaled ionization potential depression of the 1s state of H-like C^{5+}, O^{7+}, and Al^{12+} ions as a function of the free electron density. For a fixed electron temperature of kT = 200 eV, the scaled ionization potential depression is significantly increased with the increase in n_e and becomes less pronounced with an increase in nuclear charge. For example, the scaled ionization potential depression is about 1.74 and 0.68 eV for C^{5+} and Al^{12+} ions at free electron density $n_e = 2.0 \times 10^{23}$ cm^{-3}, respectively.

A simple analytical expression for the photoionization cross section of an atomic subshell with quantum numbers nl within the framework of the Kramers approach[50] is written in the following form:

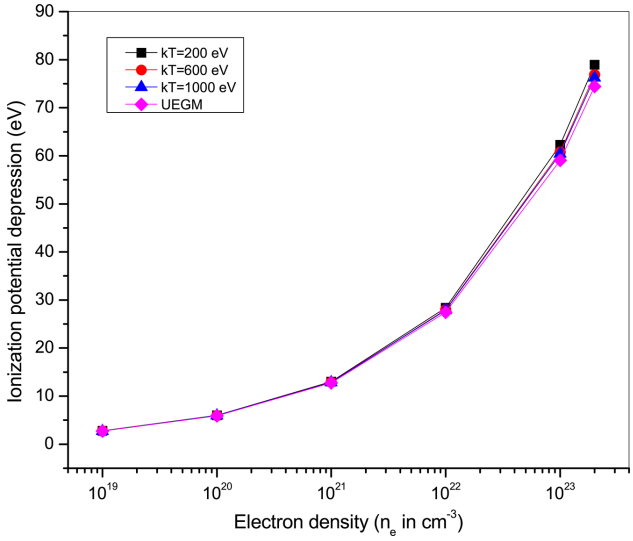


FIG. 6. Variation of ionization potential depression of the 1s state of H-like O^{7+} with respect to the free electron density.

Published under an exclusive license by AIP Publishing

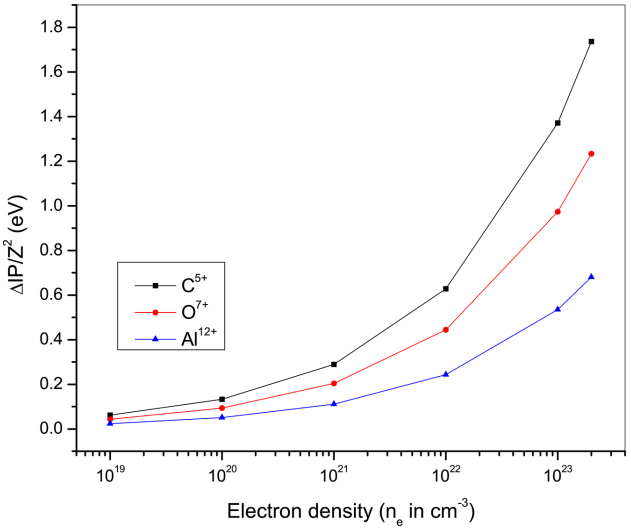


FIG. 7. Ionization potential depression ΔIP (scaled by Z^{-2}) of the 1s state of H-like ions as a function of the free electron density at kT = 200 eV.

$$\sigma_{nl}^{Kr}(\omega) = \frac{64\pi}{3\sqrt{3}} N_{nl} \frac{a_0^2}{cZ^2} \sqrt{\frac{Ry}{I_{nl}}} \left(\frac{I_{nl}}{\hbar\omega}\right)^3, \tag{13}$$

where ω is the ionizing radiation frequency, Z is the nuclear charge, I_{nl} is the ionization potential of the nl state, and this expression corresponds to the photoionization cross section for hydrogen atom in the ground state if it is assumed that $Z = N_{nl} = 1$ and $I_{nl} = Ry$ (Ry = 13.6 eV). Therefore, in Fig. 8, we have compared the photoionization cross sections for the $1s_{1/2}$ state of isolated H-like O^{7+} ion as a function of the photon energy with the photoionization cross sections obtained from the Kramers formula. The exact known value for photoionization cross section for hydrogen atom in the ground state at the ionization threshold is 6.3 Mb. If Z-scaling is used for H-like O^{7+} ion, the cross section comes out to be 0.098 Mb. The RCI and Kramers photoionization cross sections at the ionization threshold are 0.098 and 0.123 Mb. It is worth mentioning that our RCI cross section is in excellent agreement with the exact value of cross section for H-like O^{7+} ion. Additionally, in Fig. 9, a comparison of the RCI photoionization cross sections for the $1s_{1/2}$ state of H-like O^{7+} ion embedded in dense plasma has been made with the cross sections obtained from the Kramers formula at different n_e and kT. One can easily observe that the ionization thresholds are moving to the low energy region with the increase in electron densities. In dense plasma environment, the ions experience the field from the neighboring ions and free electrons, and the ionization potentials are lowered with the increase in plasma shielding. The Kramers formula overestimates our present RCI threshold value of the cross section by about 26% for the case of isolated O^{7+} ion and by about 250% for $n_e = 1.0 \times 10^{19}\,cm^{-3}$ and kT = 200 eV. However, the cross sections obtained from the Kramers formula match well with the RCI ones at higher photon energies.

Figures 10(a)–10(d) depict the photoionization cross sections for the $1s_{1/2}$, $2s_{1/2}$, $2p_{3/2}$, and $3p_{1/2}$ states of H-like O^{7+} ion under dense plasmas as a function of the photon energy using the analytical b-potential and UEGM potentials. The photoionization cross sections for the above-mentioned states have been studied at different values of free electron density and electron temperature. The photoionization cross section of hydrogenic systems is generally monotonically decreasing function of photon energy.

However, the photoionization cross sections of H-like O^{7+} show very different behavior, especially near the ionization threshold. When going from high to low photon energies, for a dense plasma, the cross section has a maximum, then a minimum and then strongly increases. With the increase in photon energy, a minimum known as the Cooper minimum appears when the electron density is sufficiently high. It is due to the fact that the sufficient plasma screening modifies the core potential, and the normal hydrogenic behavior is anticipated to break

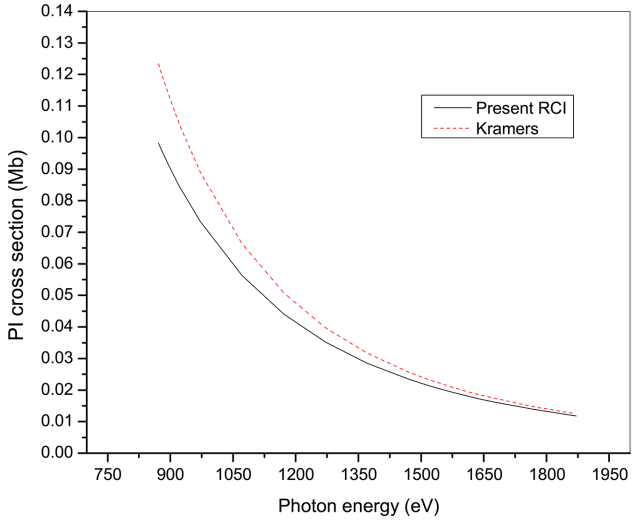


FIG. 8. The RCI and Kramers photoionization cross sections for the ground state of isolated H-like O^{7+} ion as a function of the photon energy.

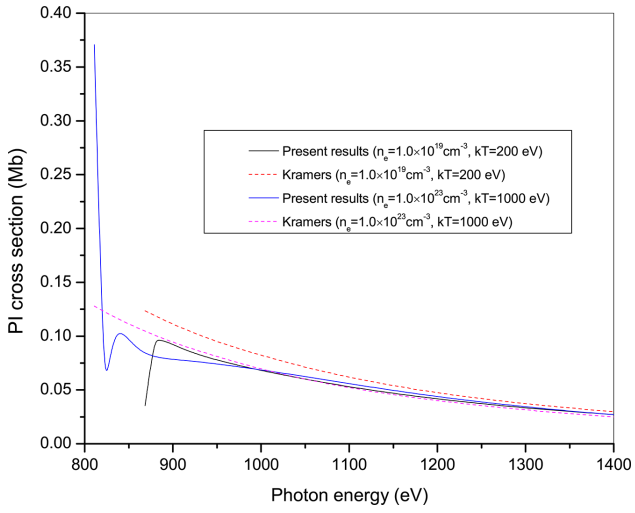


FIG. 9. The RCI and Kramers photoionization cross sections for the ground state of H-like O^{7+} ion immersed in dense plasma as a function of the photon energy.

Published under an exclusive license by AIP Publishing

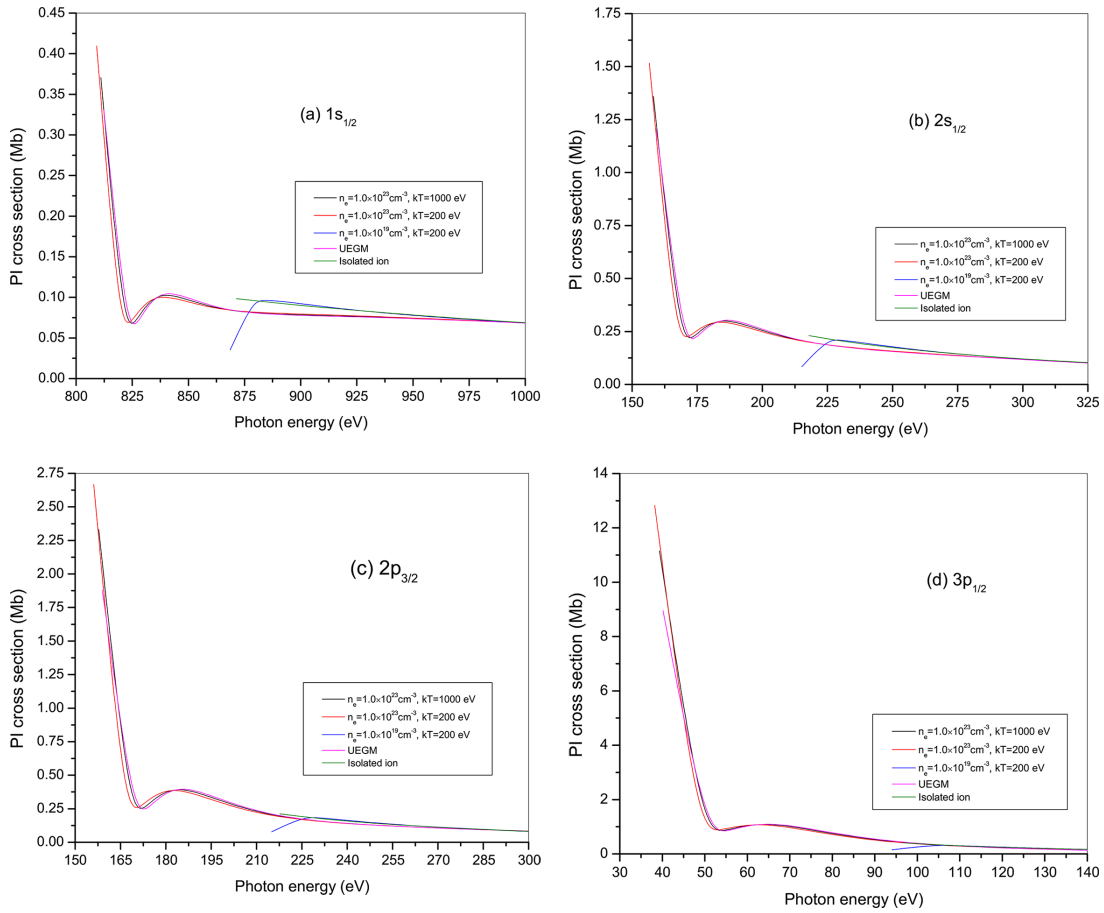


FIG. 10. (a) Photoionization cross sections of $1s_{1/2}$ state as a function of the photon energy obtained from the present RCI calculations with analytical b-potential[40] and UEGM[7] potentials for various values of plasma screening. (b) Photoionization cross sections of $2s_{1/2}$ state as a function of the photon energy obtained from the present RCI calculations with analytical b-potential[40] and UEGM[7] potentials for various values of plasma screening. (c) Photoionization cross sections of $2p_{3/2}$ state as a function of the photon energy obtained from the present RCI calculations with analytical b-potential[40] and UEGM[7] potentials for various values of plasma screening. (d) Photoionization cross sections of $3p_{1/2}$ state as a function of the photon energy obtained from the present RCI calculations with analytical b-potential[40] and UEGM[7] potentials for various values of plasma screening.

down near the ionization threshold. The photoionization cross section obeys usual monotonically decreasing behavior in the region of higher photon energy.

One can easily observe that the ionization thresholds are moving to the low energy region with the increase in electron densities. The photoionization cross section values increase with an increase in n_e for a fixed kT, whereas those decrease with the increase in kT for a fixed n_e. Taking the example of photoionization for the $1s_{1/2}$ state of O^{7+} ion, our evaluation shows that for $kT = 200$ eV, the photoionization cross sections at the ionization threshold increase by about 64.02% and 316.61%, respectively, at 1.0×10^{19} and 1.0×10^{23} cm^{-3}, as compared to the isolated ion cross sections. Furthermore, the plasma shielding effect is more pronounced for higher lying states as compared to lower lying states. Furthermore, the radiative recombination cross sections are calculated by making use of the Milne relation. It can be seen from Fig. 11 that the plasma screening effect on the radiative recombination cross sections is similar to that of photoionization cross sections.

IV. CONCLUSION

To summarize, we have investigated the photoionization cross sections of the H-like O^{7+} ion embedded in a dense plasma environment. The analytical b-potential and UEGM potential have been incorporated within the relativistic configuration interaction method for including the plasma screening effects. The bound and continuum state wavefunctions are determined by solving the modified Dirac equations of atoms/ions embedded in dense plasma environment. The calculations for excitation energies, ionization potentials, radiative parameters, photoionization, and radiative recombination cross sections of O^{7+} ion are carried out in the density and temperature ranges of $n_e = 1.0 \times 10^{19}$–2.0×10^{23} cm^{-3} and $T_e = 200$–1000 eV, respectively. The photoionization cross sections evaluated in the dipole

Published under an exclusive license by AIP Publishing

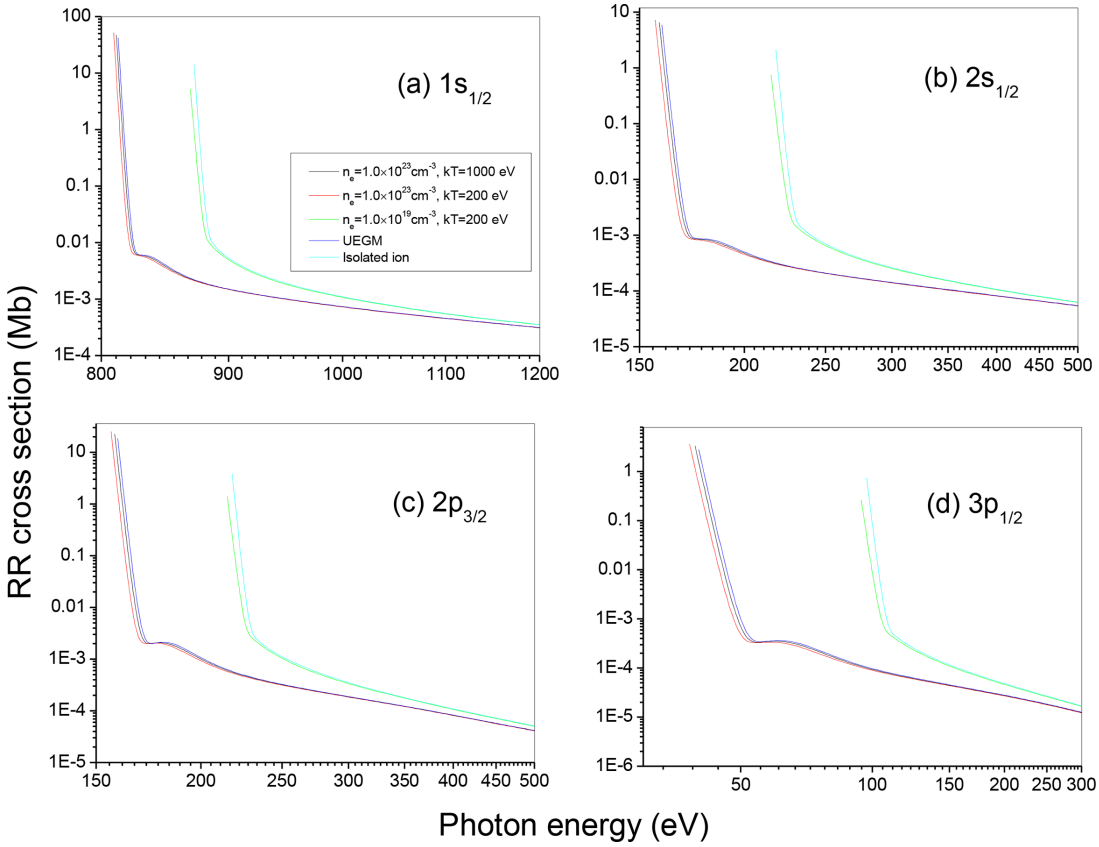


FIG. 11. Radiative recombination cross sections as a function of the photon energy obtained from the present RCI calculations with analytical b-potential[40] and UEGM[7] potentials for various values of plasma screening.

approximation are obtained from the bound-free transition matrix element. We found a significant decrease in the ionization potential with an increase in free electron densities that has apparently modified the continuum state wavefunctions and photoionization cross sections, which has been highlighted for the first time for H-like O^{7+} ion. The photoionization cross sections increase with an increase in n_e for a fixed kT, while decrease with an increase in kT for a fixed n_e. Moreover, a similar trend was exhibited by the radiative recombination cross sections. We believe that the findings of the present study will have potential implications in the analysis of spectra observed in laboratory and astrophysical plasmas.

ACKNOWLEDGMENTS

This work was carried out under the Project No. EMR/2016/001203, sponsored by SERB, the Department of Science and Technology, Government of India, at Deen Dayal Upadhyaya College, University of Delhi, India. Dishu Dawra gratefully acknowledges the Department of Physics and Astrophysics for Departmental fellowship. Mayank Dimri is thankful to DST for assisting him as a Junior Research Fellow. The authors are grateful to Professor Anil K. Pradhan and Professor Sultana N. Nahar for their valuable guidance.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]C. Gao, J. Zeng, and J. Yuan, High Energy Density Phys. **7**, 54 (2011).
[2]J. Pang, G. Han, Z. Q. We, and S. C. Li, J. Phys. B **35**, 2117 (2002).
[3]Y. Li, J. Wu, Y. Hou, and J. Yuan, J. Phys. B **41**, 145002 (2008).
[4]T. C. Killian, T. Pattard, T. Pohl, and J. M. Rost, Phys. Rep. **449**, 77 (2007).
[5]R. Piron and T. Blenski, Phys. Rev. E **83**, 026403 (2011).
[6]M. S. Murillo, J. Weisheit, S. B. Hansen, and M. Dharma-Wardana, Phys. Rev. E **87**, 063113 (2013).
[7]B. Saha and S. Fritzsche, J. Phys. B **40**, 259 (2007).
[8]M. Nantel, G. Ma, S. Gu, C. Y. Cote, J. Itatani, and D. Umstadter, Phys. Rev. Lett. **80**, 4442 (1998).
[9]S. Skupsky, Phys. Rev. A **21**, 1316 (1980).
[10]O. Renner and F. B. Rosmej, Matter Radiat. Extremes **4**, 024201 (2019).
[11]B. Saha, P. K. Mukherjee, and G. H. F. Diercksen, Astron. Astrophys. **396**, 337 (2002).
[12]D. Salzman, *Atomic Physics in Hot Dense Plasma* (Oxford University Press, New York, 1998).

Published under an exclusive license by AIP Publishing

[13]F. B. Rosmej, L. A. Vainshtein, V. A. Astapenko, and V. S. Lisitsa, Matter Radiat. Extremes **5**, 064202 (2020).
[14]R. Geneaux, H. J. B. Marroux, A. Guggenmos, D. M. Neumark, and S. R. Leone, Philos. Trans. R. Soc. A **377**, 20170463 (2019).
[15]See https://flash.desy.de/ for "FLASH," 2021.
[16]See https://www.xfel.eu/ for "EU-XFEL," 2021.
[17]See https://lcls.slac.stanford.edu/ for "LCLS," 2021.
[18]P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F.-J. Decker, and Y. Ding, Nat. Photonics **4**, 641 (2010).
[19]F. B. Rosmej and R. W. Lee, Europhys. Lett. **77**, 24001 (2007).
[20]S. Ichimaru, Rev. Mod. Phys. **54**, 1017 (1982).
[21]M. S. Murillo and J. C. Weisheit, Phys. Rep. **302**(1), 1–65 (1998).
[22]S. Bhattacharyya, J. Saha, and T. Mukherjee, Phys. Rev. A **91**, 042515 (2015).
[23]F. B. Rosmej, J. Phys. B **51**, 09LT01 (2018).
[24]S. Sahoo and Y. K. Ho, Phys. Plasmas **13**, 063301 (2006).
[25]Y. Y. Qi, Y. Wu, and J. G. Wang, Phys. Plasmas **16**, 033507 (2009).
[26]S. Sahoo and Y. K. Ho, J. Quant. Spectrosc. Radiat. Transfer **111**, 52 (2010).
[27]C. Y. Lin and Y. K. Ho, Phys. Plasmas **17**, 093302 (2010).
[28]C. Y. Lin and Y. K. Ho, Phys. Scr. **T144**, 014051 (2011).
[29]C. Y. Lin and Y. K. Ho, Eur. Phys. J. D **57**, 21 (2010).
[30]T. N. Chang and T. K. Fang, Phys. Rev. A **88**, 023406 (2013).
[31]M. Das, Phys. Plasmas **21**, 012709 (2014).
[32]L. Y. Xie, J. G. Wang, and R. K. Janev, Phys. Plasmas **21**, 063304 (2014).
[33]Y. Y. Qi and L. N. Ning, Phys. Plasmas **21**, 033301 (2014).
[34]C. F. Clauser and R. O. Barrachina, J. Phys. B **49**, 235702 (2016).
[35]Y. Y. Qi, J. G. Wang, and R. K. Janev, Phys. Plasmas **24**, 062110 (2017).
[36]H. Hu, Phys. Plasmas **24**, 052112 (2017).
[37]G. P. Zhao, L. Y. Xie, L. Liu, J. G. Wang, and R. K. Janev, Phys. Plasmas **25**, 083302 (2018).
[38]Z. B. Chen, Y. Y. Qi, H. Y. Sun, G. P. Zhao, P. F. Liu, and K. Wang, J. Quant. Spectrosc. Radiat. Transfer **253**, 107170 (2020).
[39]Z. B. Chen and K. Wang, J. Quant. Spectrosc. Radiat. Transfer **245**, 106847 (2020).
[40]X. Li and F. B. Rosmej, Phys. Lett. A **384**, 126478 (2020).
[41]A. K. Pradhan and S. N. Nahar, *Atomic Astrophysics and Astronomy* (Cambridge University Press, Cambridge, 2011).
[42]K. Kilimann and W. Ebeling, Z. Naturforsch., A **45**, 613 (1990).
[43]M. F. Gu, Can. J. Phys. **86**, 675 (2008).
[44]D. Dawra, M. Dimri, A. K. Singh, A. K. S. Jha, and S. S. Singh, J. At. Mol. Condens. Matter Nano Phys. **6**, 103 (2019).
[45]A. Singh, D. Dawra, M. Dimri, A. K. Jha, and M. Mohan, Phys. Plasmas **26**, 062704 (2019).
[46]A. Singh, M. Dimri, D. Dawra, A. K. Jha, N. Verma, and M. Mohan, Radiat. Phys. Chem. **156**, 174 (2019).
[47]A. Singh, M. Dimri, D. Dawra, A. K. Jha, and M. Mohan, Can. J. Phys. **97**, 436 (2019).
[48]A. K. Singh, D. Dawra, M. Dimri, A. K. Jha, R. K. Pandey, and M. Mohan, Phys. Lett. A **384**, 126369 (2020).
[49]A. Kramida, Yu. Ralchenko, J. Reader, and NIST ASD Team, *NIST Atomic Spectra Database (ver. 5.6.1)* (National Institute of Standards and Technology, Gaithersburg, MD, 2021).
[50]V. I. Kogan, A. B. Kukushkin, and V. S. Lisitsa, Phys. Rep. **213**, 1–116 (1992).

Published under an exclusive license by AIP Publishing

Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Junctionless Accumulation Mode Ferroelectric FET (JAM-FE-FET) for High Frequency Digital and Analog applications

Snehlata Yadav
Delhi Technological University
Sonam Rewari (rewarisonam@gmail.com)
Delhi Technological University https://orcid.org/0000-0001-8946-9849
Rajeshwari Pandey
Delhi Technological University

Research Article

Keywords: Ferroelectric, Negative capacitance, Junctionless Accumulation mode (JAM), RF parameters, HZO

Posted Date: November 10th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-943468/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Junctionless Accumulation Mode Ferroelectric FET (JAM-FE-FET) for high frequency Digital and Analog applications

Snehlata Yadav[1], Sonam Rewari[1], Rajeshwari Pandey[1]
snehlata_2k19phdec03@dtu.ac.in, rewarisonam@gmail.com, rajeshwaripandey@gmail.com
[1] Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi-110042, India

Abstract:
In this paper, a Junctionless Accumulation Mode Ferroelectric Field Effect Transistor (JAM-FE-FET) has been proposed and assessed in terms of RF/analog specifications for varied channel lengths through simulations using TCAD Silvaco ATLAS simulator, using the Shockley-Read-Hall (SRH) recombination, ferro, Lombardi CVT, fermi and LK models. Major analog metrics like transconductance (g_m), intrinsic gain (A_V), output conductance (g_d), and early voltage (V_{EA}) are obtained for the JAM-FE-FET arrangement. The proposed structure shows an improvement in parameters like gm, I_{on}/I_{off}, A_v, TGF by 6.82%, 27.95%, 5.2%, 38.83% respectively. Further, frequency analysis of the proposed device is performed and several critical RF parameters like f_T, TFP, GFP, and GTFP have been observed to be enhanced by 6.89%, 11.38%, 13.65%, 12.01% respectively. Thus, the Junctionless accumulation mode ferroelectric FET (JAM-FE-FET) arrangement has been found to have superior analog and RF performance when compared to Junctionless ferroelectric FET(JL-FE-FET). As a result, the JAM-FE-FET device presented here can be contemplated a good contender for applications in high-frequency systems.

Keywords: Ferroelectric, Negative capacitance, Junctionless Accumulation mode (JAM), RF parameters, HZO

1. Introduction:

Various MOSFET structures have been realized over the last few decades, and their scaling has been quite successful down to the nano scale, resulting in a significant increase in performance [1],[2]. However, when the size of such transistors is reduced, junctions become closer, which is difficult due to the significant doping concentration gradients required. In order to combat this problem, J. E. Lilienfeld in the 1920s introduced the concept of Junctionless transistors [3]. It was successfully fabricated at the Tyndall Institute by Colinge et al [4]. The major characteristic feature of this device is the absence of p-n junction which avoids the requirement of gradients in doping concentration[5]. Various analytical study of surface potential for junctionless transistor has also been done[6],[7]. It has also been reported that cylindrical surrounding gate MOSFETs shows good switching performance and also can be used for microwave frequency applications[8],[9],[10]. However, the junctionless transistor poses various limitations such as degraded mobilities due to high doping concentrations.

Also, some of the significant obstacles in the device include the higher gate work function for completely depleted channel region in order to insure turning off the device. To address the aforementioned issues, a new modified structure known as the JAM FET was introduced [11], which has highly doped S/D regions and decreased doping in the channel, resulting in reduced mobility deterioration [12]. Another major issue that has arisen as a result of shrinking and the increased density of transistors on a chip is higher power consumption and heat dissipation, both of which slow down data processing rate [13],[14]. This can be enhanced by overcoming the lower working voltage restriction known as 'Boltzmann's Tyranny,' which allows the transistor operation voltage to be reduced, lowering power consumption. One such promising device is the ferroelectric field effect transistors (FE-FETs) [15]. Various investigations to understand the behavior of FE-FETs have been conducted in the past [16],[17],[18]. The majority of the research, however, concentrated on hysteretic behavior in memory applications [19],[20].

The discovery of ferroelectric properties in hafnium oxide (HfO_2) [21] in 2011 has gained a lot of attention [22],[23] because of its CMOS compatibility. Higher remnant polarization and large coercive field, at minimal thickness [24],[25], and superior performance characteristics have been demonstrated in ferroelectric HfO_2[26]. Due to its lower annealing temperature [27] and customizable ferroelectric properties, zirconium-doped HfO_2 (HZO) has popped up as a potential material [28],[29]. Apart from realizing memory, ferroelectric hafnium oxide is the sole material system that may be utilized to realize a new form of steep slope device called a Negative Capacitance Field Effect Transistor (NCFET) that was proposed in 2008 [30]. Negative capacitance FET has been studied both theoretically and practically for digital applications due to its ability to produce a subthreshold swing (SS) of 60 mV/decade [31],[32]. Although there are several demonstrations and investigations of digital and DC behavior of NCFET in the reported literature, the high frequency (RF) capabilities of NCFETs are yet to be fully explored. For RF applications of NCFET in future, research in this direction is crucial. Most of the previous research on NCFET have focused on Metal ferroelectric metal insulator semiconductor (MFMIS) configuration structure

[33],[34],[35],[36]. This structure simplifies modeling since the ferroelectric–insulator interface has a uniform potential, but the leakage currents would render the negative capacitance of ferroelectric FET unstable. Because of this instability, it is difficult to bias MFMIS devices and circuits, making the MFMIS structure unsuitable for RF applications. A metal–ferroelectric–insulator–semiconductor (MFIS) structure, on the other hand, mitigates these issues [37] and produces different overall device properties as a result of a spatially changing ferroelectric potential than MFMIS devices. Thus, the MFIS structure have always been preferred for integrating in advanced technologies. RF performance of MFIS NCFETs has also been investigated in few researches. Experiment in [38] revealed a modest improvement in f_T, while simulations by [39] revealed circuit-level figures of merit. However, no report has been published on the RF performance of JAM- FE-FET. Quantum confinement is also investigated in relation to ferroelectric thickness (t_{FE}) and channel thickness (t_{ch}). The quantum confinement effect reduces when the valus of t_{FE} and t_{ch} are greater than 7 nm and 15 nm respectively [40]. In our study, quantum mechanical effect has not been considered and therefore, the channel thickness and ferroelectric thickness are restricted to 20 nm and 10nm respectively. Thus, motivated by these coexisting research findings, a novel device structure, JAM-FE-FET is reported in this paper. This study takes the entire RF capabilities of NCFET with MFIS structure and incorporating the benefits of JAM FET.
Following the introduction, the next section delves into the device structure and simulation methods. The suggested device's performance parameters are examined in Section 3. Conclusions are drawn in the final section.

2. Proposed Structure & Simulation Parameters:

The schematic illustration of a JAM-FE-FET is shown in Fig. 1(a). All simulations are run on the TCAD Silvaco ATLAS simulator, with the Shockley-Read-Hall (SRH) recombination, ferro model, Lombardi CVT model, fermi and Landau-Khalatnikov (LK) models [41] being used. According to previous research, the negative capacitance is caused by a unique relationship between the electric field (E_{FE}) and polarization (P) of the ferroelectric layer which is given by the following equation:

$$E_{FE} = \alpha P + \beta P^3 + \gamma P^5 + \rho \frac{dP}{dt} \quad (1)$$

where α, β, and γ are the ferroelectric material parameters and ρ is the kinetic coefficient linked to the time constant associated with change in ferroelectric polarization. The parameters for the compared devices are enlisted in Table 1. The proposed structure comprising a gate-stack of TiN/HZO/SiO_2 is employed in the simulations. The channel length (L) is varied from 90 nm to 32 nm and channel thickness (t_{ch}) is taken 20 nm. A ferroelectric layer with a thickness (T_{FE}) of 10 nm and an insulator layer thickness (t_{IL}) of 0.9 nm in the gate stack. In JAM-FE-FET, the doping in the silicon channel is 1e17 cm^{-3} with n-type dopants, whereas the source and drain regions are doped strongly with 1e19 cm^{-3} n-type dopants. The doping level for the entire simulation is considered to be uniform. Titanium Nitride (TiN) having work function (ϕ_m) 4.65eV is used as gate material. Thereafter, numerical approaches such as Gummel and Newton, have been contemplated to improve convergence. In addition to these, the specific properties of the ferroelectric material (HZO) are listed in Table 2, that induces the negative capacitance effect. The drain to source voltage (V_{ds}) was set to 50 mV and the gate to source voltage (V_{gs}) was varied from 0 to 1 V to obtain the transfer characteristics. Fig. 1(b) shows the drain current with and without quantum model. As seen in Fig. 1(b), quantum effects have no influence on the transfer characteristics and have thus been ignored in this study.

3. Device fabrication and calibration:

The fabrication of ferroelectric negative capacitance field effect transistor has been explained by D. Kwon et al. in 2019 [42]. The transistor can be fabricated by taking a silicon substrate followed by the active region formation using photolithography.

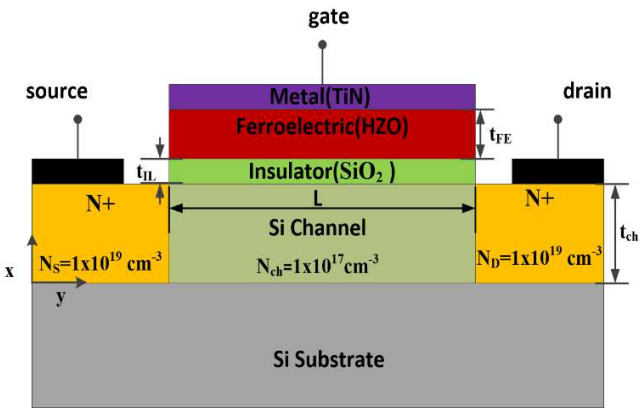


Fig. 1(a). A schematic representation of JAM-FE-FET

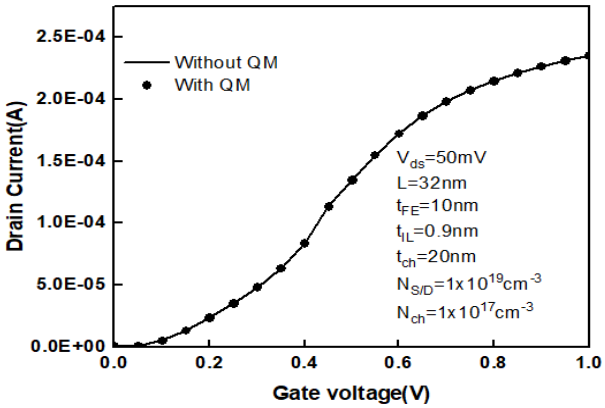


Fig. 1(b). The effect of the Quantum Model (QM) on the drain current

Table 1. Various parameters employed for the device simulation

Parameter	Symbol	JL-FE-FET	JAM-FE-FET
Channel length	L	90,45,32 nm	90,45,32 nm
Channel thickness	t_{ch}	20 nm	20 nm
Ferroelectric thickness	t_{FE}	10 nm	10 nm

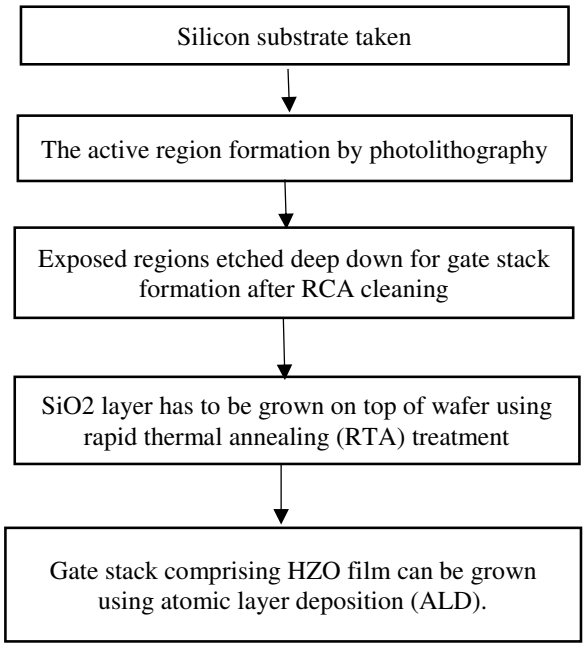
Insulator thickness	t_{IL}	0.9 nm	0.9 nm
Source/Drain doping	$N_{S/D}$	$1x10^{19}$ cm^{-3}	$1x10^{19}$ cm^{-3}
Channel doping	N_{ch}	$1x10^{19}cm^{-3}$ (n-type)	$1x10^{17}cm^{-3}$(n-type)
Metal work function	ϕ_m	4.65 eV	4.65 eV

Table 2. Properties of ferroelectric material (HZO)

Parameter	Symbol	Values
Spontaneous Polarization	Ps	10-40μC/cm^2
Remnant Polarization	Pr	1-40μC/cm^2
Coercive Field	Ec	1-2MV/cm
Dielectric constant	ε	30

Next, the exposed regions have to be etched deep down for the gate stack formation after RCA cleaning. On top of the wafer, rapid thermal annealing (RTA) treatment can be used to grow SiO2 layer. Further, the ferroelectric gate oxide, an HZO film deposition using atomic layer deposition (ALD) has to be performed. For short channel devices, electron beam lithography can be used to define the gate region. Ion implantation can then be performed on the exposed source/drain (S/D) areas. Next, the post metallization anneal in N_2 ambient can be performed for dopant activation and finally metal contacts of sputtered TiN gate electrode can be formed at the top. The steps of fabrication process is shown by a flowchart in Fig. 2(a).

The proper calibration of this research work is done with the experimental ferroelectric research under the same device dimensions [42]. The Shockley-Read-Hall (SRH) recombination, ferro model, Lombardi CVT model, fermi and Landau-Khalatnikov (LK) models are used for calibration. The transfer characteristic curves in Fig. 2(b), are in good agreement with one other.



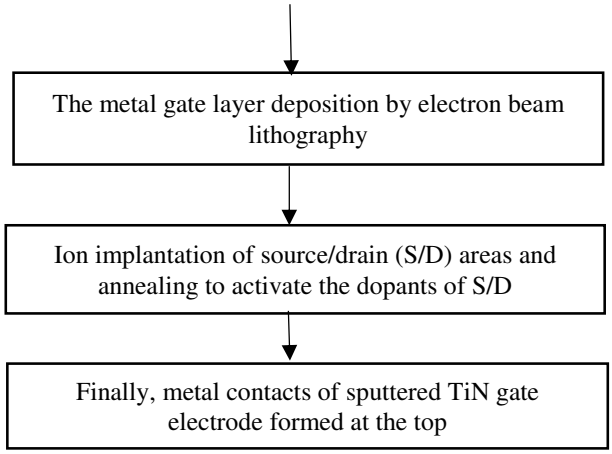


Fig. 2(a). Fabrication flowchart of proposed JAM-FE-FET

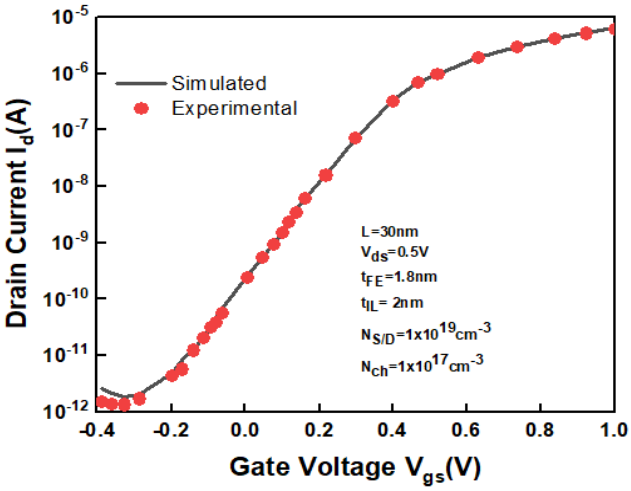


Fig. 2(b). Calibrated transfer characteristics of simulation data with experimental data[42]

4. Results & Discussions:

Fig. 3 demonstrates how the drain current of JAM-FE-FET and JL-FE-FET varies with gate voltage for various channel lengths. Fig. depicts that JAM-FE-FET have higher drain current over JL-FE-FET. The greater mobility of carriers in the channel is responsible for this improvement. The drain current (I_{on}) increases dramatically when the channel length L is scaled down from 90 nm to 32 nm, as shown in Fig. 3. The results for I_{on} increment in short channel can be intuitively explained through velocity saturation theory in which the inversion region current is in proportion to the device total oxide capacitance [43].

Table 3. I_{on}/I_{off} Ratio

Parameter	L=32nm		L=45nm		L=90nm	
	JAM-FE-FET	JL-FE-FET	JAM-FE-FET	JL-FE-FET	JAM-FE-FET	JL-FE-FET
I_{on}	2.35×10^{-4}	2.21×10^{-4}	2.09×10^{-4}	1.95×10^{-4}	1.54×10^{-4}	1.43×10^{-4}
I_{off}	6.58×10^{-10}	7.96×10^{-10}	6.68×10^{-10}	8.04×10^{-10}	6.9×10^{-10}	8.37×10^{-10}
I_{on}/I_{off}	35.7×10^{4}	27.7×10^{4}	31.2×10^{4}	24.2×10^{4}	22.3×10^{4}	17.08×10^{4}

In addition to suppressing I_{off} by increasing total oxide capacitance, drain coupling in NCFET at lower gate voltage causes a repression of channel potential and a rise in energy barrier that grows greater as capacitance matching improves, further suppressing I_{off}.

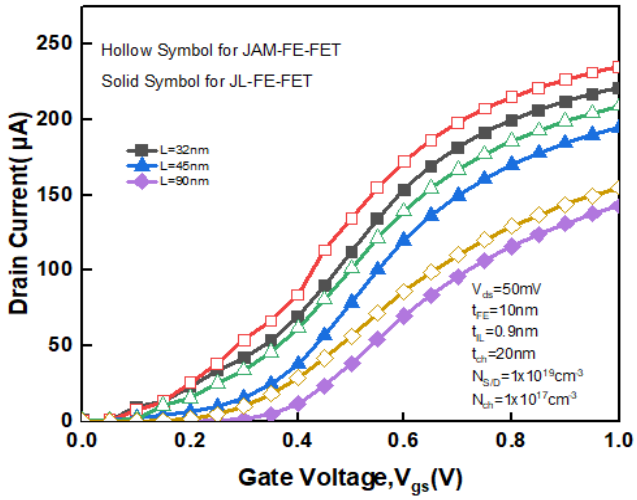


Fig. 3. Drain current variation with V_{gs}

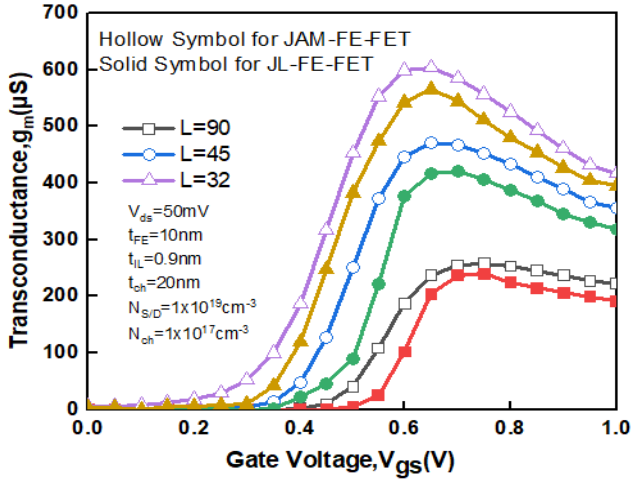


Fig. 4. Transconductance variation with V_{gs}

Transconductance is a measurement of the relationship between the deviation in drain current and the change in V_{gs} at constant V_{ds}. As illustrated in Fig. 4, gm is calculated using the I_d-V_{gs} curve derivative. Since the gate control over channel has been enhanced, and short channel effects have been decreased, the JAM-FE-FET has the highest transconductance value when compared to JL-FE-FET.

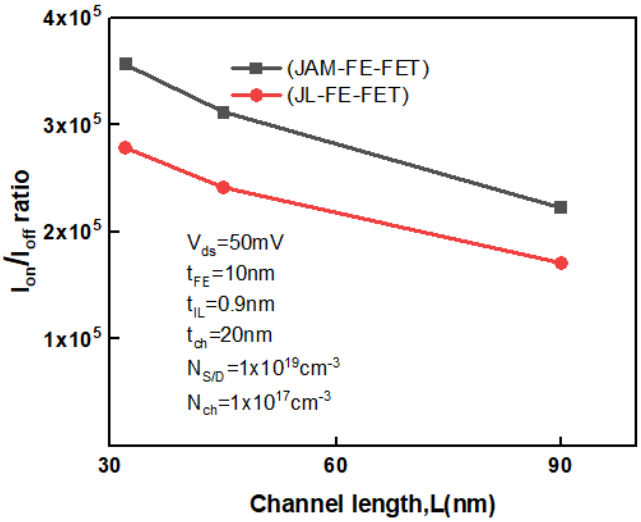


Fig. 5. I_{on}/I_{off} variation with channel length

Additionally, gate-stack construction improves average carrier velocity, which leads to higher electron mobility and, eventually, increased gm.

Fig. 5 shows the I_{on}/I_{off} ratio variation for JAM-FE-FET and JL-FE-FET for different channel length. I_{on} and I_{off} for these devices are obtained at V_{gs}= 1V and V_{gs}= 0V respectively for V_{ds}= 50mV. It is one of the most critical parameters for the digital applications. The I_{on}/I_{off} ratio for the compared devices are shown in Table 3.

It can be assessed that, I_{on}/I_{off} ratio for channel length L= 90 nm is low for digital devices. However, an improvement in I_{on}/I_{off} ratio of JAM-FE-FET over JL-FE-FET for channel length L= 32 nm has been obtained due to increased mobility in JAM configuration leading to increase in drain current with reduction in current leakage and consequently greater switching ratio.

The subthreshold swings for the compared devices are shown in Fig. 6. It has been discovered that both the devices have steep SS behavior (<60mV/dec). Subthreshold slope values fewer than 60 mv/dec have also been seen in previous research studies[44]. Hence the device can be switched quickly over a wide range of current. It can also be observed that as the gate length is reduced the subthreshold value increases.

SS can be expressed as:

$$SS = \frac{\partial V_{gs}}{\partial log_{10} I_d} \tag{2}$$

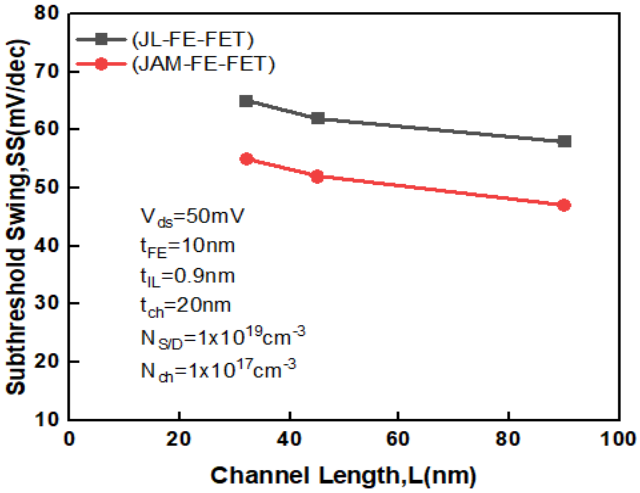


Fig. 6. SS variation with channel length

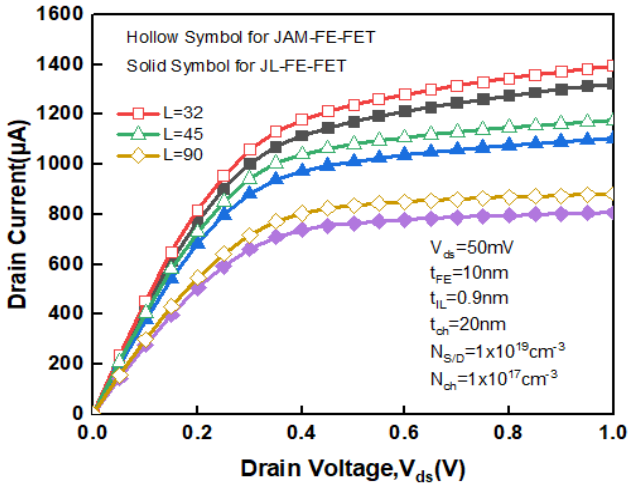


Fig. 7. Drain current variation with V_{ds}

For V_{gs}= 1V, Fig. 7 depicts the drain current variation with drain voltage. The figure clearly indicates that, similar to JL-FE-FET, in the linear region drain current is almost identical, but in the saturation region, device with shorter gate lengths display greater saturation currents.

Fig. 8 depicts the output conductance variation for the compared device structures. The output conductance can be calculated by varying the drain current with the drain to source voltage while maintaining the constant gate to source voltage. It can be expressed as [45]:

$$g_d = \frac{\partial I_d}{\partial V_{ds}} \tag{3}$$

It can be inferred from the figure that g_d is higher in the linear region and keep up a constant value in the saturation region. Thus, the driving capability of the proposed device is greater than the compared one.

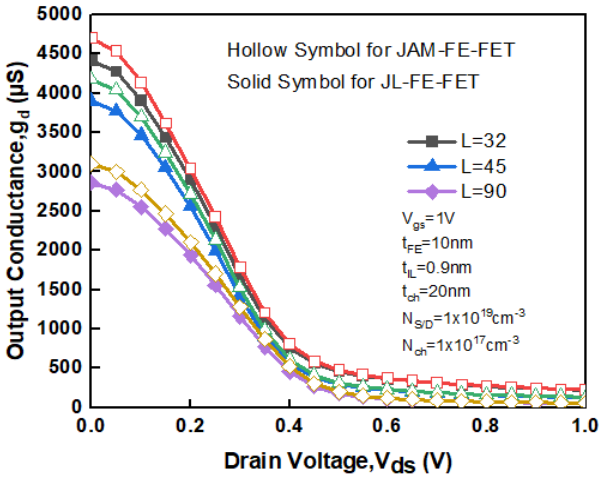


Fig. 8. Output conductance variation with V_{ds}

It can also be assessed from the graph that the output conductance increases when the scaling of channel length is done from 90 nm to 32 nm due to the suppressed short channel effects and the increased gate controllability.

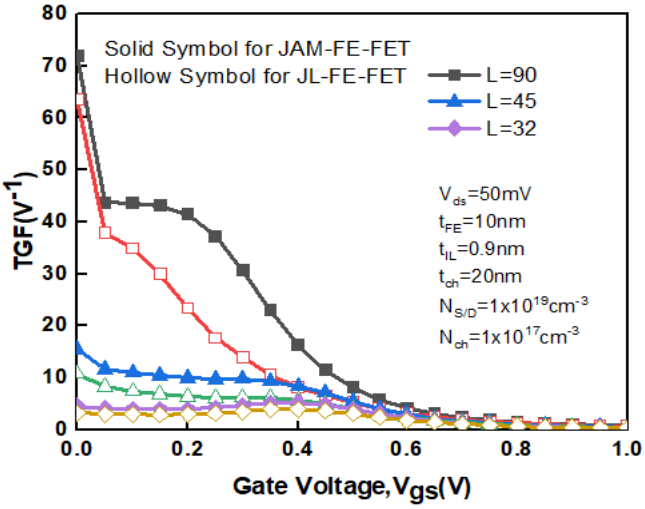


Fig. 9. TGF as a function of V_{gs}

Transconductance Generation factor (TGF) can be defined as the accessible gain per unit power loss. It can be expressed as

$$TGF = \frac{g_m}{I_d} \tag{4}$$

The device that operates at lower supply voltage performs better for higher TGF values. It is clearly evident from Fig.9 that the proposed device structure attains the maximum TGF value. Since the drain current is higher, it corresponds to high value of transconductance and eventually high TGF.

Immunity to the channel length modulation (CLM) is provided by Early Voltage (V_{EA}) [45]. Early Voltage is expressed as:

$$V_{EA} = \frac{I_d}{g_d} \tag{5}$$

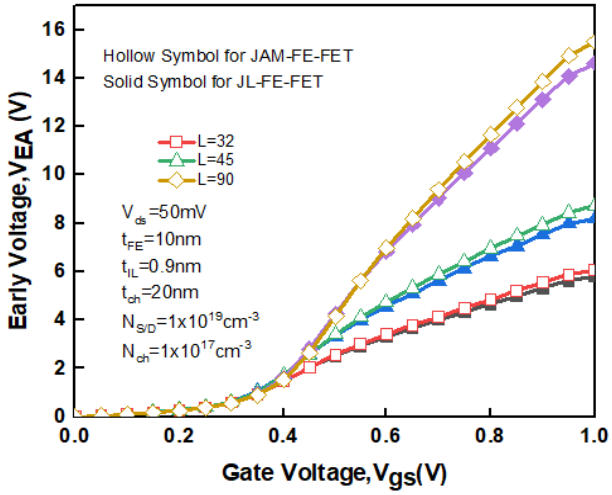


Fig. 10. Early Voltage variation with V_{gs}

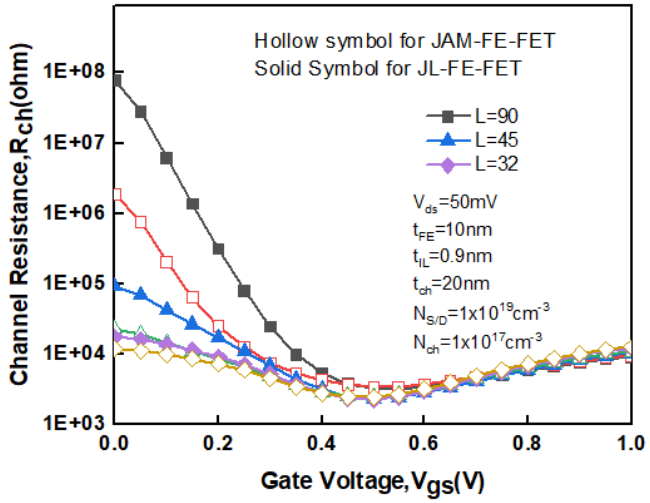


Fig. 11. Channel resistance variation with V_{gs}

The drain current rise, I_d in JAM-FE-FET is greater than the decrease in transconductance, resulting in a higher V_{EA} than in JL-FE-FET. For varied channel lengths, Fig. 10 depicts the early voltage shift as a function of gate bias. A greater V_{EA} indicates that the device will have higher gain and can be used in amplifiers [46]. Fig. 11 shows the channel resistance variation with the gate voltage. The channel resistance should be minimum for higher driving current. As can be observed from the figure that the JAM-FE-FET exhibits lower channel resistance owing to its increased carrier density and velocity in the channel region.

The next figure of merit is the intrinsic dc gain which can be defined as the transconductance (g_m) to output conductance (g_d) ratio, i.e., g_m/g_d [46]. Since g_d is extracted from static I_d–V_{ds} curve, so it is the low frequency or quasi-static result. It is a critical parameter for practical transconductance amplifiers. The change of g_m/g_d with gate voltage is shown in Fig. 12 for L= 32 nm, L= 45 nm and L= 90 nm for the compared devices. It can be observed from Fig. 12 that the intrinsic gain is higher at channel length 90 nm because of low output conductance which is essential for analog applications.

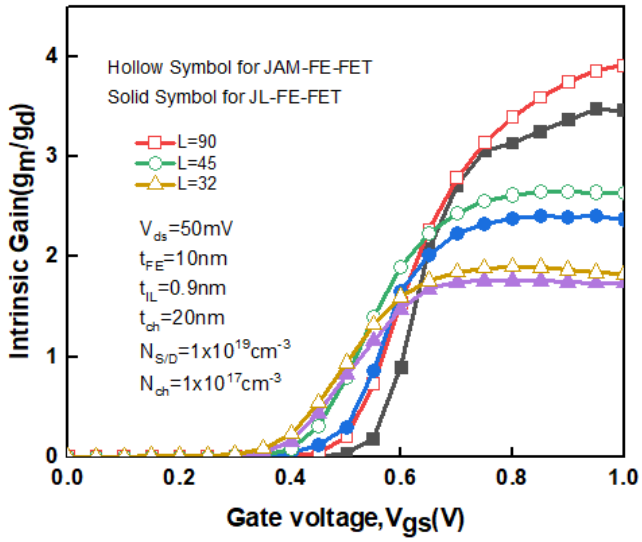


Fig. 12. Intrinsic gain variation with V_{gs}

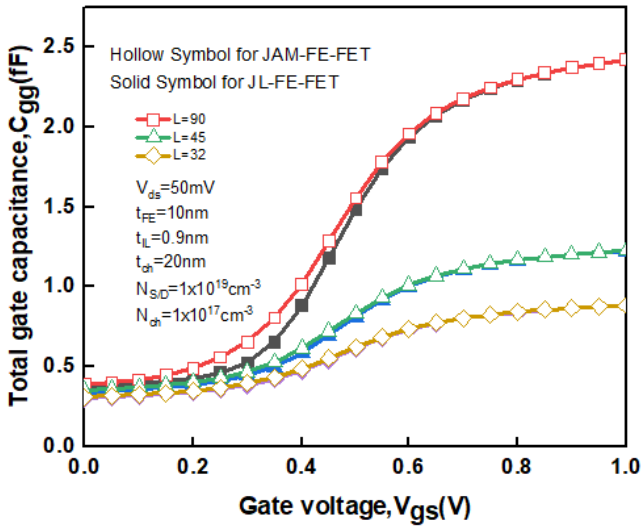


Fig. 13. C_{gg} variation with V_{gs}

The capacitive behavior of a device controls its high frequency functioning. Fig. 13 shows how a change in gate bias (V_{gs}) affects total gate capacitance (C_{gg}) in JAM-FE-FET and JL-FE-FET. It can be observed that JAM-FE-FET reveals profiles that are considerably closer to those of JL-FE-FET.

Cut off frequency denoted as f_T is an intrinsic property of the device and represents a figure of merit for high frequency operation. It can be expressed as:

$$f_T = g_m/2\pi C_{GG} \quad (6)$$

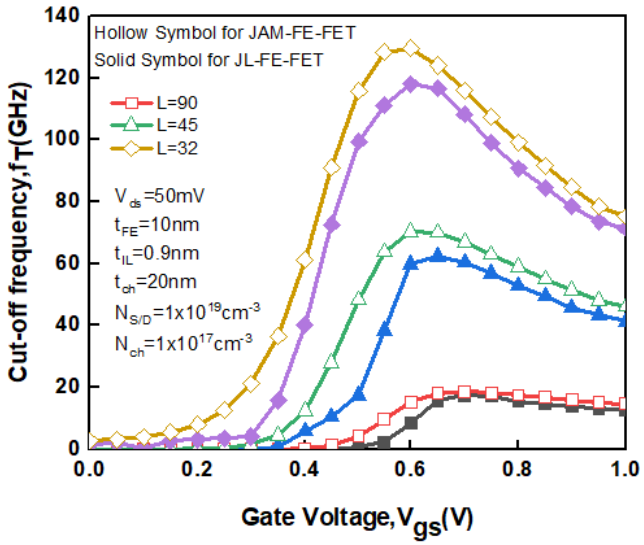


Fig. 14. f_T variation with V_{gs}

Cut-off frequency for different channel lengths with respect to gate voltage is shown in Fig. 14. The high value of cut-off frequency is observed because of high g_m value and comparable C_{gg}. However the cut off frequency further reduces for high gate voltages due to the increase in total gate capacitance.

The transducer power gain (G_T) can be defined as the P_{load} to the P_{source} ratio, where P_{load} is the average power given to the load, and P_{source} is the average power available from the source. It can be expressed as:

$$G_T = \frac{P_{load}}{P_{source}} \quad (7)$$

Maximum-Transducer-Power-Gain (MTPG) is described as a power gain that can be obtained when load is driven with the identical inputs.

The MTPGs for JL-FE-FET and JAM-FE-FET are shown in Fig. 15 for channel lengths varying from 30 nm to 90 nm. From the Fig. 15, it can be clearly interpreted that JAM-FE-FET has a higher gain owing to its architecture which overcomes the deterioration in mobility. It can also be inferred from the figure that as the channel length is scaled down from 90 nm to 45 nm the transducer gain increases drastically.

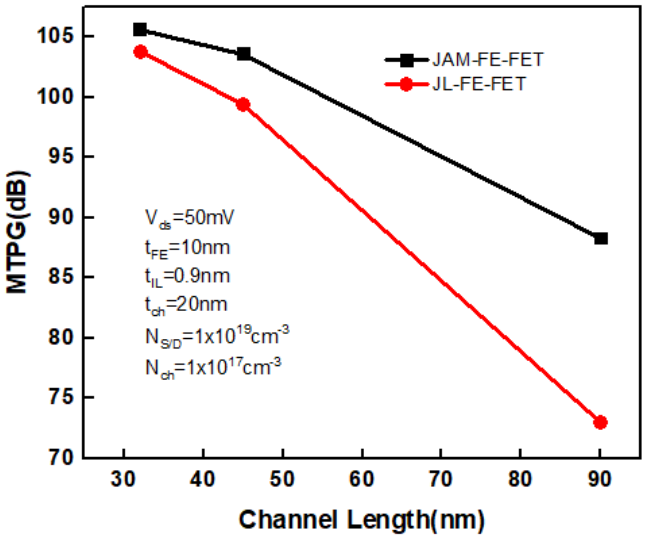


Fig. 15. MTPG variation with channel length

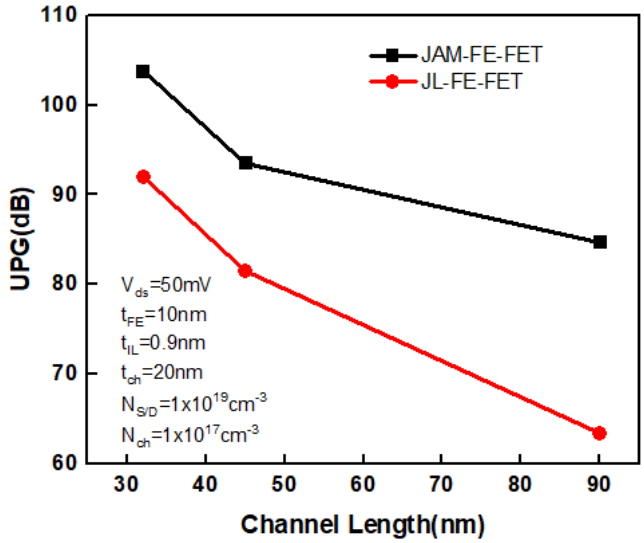


Fig. 16. UPG variation with channel length

Variation in Unilateral Power Gain (UPG) for JAM-FE- FET and JL-FE-FET with respect to V_{gs} is depicted in Fig. 16. When the gate bias is applied, the structure generates a larger electric field in addition to enhanced capacitance, resulting in increased electron velocity and therefore superior saturation velocity. As a result, it's reasonable to conclude that JAM-FE-FET has a greater UPG. It can also be inferred from the figure that as the channel length is increased from 45 nm to 90 nm there is a gradual decline in the unilateral power gain. Overall, the JAM-FE-FET possess high UPG.

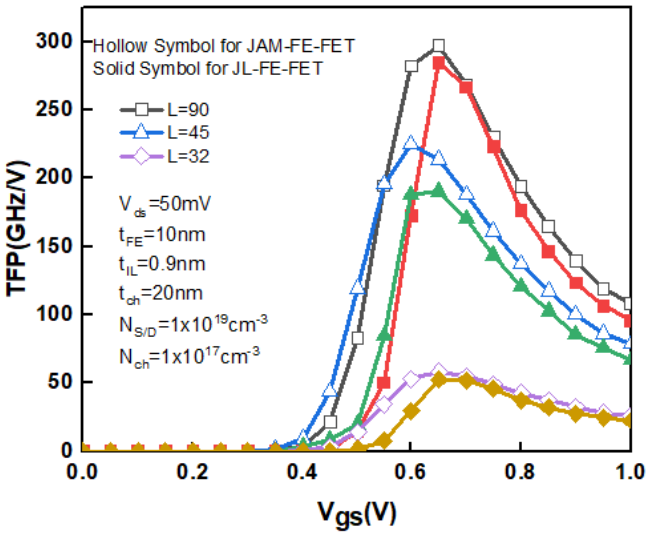


Fig. 17. TFP variation with V_{gs}

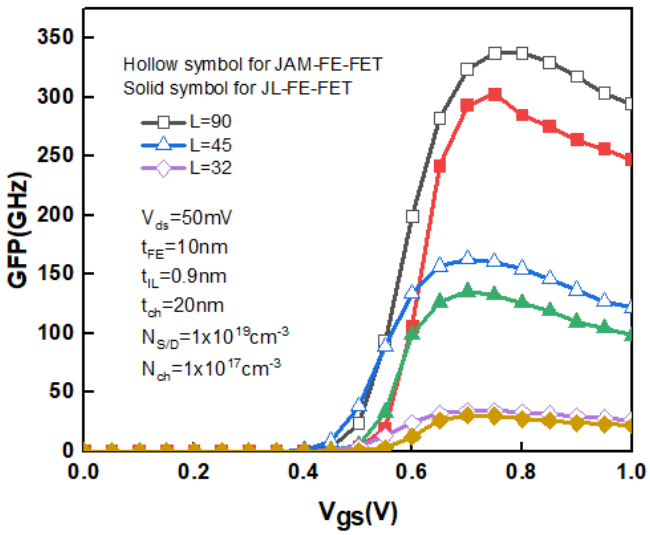


Fig. 18. GFP variation with V_{gs}

The total oxide capacitance also explains these trends. A better capacitance matching in Negative Capacitance FET helps in achieving a high oxide capacitance through which a higher TFP can be obtained. The Gain Frequency Product (GFP) is another performance parameter for high frequency operational amplifier applications. TFP and GFP are expressed as follows:

$$TFP = \frac{g_m f_T}{I_D} \tag{8}$$

$$GFP = \frac{g_m}{g_d} \times f_T \tag{9}$$

GFP increases with V_{gs} due to increment in f_T and intrinsic gain. The curves begin to decline owing to the charge carriers' saturation mobility, which is what causes parasitic capacitances to exist.

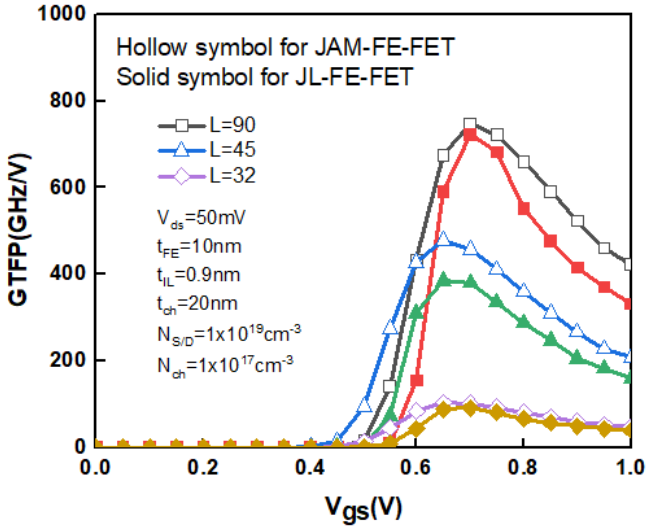


Fig. 19. GTFP variation with V_{gs}

Table 4. Performance parameters at channel length L = 32 nm

Parameter	JAM-FE-FET	JL-FE-FET	%Improvement
I_d(μA)	235.21	221.11	6.3
g_m(μS)	603.52	564.98	6.82
g_d(μS)	4700	4420	6.33
I_{on}/I_{off}(×10^6)	0.357	0.279	27.95
SS (mV/dec)	54	65	20.37
TGF(V^{-1})	4.54	3.27	38.83
V_{EA}(V)	6.07	5.80	4.65
Av (dB)	1.82	1.73	5.20
C_{gg}(fF)	0.83	0.82	1.21
f_T (GHz)	124	116	6.89
MTPG (dB)	105.63	103.78	1.78
UPG (dB)	103.79	91.97	12.85
TFP(GHz/V)	58.50	52.52	11.38
GFP(GHz)	34.45	30.31	13.65
GTFP(GHz/V)	101.8	90.88	12.01

Fig.17 and Fig.18 shows the comparison of TFP and GFP of both the device configuration. It can be observed that JAM-FE-FET possess higher values for both parameters. The RF parameter which determines the entire performance of the device is Gain Transconductance Frequency Product (GTFP). It is expressed as

$$GTFP = \frac{g_m}{g_d} \times \frac{g_m}{I_D} \times f_T \tag{10}$$

Higher the GTFP value, better is the device performance. So, from the Fig.19 it is clear that the proposed device achieves the high GTFP value which indicates better performance. This enhancement is due to the electric field reduction in JAM-FE-FET owing to its gate stack architecture resulting in an improvement of the parameters. Table 4. Gives the Performance parameters at channel length L = 32 nm

5. Conclusion

In this research article, a simulation-based comparative analysis is done for proposed JAM-FE-FET with JL-FE-FET. The SILVACO ATLAS 3D simulator was used to test the

performance of proposed device in terms of analog and RF characteristics. In terms of switching ratio, the suggested JAM-FE-FET device achieves the best results. In comparison to JL-FE-FET, the subthreshold swing is also lowered by 20.37%. The proposed device's SS improves as a result of the negative capacitance effect in the MFIS structure. The proposed device also shows an improvement in terms of transconductance and TGF by 6.82% and 38.83% respectively. The I_{on}/I_{off} ratio also shows a significant improvement of 27.95% owing to the device architecture which overcomes the mobility degradation in the channel region. Also, the capacitive behavior of the proposed device is in close agreement with the JL-FE-FET. However, the enhanced behavior of transconductance results in high cut-off frequency which is a critical parameter in RF applications. Various RF parameters like f_T, TFP, GFP and GTFP are enhanced by 6.89%, 11.38%, 13.65%, and 12.01% respectively. Therefore, from the above obtained results the proposed JAM-FE-FET can be viewed as a promising device for applications in high frequency systems.

6.ACKNOWLEDGMENT

Not Applicable

7. FUNDING STATEMENT

Not Applicable

8. CONFLICT OF INTEREST

There are no conflicts of interest amongst the authors

9. AUTHOR CONTRIBUTIONS

The authors have contributed mutually regarding this paper.

10. AVAILABILITY OF DATA AND MATERIAL

Not Applicable

11. COMPLIANCE WITH ETHICAL STANDARDS

Not Applicable

12. CONSENT TO PARTICIPATE

All the authors have complete consent to participate.

13. CONSENT FOR PUBLICATION

All the authors have complete consent for publication.

14. Ethics approval

All the ethics have been followed

15. Compliance with Ethical Standards section

* Disclosure of potential conflicts of interest

There are no conflicts of interest amongst the authors

* Research involving Human Participants and/or Animals

No animals or human beings were harmed during this research

* Informed consent

All the authors have Informed consent

References:

[1] IRDS Systems and Architectures Team, "International Roadmap for Devices and Systems: Executive Summary," *IEEE Adv. Technol. Humanit.*, p. 63, 2020, [Online]. Available: https://irds.ieee.org/images/files/pdf/2020/2020IRDS_ES.pdf.

[2] C. Hong, J. Zhou, J. Huang, R. Wang, W. Bai, and Y. Chen, "A General and Transformable Model Platform for Emerging Multi-Gate MOSFETs," *IEEE Electron Device Lett.*, vol. 38, no. 8, pp. 1015–1018, 2017, doi: 10.1109/LED.2017.2722227.

[3] J. E. LILIENFELD, "METHOD AND APPARATUS FOR CONTROLLING ELECTRIC CURRENTS," *Patent*, 1926.

[4] R. T. Doria *et al.*, "Analog Operation and Harmonic Distortion Temperature Dependence of nMOS Junctionless Transistors," *ECS Trans.*, vol. 31, no. 1, pp. 13–20, 2010, doi: 10.1149/1.3474137.

[5] N. Gupta, J. B. Patel, and A. K. Raghav, "A study of conventional and junctionless MOSFET using TCAD simulations," in *International Conference on Advanced Computing and Communication Technologies, ACCT*, 2015, vol. 2015-April, pp. 53–56, doi: 10.1109/ACCT.2015.51.

[6] P. Kumar, M. Vashishath, and N. Gupta, "Analytical Modeling and Performance Analysis of Surface Potential for Junctionless MOSFET," 2021, doi: 10.1088/1742-6596/1916/1/012026.

[7] A. K. Raghav, A. Sharma, R. Gupta, and N. Gupta, "Subthreshold modeling of dual-halo dual-dielectric triple-material surrounding-gate (DH-DD-TM-SG) MOSFET for improved leakages," *J. Eng. Res.*, vol. 8, no. 2, pp. 178–190, 2020.

[8] P. Kumar, N. Gupta, N. Sachdeva, T. Sachdeva, and M. Vashishath, "Performance Investigation of Dual-Halo Dual-Dielectric Triple Material Surrounding Gate MOSFET with High-κ dielectrics for Low Power Applications," *J. Semicond. Technol. Sci.*, vol. 20, no. 3, 2020, doi: 10.5573/JSTS.2020.20.3.297.

[9] N. Gupta and P. Kumar, "Elicitation of scattering parameters of dual-halo dual-dielectric triple-material surrounding gate (DH-DD-TM-SG) MOSFET for microwave frequency applications," *Adv. Electr. Electron. Eng.*, vol. 19, no. 1, pp. 66–73, 2021, doi: 10.15598/aeee.v19i1.3788.

[10] P. Kumar, M. Vashisht, N. Gupta, and R. Gupta, "Subthreshold Current Modeling of Stacked Dielectric Triple Material Cylindrical Gate All Around (SD-TM-CGAA) Junctionless MOSFET for Low Power Applications," *Silicon*, 2021, doi: 10.1007/s12633-021-01399-4.

[11] T. Kyun Kim *et al.*, "First Demonstration of Junctionless Accumulation-Mode Bulk FinFETs With Robust Junction Isolation," *IEEE ELECTRON DEVICE Lett.*, 2013, doi: 10.1109/LED.2013.2283291.

[12] J. Hun Choi *et al.*, "Origin of Device Performance Enhancement of Junctionless Accumulation-Mode (JAM) Bulk FinFETs With High-κ Gate Spacers," *IEEE ELECTRON DEVICE Lett.*, vol. 35, no. 12, 2014, doi: 10.1109/LED.2014.2364093.

[13] S. Borkar, "DESIGN CHALLENGES OF TECHNOLOGY SCALING," *IEEE Micro*, vol. 19, no. 4, pp. 23–29, 1999.

[14] R. K. Cavin, V. V Zhirnov, J. A. Hutchby, and G. I. Bourianoff, "Energy Barriers, Demons, and Minimum Energy Operation of Electronic Devices," *Proc. SPIE*, vol. 5844, pp. 1–9, 2005, doi: 10.1117/12.613118.

[15] A. I. Khan, C. W. Yeung, C. Hu, and S. Salahuddin, "Ferroelectric Negative Capacitance MOSFET: Capacitance Tuning & Antiferroelectric Operation," in *International Electron Devices Meeting*, 2011, pp. 5–7.

[16] A. Rusu, "Metal-Ferroelectric-Metal Oxide-Semiconductor Field Effect Transistor with Sub-60mV/decade Subthreshold Swing and Internal Voltage Amplification," 2010.

[17] A. Vinod, P. Kumar, and B. Bhowmick, "Impact of ferroelectric on the electrical characteristics of silicon–germanium based heterojunction Schottky barrier FET," *AEU - Int. J. Electron. Commun.*, vol. 107, pp. 257–263, 2019, doi: 10.1016/j.aeue.2019.05.030.

[18] Y. G. Xiao *et al.*, "Use of negative capacitance to simulate the electrical characteristics in double-gate ferroelectric field-effect transistors," *Curr. Appl. Phys.*, vol. 12, pp. 1591–1595, 2012, doi: 10.1016/j.cap.2012.05.008.

[19] J. J. Zhang, J. Sun, J. Zheng, and Y. Kuk, "A model for the C-V characteristics of the metal-ferroelectric-insulator-semiconductor structure," *Solid State Electron.*, vol. 53, no. 2, pp. 170–175, 2009, doi: 10.1016/j.sse.2008.10.012.

[20] S. Huang, X. Zhong, Y. Zhang, Q. Tan, J. Wang, and Y. Zhou, "A Retention Model for Ferroelectric-Gate Field-Effect Transistor," *IEEE Trans. Electron Devices*, vol. 58, no. 10, 2011, doi: 10.1109/TED.2011.2162629.

[21] T. S. Böscke, J. Müller, D. Bräuhaus, U. Schröder, and U. Böttger, "Ferroelectricity in hafnium oxide thin films," *Cit. Appl. Phys. Lett*, vol. 99, p. 102903, 2011, doi: 10.1063/1.3634052.

[22] S. Dünkel *et al.*, "A FeFET Based Super-Low-Power Ultra-Fast Embedded NVM Technology for 22nm FDSOI and Beyond," 2017.

[23] E. Yurchuk *et al.*, "A 28nm HKMG super low power embedded NVM technology based on ferroelectric FETs," *IEEE Trans. Electron Devices*, vol. 63, no. 9, pp. 3501–3507, 2016, doi: 10.1109/TED.2016.2588439.

[24] T. Mikolajick *et al.*, "Next Generation Ferroelectric Memories enabled by Hafnium Oxide," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, vol. 2019-Decem, pp. 354–357, 2019, doi: 10.1109/IEDM19573.2019.8993447.

[25] M. Hoffmann, U. Schroeder, and T. Schenk, "Stabilizing the ferroelectric phase in doped hafnium oxide," *J. Appl. Phys*, vol. 118, p. 72006, 2015, doi: 10.1063/1.4927805.

[26] U. Schroeder, Ekaterina Yurchuk1, and Johannes Müller, "Impact of different dopants on the switching properties of ferroelectric hafnium oxide," *Jpn. J. Appl. Phys.*, 2014, doi: 10.1088/0169-5983/46/4/041001.

[27] J. Y. Kim, M. J. Choi, and H. W. Jang, "Ferroelectric field effect transistors: Progress and perspective," *APL Mater.*, vol. 9, no. 2, 2021, doi: 10.1063/5.0035515.

[28] D. H. Min, S. Y. Kang, S. E. Moon, and S. M. Yoon, "Impact of thickness control of Hf0.5Zr0.5O2 Films for the metal-ferroelectric-insulator-semiconductor capacitors," *IEEE Electron Device Lett.*, vol. 40, no. 7, pp. 1032–1035, 2019, doi: 10.1109/LED.2019.2917032.

[29] T. Ali *et al.*, "Silicon doped hafnium oxide (HSO) and hafnium zirconium oxide (HZO) based FeFET: A material relation to device physics," *Cit. Appl. Phys. Lett*, vol. 112, p. 222903, 2018, doi: 10.1063/1.5029324.

[30] S. Salahuddin and S. Datta, "Use of negative capacitance to provide voltage amplification for low power nanoscale devices," *Nano Lett.*, vol. 8, no. 2, pp. 405–410, 2008, doi: 10.1021/nl071804g.

[31] D. J. R. Appleby *et al.*, "Experimental Observation of Negative Capacitance in Ferroelectrics at Room Temperature," *Nano Lett.*, vol. 14, no. 7, 2014, doi: 10.1021/nl5017255.

[32] J. Jo and C. Shin, "Negative capacitance field effect transistor with hysteresis-free sub-60-mV/decade switching," *IEEE Electron Device Lett.*, vol. 37, no. 3, pp. 245–248, 2016, doi: 10.1109/LED.2016.2523681.

[33] S. Mehrotra and S. Qureshi, "Analog/RF performance of thin (~10 nm) HfO2 ferroelectric FDSOI NCFET at 20 nm gate length," *2018 IEEE SOI-3D-Subthreshold Microelectron. Technol. Unified Conf. S3S 2018*, pp. 1–3, 2019, doi: 10.1109/S3S.2018.8640153.

[34] Y. Li, Y. Kang, and X. Gong, "Evaluation of Negative Capacitance Ferroelectric MOSFET for Analog Circuit Applications," *IEEE Trans. Electron Devices*, vol. 64, no. 10, pp. 4317–4321, 2017, doi: 10.1109/TED.2017.2734279.

[35] Y. Liang, X. Li, S. K. Gupta, S. Datta, and V. Narayanan, "Analysis of DIBL Effect and Negative Resistance Performance for NCFET Based on a Compact SPICE Model," *IEEE Trans. Electron Devices*, vol. 65, no. 12, pp. 5525–5529, 2018, doi: 10.1109/TED.2018.2875661.

[36] R. Singh *et al.*, "Evaluation of 10-nm Bulk FinFET RF Performance - Conventional Versus NC-FinFET," *IEEE Electron Device Lett.*, vol. 39, no. 8, pp. 1246–1249, 2018, doi: 10.1109/LED.2018.2846026.

[37] A. I. Khan, U. Radhakrishna, K. Chatterjee, and D. Antoniadis, "Negative Capacitance Behavior in a

Leaky Ferroelectric," *IEEE Trans. Electron Devices*, vol. 63, no. 11, pp. 4416–4422, 2016, doi: 10.1109/TED.2016.2612656.

[38] K. S. Li *et al.*, "Negative-Capacitance FinFET Inverter, Ring Oscillator, SRAM Cell, and Ft," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, vol. 2018-Decem, pp. 31.7.1-31.7.4, 2019, doi: 10.1109/IEDM.2018.8614521.

[39] Y. C. Lu and V. P. H. Hu, "Evaluation of analog circuit performance for ferroelectric SOI MOSFETs considering interface trap charges and gate length variations," *Silicon Nanoelectron. Work. SNW 2019*, vol. 8, pp. 9–10, 2019, doi: 10.23919/SNW.2019.8782942.

[40] N. Pandey and Y. S. Chauhan, "Analytical Modeling of Short-Channel Effects in MFIS Negative-Capacitance FET Including Quantum Confinement Effects," *IEEE Trans. Electron Devices*, vol. 67, no. 11, pp. 4757–4764, 2020.

[41] D. S. Software, "Atlas User Manual," no. 408, pp. 567–1000, 2019.

[42] D. Kwon *et al.*, "Negative Capacitance FET with 1.8-nm-Thick Zr-Doped HfO2 Oxide," *IEEE Electron Device Lett.*, vol. 40, no. 6, pp. 993–996, 2019, doi: 10.1109/LED.2019.2912413.

[43] Y. Tsividis and C. McAndrew, "Operation and Modeling of the Mos Transistor.Pdf." p. 750, 2011.

[44] M. Kobayashi and T. Hiramoto, "On device design for steep-slope negative-capacitance field-effect-transistor operating at sub-0.2V supply voltage with ferroelectric HfO2 thin film," *Cit. AIP Adv.*, vol. 6, p. 25113, 2016, doi: 10.1063/1.4942427.

[45] K. P. Pradhan, S. K. Mohapatra, P. K. Sahu, and D. K. Behera, "Impact of high-k gate dielectric on analog and RF performance of nanoscale DG-MOSFET," *Microelectronics J.*, vol. 45, pp. 144–151, 2014, doi: 10.1016/j.mejo.2013.11.016.

[46] S. Rewari, V. Nath, S. Haldar, • S S Deswal, and • R S Gupta, "Improved analog and AC performance with increased noise immunity using nanotube junctionless field effect transistor (NJLFET)," *Appl. Phys. A*, 2016, doi: 10.1007/s00339-016-0583-9.

Ain Shams Engineering Journal xxx (xxxx) xxx

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Landslide hazard assessment using analytic hierarchy process (AHP): A case study of National Highway 5 in India

Sandeep Panchal *, Amit Kr. Shrivastava

Delhi Technological University, Main Bawana Road, Samaypur Badli, Delhi, India

ARTICLE INFO

Article history:
Received 21 July 2021
Revised 14 October 2021
Accepted 25 October 2021
Available online xxxx

Keywords:
Landslide hazard zonation
Weighted linear combination (WLC)
Analytic hierarchy process (AHP)
Landslide susceptibility

ABSTRACT

Slope failure along highways is a crucial problem in hilly regions. Landslide hazard maps are very efficient and effective tools for planning and management of landslide disasters. Aim of this study is to prepare a landslide hazard map along national highway 5 (197.600–283.200 Km) using analytic hierarchy process (AHP) model. The different causative factors of landslides considered in this study are slope, aspect, curvature, relative relief, fault density, drainage density, geology, topographic wetness index (TWI), distance from road and lithology. The causative factors are divided into sub-factors and weightage are assigned according to analytic hierarchy process (AHP). The causative factor layers are overlaid using weighted linear combination (WLC) technique and a landslide hazard map is prepared. A landslide inventory of 215 landslides is used for validation of the landslide hazard map. The map shows a prediction rate of 0.825 on area under curve (AUC) technique. The study can be used by the construction planners and decision makers.

© 2021 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Ain Shams University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Road network plays an important role in socio-economic development of a nation. The roads in hilly regions are generally subjected to heavy landslides and slope failure events. Landslides are the most destructive phenomenon that occur in hilly regions frequently. The failure of slopes along highways results into heavy traffic jamming and causes inconvenience to the passengers. Landslides are responsible for economic losses and sometimes cause the loss of lives too.

The construction of roads in the hilly region accelerates the failure of slopes due to cutting works [47,1]. The anthropogenic activities during highway construction disturbs the natural slopes and make them more vulnerable to the failures. So, the landslide hazard assessment along highways is essential for proper planning and management of road infrastructure [47].

Landslide hazard maps are important tools in planning and mitigation of landslide disaster. A landslide hazard map shows the possibility of occurrence of landslide in a particular region. Landslide hazard assessment studies require good understanding of the causative factors of landslides [5,47]. The researchers have considered slope, aspect, curvature, lithology, geology and drainage characteristics of the region as causative factors of landslides [8,7]. The triggering factors like earthquake, rainfall, erosion etc. accelerates the frequency and speed of the landslide disaster [4,51]. Geographic information system (GIS) helps in managing the spatial and temporal data effectively. The main advantage of GIS is that it gives utility of changing the input so that output can be varied. Landslide hazard zonation and landslide susceptibility mapping in GIS environment are performed by the various researchers [15,9,12,47].

1.1. Literature review

There are three approaches used in landslide hazard mapping in GIS environment i.e. qualitative approach, quantitative approach and semi-quantitative approach [48]. Qualitative approach gives a descriptive solution to landslide susceptibility mapping while quantitative approaches consider the mathematical estimates

* Corresponding author at: Department of Civil Engineering, Delhi Technological University, Main Bawana Road, Samaypur Badli, Delhi, India.
E-mail addresses: 2290sandy@gmail.com (S. Panchal), aksrivastava@dce.ac.in (A.Kr. Shrivastava).

Peer review under responsibility of Ain Shams University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.asej.2021.10.021
2090-4479/© 2021 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Ain Shams University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: S. Panchal and Amit Kr. Shrivastava, Landslide hazard assessment using analytic hierarchy process (AHP): A case study of National Highway 5 in India, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2021.10.021

[31,42]. The expert based methods for landslide hazard assessment falls in qualitative approach category [3]. The opinion of expert is considered to assign the relative weightage to the different causative factors of landslides. The factors which are responsible for occurrence of landslides are selected and their impact on landslide occurrence is evaluated based on the expertise of expert. The result in the expert based method depends upon the subjective judgment of the expert which may not be right always.

The quantitative approach involves in establishing a mathematical relationship between occurrence of landslide and causative factors [2]. The impact of causative factors and sub-factors is simulated using mathematical models like frequency ratio [36,33,41], logistic regression [36,33,34], bivariate [45,20,37] and multivariate statistical models [36,43,2,47]. While the accuracy of the quantitative methods is higher as compared to the qualitative techniques but it depends upon the availability of an accurate landslide inventory. The semi-quantitative or semi-qualitative approaches are the hybrid of both the approaches. The methods like analytic hierarchy process (AHP), analytic network process (ANP), fuzzy-AHP etc. are the semi-qualitative or semi-quantitative approaches [17,6].

Analytic hierarchy process (AHP) helps in breaking a complex problem in the different simple criteria and the criteria are given weightage according to their relative importance [49,50,38,40]. A complex problem is the one which depends upon multiple causative factors. In our study, landslide is the complex natural problem which depends upon the multiple causative factors like slope, aspect, curvature, drainage characteristics, geological and lithological characteristics of the study area.

AHP has been used for landslide hazard mapping by many researchers [18,14,27,21,16]. AHP helps in finding a consistent solution for the problems [49,50]. Table 1 shows the findings of the literature review and Table 2 shows the source of input data used in this study.

In this study, landslide hazard assessment is performed using analytic hierarchy process (AHP) due to the following reasons:

a. The historical landslide data/landslide inventory is available in point form for the study area.
b. Due to the pandemic, the movement is restricted throughout the study area. So, it is not possible to arrange frequent field visits for preparation of landslide inventory.
c. Analytic hierarchy process (AHP) helps in checking the consistency of weightage given by experts. So, this technique helps in avoiding the inconsistent solutions.
d. The output can be improved by varying the input weightage in analytic hierarchy process (AHP) which helps in improving the accuracy of the output.

The study area selected for this study is National Highway 5 (NH-5) from Kalka, Haryana to Shimla, Himachal Pradesh (H.P.), India. NH-5 is an important highway corridor from the point of view of transfer of goods and passengers. The highway faces frequent landslide events which are triggered in rainy season. The route is damaged due to heavy landslides at different locations.

2. Study area

National Highway 5 is 660.2 KM long and passes through plain and hilly terrains. The highway is very important from strategic and social point of view. It connects the remote areas of the hilly regions. It establishes the connection between India and Tibet which makes it important for international trade too. The stretch selected for the study lies from 197 Km to 283 Km along National Highway 5. The selected stretch of highway connects Haryana province to the capital of Himachal Pradesh (H.P.) province and facilitates the transportation of goods and crops from the plain area to the hilly terrain of high elevation. The highway has been functioning since the British period. The up-gradation to four lanes of Kalka to Shimla stretch is under progress. The highway stretch consists of few tunnels which reduces the distance through hilly terrain.

A buffer of 1 KM at both sides of the road is taken. The study area lies between 30°50′17.88″N, 76°56′8.52″E to 31°06′17.28″N, 77°10′24.24″E. The altitude of the study area varies from 591 m to 2459 m above mean sea level. The temperature of the study area varies from 6.2 °C to 36.7 °C (District survey document, Shimla). Annual rainfall of the Solan district is 1413 mm and average annual rainfall of Shimla district is 1575 mm. Fig. 1 shows the details of the study area.

2.1. Geological setting

The study area lies in the districts of Solan and Shimla in Himachal Pradesh (H.P.). Solan. Dolomitic granite stones covers the region with its rugged topography. The Tethys Tectogen are found in Shimla region which are characterized by the diversity in the stratigraphy. The study area is covered with Miocene, Eocene and Neoproterozoic aged deposits. The region has sandstones, shale,

Table 1
Summary of the previous literature.

Method	Summary	Findings from the literature review
Qualitative techniques	The weightage to the causative factors and sub-factors are given by experts based on their judgment [29,48,44]. The evaluator selects the causative factors and their relative importance [48]. These techniques are the simplest to implement as no historical landslide data is required.	The accuracy of the susceptibility and hazard map depends upon the subjective judgment of the expert [26,48]. The hazard maps can vary from expert to expert [48]. Generally, accuracy of the expert based methods is lesser as compared to quantitative techniques [30].
Quantitative techniques	The quantitative techniques establish a mathematical relationship between occurrence of landslides and the causative factors [2]. Frequency ratio, info value, weight of evidence and logistic regression are the examples of quantitative methods [42]. These methods depend on the spatial distribution of landslides and their relationship with causative factors [48,29].	These methods can be applied to the large areas [48]. These techniques give more accurate results as compared to the qualitative techniques [29]. The weightage can be derived statistically from the distribution of landslides. A well distributed landslide inventory is required for calculation of weightage of causative factors [42,48]. The preparation of landslide inventory requires large efforts. The methods are not suitable for inaccessible regions.
Semi-quantitative techniques	The semi-quantitative techniques are a hybrid of qualitative and quantitative techniques. The human perception is considered in these studies and validated mathematically [30]. Analytic hierarchy process, fuzzy logic etc. are the examples of this technique.	The human perception is included in these methods. The consistency of the causative factors can be checked and validated using mathematical relationship [48]. While there is no need of landslide inventory but the use of landslide inventory can improve the consistency and accuracy of the results [30].

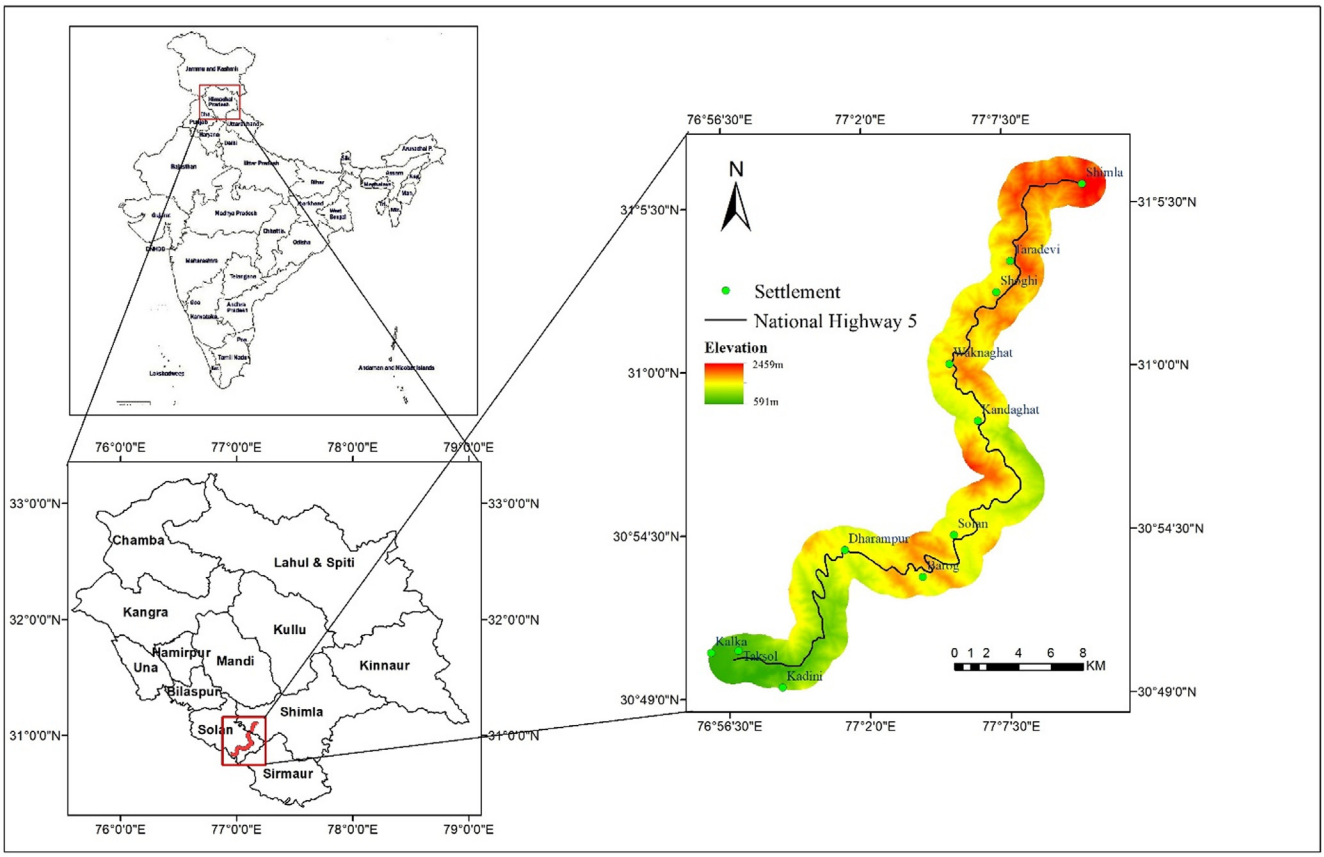


Fig. 1. Study area.

Table 2
Data Used in the Study.

Extracted Data	Scale	Source
Landslide Points		Literature, Field Survey, GSI Practical Sheet, Google Earth
Aspect		Digital Elevation Model (DEM)
Curvature		Digital Elevation Model (DEM)
Relative Relief		Digital Elevation Model (DEM)
Fault Density	1:50,000	Ground Water Prospects Map
Geology	1:500,000	Geology Map of Himachal Pradesh
Lithology	1:50,000	Ground Water Prospects Map
Drainage Density	1:50,000	Survey of India Toposheet
Road	1:250,000	U.S. Army Corps of Engineers

limestone, dolomite, quartzite, phyllite and siltstone. The major deposits of limestone are present in Shimla region. The lithology of the study area belongs to Jutogh group, Rampur group, Shali group and Shimla group [22].

2.2. Landslide inventory

Landslide inventory shows the distribution of the landslide data [13,19]. The landslide inventory in this study is prepared using historical data, google earth and field survey. The landslide information is obtained from the literature, Survey of India sheets and inventories prepared by National Remote Sensing Centre (NRSC). The landslide inventory consists of the information about location of landslide, triggering factor, activity of landslide, type, classification and geology of the failed mass. Landslides are also detected from the google earth from LISS imagery. Landslides are presented in the form of points. A total number of 214 landslides are considered in the study. Fig. 2 shows the images of few landslides occurred in the area. Fig. 2 shows the images of few landslides occurred in the study area.

3. Materials and methods

The data required for this study is obtained from the satellite imagery, Survey of India (SOI) toposheets, field surveys and GSI practical sheets. The digital elevation model (DEM) is obtained from Advanced Spaceborn Thermal Emission and Reflection Radiometer (ASTER). The digital elevation model has a grid size of 30*30 m. The data related to slope, aspect, curvature and relief is extracted from ASTER DEM. The geology and lithology data is digitized from SOI toposheets. A detailed flow chart is shown in the Fig. 3.

3.1. Causative factors of landslides

Occurrence of the landslides depends upon the various factors. The factors considered in this study are divided into three categories i.e. DEM based factors, drainage and lineament characteristics and geological characteristics of the region. Fig. 4 shows the details of causative factors.

3.1.1. Slope gradient

Slope gradient is one of the most important factors which affect the occurrence of the landslides. It directly impacts the stability of slopes as it influences the shear force [23]. Steeper the slope more will be the probability of occurrence of the landslides [11]. The slope gradient in the region varies from gentle to very steep. In



Fig. 2. Landslide occurred in the study area [Landslide 1: Long: 77°9′6.893″E, Lat: 31°2′55.57″N, Landslide 2: Long:77°6′44.045″E, Lat:30°58′45.342″N, Landslide 3: Long: 77°4′42.711″E, Lat: 30°53′23.058″N, Landslide 4: Long: 76°56′58.685″E, Lat: 30°51′1.586″N].

57% cases the slope is more than 60°. The instability in the mass is induced due to the steep angle of the slope. Slope gradient also controls the flow of water [35]. The slope has been divided into five categories.

3.1.2. Slope aspect

Aspect represents the direction of the slope. Aspect affects the exposure of sun, erosion characteristics and flow of water [47]. Aspect map is divided into 10 categories.

3.1.3. Curvature

Curvature is defined as slope of slope. It is the second derivative of the surface. The positive value of curvature indicates the convexity of the surface while the negative value shows the concavity of the surface. 0 value of curvature indicates the flat surface. In landslide hazard zonation, curvature indirectly represents the effect of water [28].

3.1.4. Relative relief

Relative relief of the study area shows the variation of the elevation. It affects the natural conditions which affect the occurrence of the landslides [47]. The relative relief varies from 0 m to 254 m.

3.1.5. Geology

The study area has diverse geological characteristics. The region is covered with lower Shivalik formations, Subathu formation, Jaunsar group, Jutogh group, Dharmashala group and Balani formations. Mix of Muree group, Dharmashala group, Dagshai and Kasauli formations are more prone towards landslide incidents while Jaunsar group is relatively less affected by landslides in this study area.

3.1.6. Lithology

Lithology identifies the type of rock in the study area. The erosion and stability of the rocks depend upon the lithology of the area. The rocks are of mixed type in the study area. The study area consists of shale, sandstone, siltstone, phyllite and dolomite in mixed form.

3.1.7. Fault density

Natural faults occur in the study area. The effect of faults is considered in the form of fault density. The faults increase the landslide activities [32].

3.1.8. Drainage density

Drainage characteristics of the region affect the landslide possibility. Drainage density is obtained by digitizing the streams.

3.1.9. Distance from the road

Distance from the road is also an important parameter for occurrence of landslides. During the construction of roads, the cutting activities disturb the natural slopes [46]. The slope near the toe become weak along the highways.

3.1.10. Topographic wetness index (TWI)

Topographic wetness index (TWI) represents the moisture in the soil. The TWI represents the tendency of water accumulation [24]. TWI is calculated using the following formula:

$$Topographic\ \ Wetness\ \ Index\ \ (TWI) = \ln\left(\frac{SCA}{tan\varphi}\right) \quad (1)$$

Here, SCA is specific catchment area and φ is the slope angle. SCA is defined as flow accumulation area per unit contour width. Assuming the properties of soil or rock mass uniform, SCA shows the tendency to receive the water while slope angle shows the tendency of draining of the water [24].

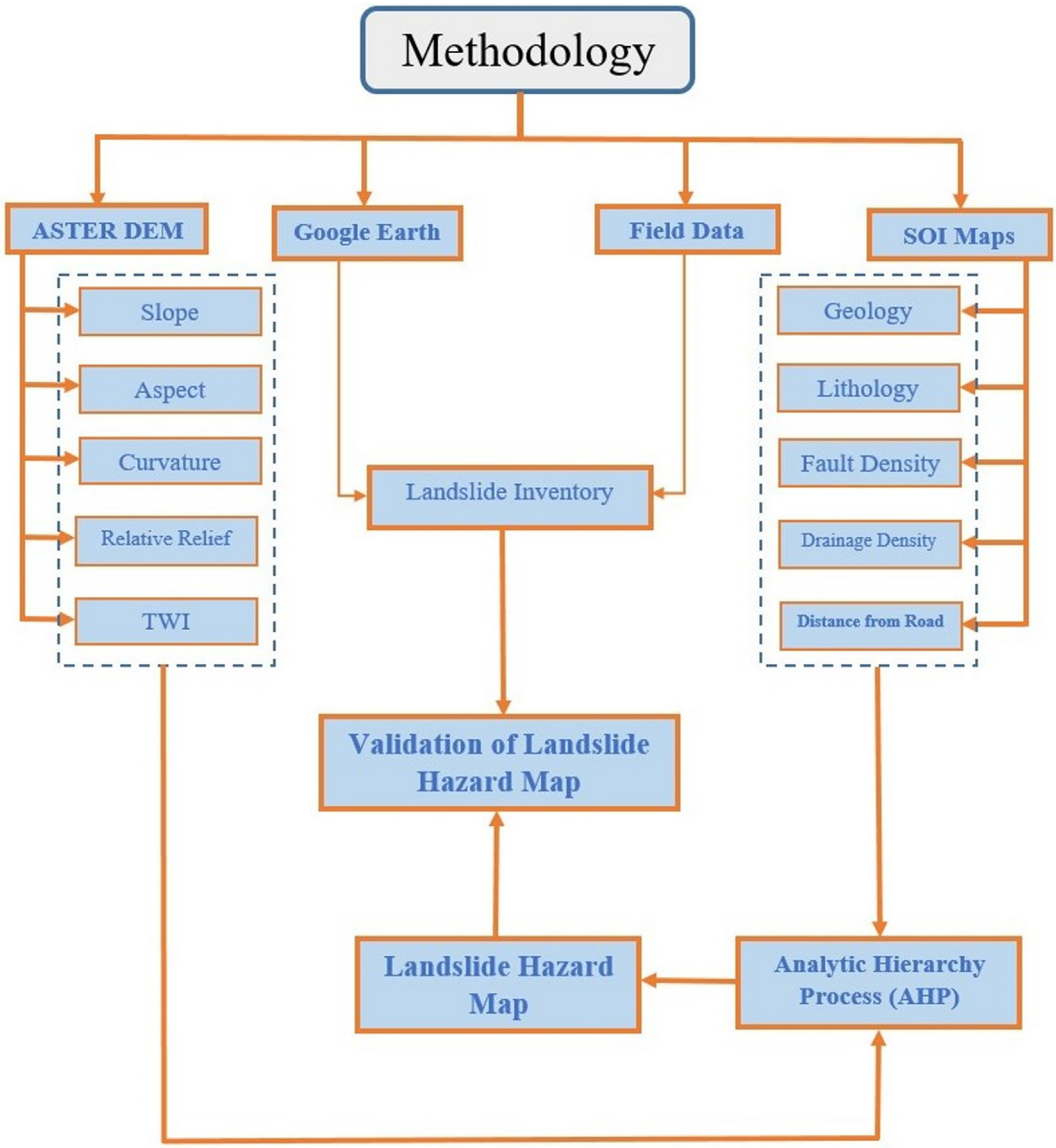


Fig. 3. Flow of the work.

3.2. Analytic hierarchy process (AHP)

Analytic hierarchy process (AHP) is a decision making tool which helps in breaking the complex problem in simple criteria. AHP is based upon three principles i.e. decomposition of problem, comparative judgment and synthesis of relative importance or rankings [49,50,39]. In AHP, the problem is broken in hierarchical criteria. These criteria are compared to each other. This process of relative comparison is called pair-wise comparison. Eigen vector method is used to calculate the rankings and after that consistency of the solution is also checked by using consistency ratio [49,50]. Table 3 shows the scale of pairwise comparison given by Saaty.

The consistency of the weights for relative importance assigned during the pairwise comparison can be checked using the equation given below.

$$Consistency Ratio(CR) = \frac{CI}{RI} \quad (2)$$

Here, CI is consistency index while RI is randomness index.

CI is calculated as follows:

$$consistency\ index(CI) = \frac{\lambda \max - n}{n-1} \quad (3)$$

Here, λmax = Major Eigen value and n = order of matrix

Randomness index values are given by Saaty which depends on the value of n. RI is the result of extensive experimentation on the large sample of dataset. Table 4 shows the randomness index (RI) for different values of n. If CR values are less than 10%, the pairwise comparison is considered as consistent. If the CR value is more than 10%, the solution is considered inconsistent and weights are reassigned in pairwise comparison matrix. The weightage of causative factors and sub-factors are shown in Table 5.

4. Results and discussion

Analytic hierarchy process (AHP) is implemented for landslide susceptibility mapping along highway corridors in GIS environment. A landslide inventory of 215 landslides is prepared from Google earth, previous literature and GSI practical sheet. 70% of the landslides from the inventory are randomly selected as training data while the rest 30% are used for testing. The relative comparison of the causative factors is done based on the landslide inventory. The weightage of the factors and sub-factors is calculated using AHP and the results are checked for consistency. The consistency ratio is kept below 10%. The consistency ratio is obtained

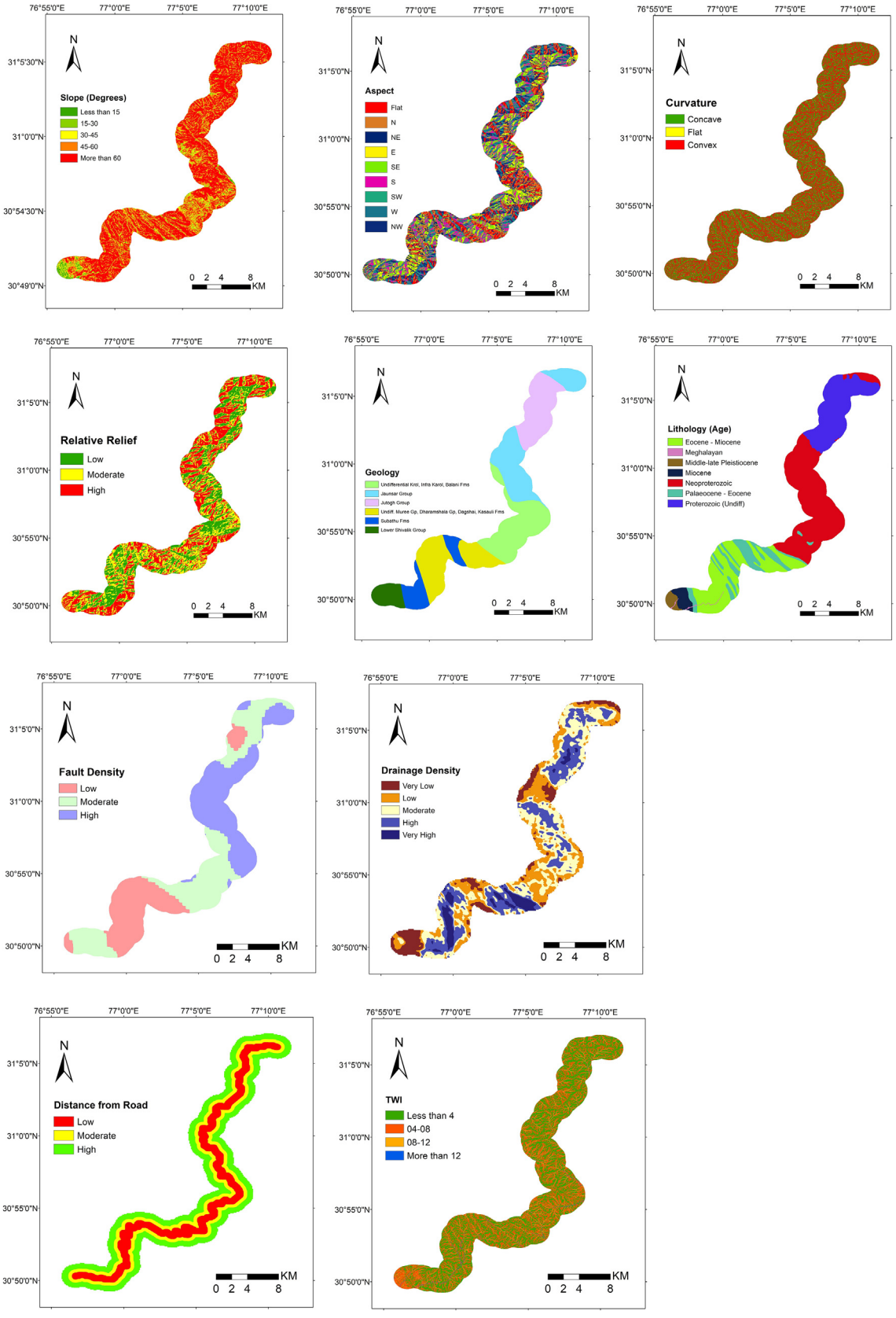


Fig. 4. Causative factors of landslides a) Slope b) Aspect c) Curvature d) Relative relief e) Geology f) Lithology g) Fault density h) Drainage density i) Distance from the road k) Topographic wetness index (TWI).

Table 3
Scale of Analytic Hierarchy Process (AHP) [49].

Degree of Preference [49]	Definition	Explanation
1	Equally Important	Both criteria are equally important or both the factors have same effect on occurrence of landslides
3	Moderately Important	One factor is more effective as compared to the other factor
5	Highly Important	One factor affects highly as compared to the other factor
7	Very Highly Important	A factor is highly dominated over other
9	Extremely Important	A factor has highest possibility of affecting the occurrence of landslide over other factor
2,4,6,8	Intermediate Values	If a compromise between two factors is required, intermediate values can be used

using the highest Eigen value and randomness index is obtained from the Table 4. Fig. 4 shows the causative factors of the landslides.

The weightage of individual factors is given in Table 5. Distance from the road had the highest impact on the occurrence of landslides and got the weightage 0.325 in AHP matrix while lithology is the second dominating causative factor with weightage of 0.208. Geology had weightage of 0.114 while aspect is the least important causative factor with AHP weightage of 0.019.

It is observed that the instability in slopes increases as the slope gradient increases. There are no landslides for the slope angles below 30°. Only 20% of the study area consists of slopes with slope angle less than 60°. The slopes which are more than 60° covers around 80% of the study area and 57% of the total landslides occur in regions with slope more than 60°. 15% of the study area is covered with the slopes with slope angle from 45° to 60°, and 28% of the total landslides occur in such regions. No landslide is observed in the area with slopes 0–15°.

Slope aspect represents the direction of slope with respect to magnetic north. It is observed that south-west, west and north-west aspects have more impact on the occurrence of the landslides due to more erosion. 22.85% of the total landslides occur in north-west aspect while south west and west direction face 14.28% and 17.15% of the total landslide event respectively.

Concave curvature has more probability of occurrence of landslides due to impact of water. Flat curvature has no event of landslides. Concave surface faces 54.54% landslide while convex surface faced 45.46% land-sliding events.

Relative relief shows the impact of changes in the vegetation with respect to the elevation. Lower relative relief values in the study area faced more landslide events as compared to the higher relative relief values. The regions with 0–76 m relative relief faced 42.42% of the total landslides while it covers only 30.62% of the total area. The region with highest relative relief i.e. 172–254 m faced only 27.27% landslides while it covers 36.30% of the total area.

Six group of geological formations are encountered in the study area. Undifferential muree group, Dharmashala group, Dagshai and Kasauli formations have highest evidences of landslides in the study area. This geological group covers 20.74% of the total area but 35.53% of the total landslides occurred in this geological formation. Balani formations, un-differential and Infra Krol are also highly susceptible to landslides as this type of formation covers 22.88% area but faced 29.21% of total landslides. Jaunsar group has least susceptibility towards landslides as it faces only 6% of total landslides in the study area.

The lithology of the region is mixed type. The lithological deposits are classified based on their age. Eocene-miocene deposits are highly prone towards landslide events. These deposits faced 42.85% of the total landslides while they cover 22.77% of the total area. Middle late Palaeocene and Meghalayan formation cover very less area and faced only one landslide event. Palaeocene-eocene deposits cover around 10% of the total area and faces around 15% of the total landslides which shows their high proneness towards landslides.

Fault density had a significant impact on the occurrence of landslides. The moderate fault density covers 39.12% of the area but responsible for the occurrence of 50% of the total landslides. The low fault density covers 27% of the total study area while 38.23% of total landslides occurred in these regions.

Drainage density had a linear impact on occurrence of landslides. As the drainage density increases, the proneness towards landslide occurrence increases. 6.83% of the total area is covered with very high drainage density while such type of the region faced 15.83% of the total landslides. There was 24.07% of the total area covered with high drainage density which faced 30.77% of the total landslides. 11.30% of the total area is covered with very low drainage density which faced only 3.84% of the total landslides.

The road considered in this study has been upgraded to four lanes recently. So, landslide events have been increased due to disturbance of natural slopes. Distance from the road is another important causative factor of landslides. The region which has low distance to road covers 30.55% of the total area but faced

Table 4
Randomness Index (R.I.) Table [49].

Number of Criteria	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R.I.	0.0	0.0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.48	1.56	1.57	1.58

Table 5
Weightage of Major Causative Factors.

Factors	Slope	Aspect	Curvature	Relative Relief	Geology	Lithology	Drainage Density	Fault Density	TWI	Road Distance	Weightage
Slope	1										0.058
Aspect	0.2	1									0.019
Curvature	0.5	3	1								0.038
Relative Relief	0.5	2	0.5	1							0.027
Geology	3	5	2	4	1						0.114
Lithology	5	7	5	6	3	1					0.208
Drainage Density	3	5	4	3	0.5	0.25	1				0.095
Fault Density	2	6	2	4	0.5	0.33	0.5	1			0.073
Topographic Wetness Index (TWI)	0.33	2	2	2	0.25	0.2	0.33	1	1		0.042
Distance from Road	5	7	6	7	4	3	7	4	7	1	0.325

Table 6
Pairwise comparison matrix

Causative Factors/Sub-factors	1	2	3	4	5	6	7	8	9	AHP Weightage
Slope Gradient										
0-<15	1									0.033
15-<30	3	1								0.063
30-<45	5	3	1							0.129
45-<60	7	5	3	1						0.262
More than 60	9	7	5	3	1					0.513
CR	0.053									
Aspect										
Flat	1									0.023
North	2	1								0.047
Northeast	7	4	1							0.204
East	4	2	0.33	1						0.082
South East	9	8	3	4	1					0.362
South	5	3	0.5	2	0.25	1				0.125
South West	4	2	0.25	1	0.2	0.5	1			0.078
West	3	0.5	0.25	0.5	0.14	0.33	0.5	1		0.046
North West	2	1	0.14	0.33	0.12	0.25	0.33	0.5	1	0.033
CR	0.025									
Curvature										
Flat	1									0.072
Concave	7	1								0.649
Convex	5	0.33	1							0.279
CR	0.068									
Relative Relief										
0-<76	1									0.637
76-<172	0.33	1								0.258
172–254	0.2	0.33	1							0.105
CR	0.04									
Geology										
Undifferential Krol, Infra Krol, Blani Fms	1									0.043
Jaunsar Group	2	1								0.064
Jutogh Group	4	3	1							0.132
Undiff Muree Group, Dharamshal Group, Dagshai and Kasauli Fms	8	7	5	1						0.474
Subathu Fms	6	5	3	0.33	1					0.256
Lower Shivalik Group	0.5	0.33	0.2	0.11	0.14	1				0.031
CR	0.043									
Lithology										
Eocence-Miocene	1									0.135
Meghalayan	0.12	1								0.01
Middle late Pleistiocene	0.11	0.50	1							0.08
Miocene	0.17	4	4	1						0.028
Neoproterozoic	0.25	6	6	4	1					0.076
Palaeocene-Eocene	2	8	9	6	4	1				0.144
Proterozoic (Undiff)	0.25	4	5	5	0.33	0.25	1			0.066
CR	0.085									
Fault Density										
Low	1									0.072
Moderate	5	1								0.279
High	7	3	1							0.649
CR	0.068									
Drainage Density										
Low	1									0.105
Moderate	3	1								0.258
High	5	3	1							0.637
CR	0.04									
Distance from Road										
Low	1									0.804
Moderate	0.12	1								0.122
High	0.11	0.5	1							0.074
CR	0.039									
Topographic Wetness Index (TWI)										
Less than 4	1									0.671
4-<8	0.2	1								0.178
8-<12	0.14	0.5	1							0.096
More than 12	0.11	0.25	0.5	1						0.055
CR	0.022									

40% of the total landslides while the region which has high distance to road covered 45.23% of the total area but faced only 33.12% of the total landslide events.

Lower values of topographic wetness index (TWI) had more impact on occurrence of landslides while higher values had very less effect on occurrence of landslides. 52.6% of the total area is covered with TWI less than 4 while 38% of the total area is covered with TWI between 4 and 8. The region covered TWI value less than 4 faces 61.76% of total landslides while area covered with TWI (from 4 to 8) faced 35.29% of the total landslides. The area with TWI values more than 12 covers 0.5% of the total area and faced no landslide event. Table 6 shows the details of pairwise comparison matrix of the sub-factors, consistency ratio and weightage of sub-factors.

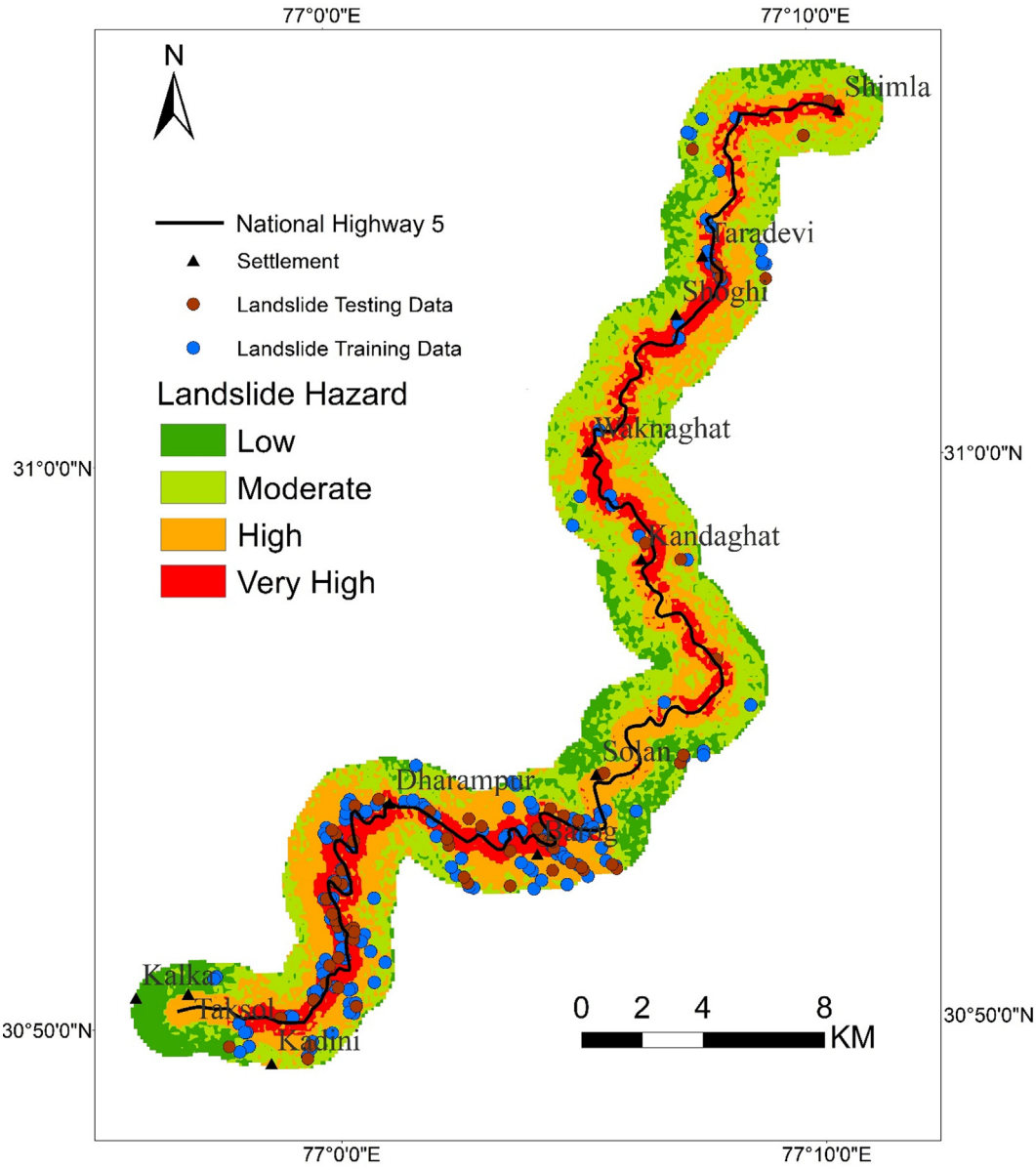


Fig. 5. Landslide Hazard Map.

The landslide hazard index (LHI) is calculated using the following equation:

$$\begin{aligned} Landslide\ \ Hazard\ \ Index\ \ (LHI) = {} & 0.058 \times Slope + 0.019 \\ & \times Aspect + 0.038 \\ & \times Curvature + 0.027 \\ & \times Relative\ \ Relief + 0.114 \\ & \times Geology + 0.208 \\ & \times Lithology + 0.095 \\ & \times Drainage\ \ Density + 0.073 \\ & \times Fault\ \ Density + 0.042 \\ & \times TWI + 0.325 \\ & \times Distance\ \ from\ \ Road \end{aligned} \quad (4)$$

Based on LHI, landslide hazard map for the study area is shown in Fig. 5. The landslide hazard index (LHI) varies from 0.00839 to 0.5211. The landslide hazard map is divided into four categories using natural breaks i.e. low (0.00839–0.1317), moderate (0.1317–0.1961), high (0.1961–0.3969) and very high (0.3969–0.5211). It is found that 21.56% of the total area lies under very high landslide hazard zone which faced 46.51% of the total landslides. 29.05% of the total region lies under high landslide hazard zone which faced 32.56% of the total landslide events. 39.85% of the total study area lies under moderate landslide hazard zone while it faced only 18.60% of total landslide events. 9.52% of the region lies under low landslide hazard zone and only 2.32% of the landslides occurred in this region.

The output of the study is validated with receiver operation characteristics (ROC) curve. The ROC curve is plotted between area of landslide hazard class and area of landslides. The prediction rate for training data is found to be 0.874 while the prediction rate for

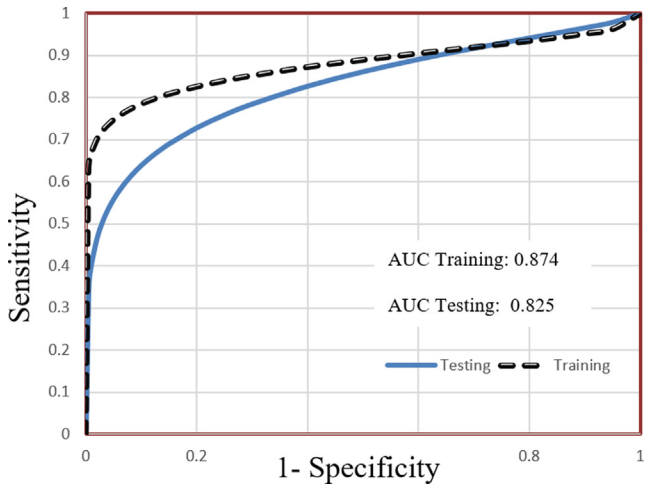


Fig. 6. ROC curve for validation.

testing data is found to be 0.825. Fig. 6 shows the prediction curve for the output.

5. Conclusion

The highway stretch considered in this study is subjected to the landslide disaster frequently. So, landslide hazard assessment is required for planning and execution of construction activities along the highway corridor. Analytic hierarchy process (AHP) has been implemented for preparation of landslide hazard map. There were 10 causative factors which are considered in this study. Distance from road, lithology and geology are the major contributing factors in landslide occurrence according to the weightage assigned using AHP. A landslide inventory containing details of 215 landslides was used to train and test the final output. The landslide inventory had also helped in deciding the pairwise comparison of the different causative factors and sub-factors. 70% of the landslides were used for training and 30% landslide data is used for testing purpose. The landslide hazard map was divided into four zones. It is found that around 79% of landslides lies in high and very high landslide hazard zones. The prediction rate observed for the landslide hazard map is found to be 0.825. The output of the study depends on the causative factors considered and their weightage. The prediction rate can be increased by varying the causative factors and their weightage. The output of the study can be used by the project planners, environmental engineers, construction managers and risk engineers. The landslide hazard map can be used for planning the construction and maintenance operations of the highway.

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] Ali S, Biermans P, Haider R, Reicherter K. Landslide susceptibility mapping by using a geographic information system (GIS) along the China-Pakistan Economic Corridor (Karakoram Highway), Pakistan. Nat Hazards Earth Syst Sci 2019;19:999–1022. doi: https://doi.org/10.5194/nhess-19-999-2019.

[2] Akgun A, Erkan O. Landslide susceptibility mapping by geographical information system-based multivariate statistical and deterministic models: in an artificial reservoir area at Northern Turkey. Arabian J Geosci 2016;9 (2):1–15. doi: https://doi.org/10.1007/s12517-015-2142-7.

[3] Aleotti P, Chowdhury R. Landslide hazard assessment: Summary review and new perspectives. Bull Eng Geol Environ 1999;58(1):21–44. doi: https://doi.org/10.1007/s100640050066.

[4] Altın TB, Gökkaya E. Landslide-Triggering Factors in Korucak Subbasin, North Anatolian, Turkey. Procedia Earth Planet Sci 2015;15:566–72. doi: https://doi.org/10.1016/j.proeps.2015.08.104.

[5] Bozzano F, Mazzanti P, Prestininzi A, Mugnozza GS. Research and development of advance technologies for landslide hazard analysis in Italy. Landslides 2010;7(3):381–5. doi: https://doi.org/10.1007/s10346-010-0208-x.

[6] Chaiyaphan C, Ransikarbum K, Sriariyanun M, Cheng YS, Rattanaporn K, Yasurin P, et al. Criteria Analysis of Food Safety using the Analytic Hierarchy Process (AHP) - A Case study of Thailand's Fresh Markets. E3S Web Conferences 2020;141:02001. doi: https://doi.org/10.1051/e3sconf/202014102001.

[7] Chen W, Xie X, Peng J, Wang J, Duan Z, Hong H. GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models, Geomatics, Natural Hazards Risk 2017;8(2):950-973 10.1080/19475705.2017.1289250

[8] Demir G, Aytekin M, Akgun A. Landslide susceptibility mapping by frequency ratio and logistic regression methods: an example from Niksar-Resadiye (Tokat, Turkey). Arabian J Geosci 2015;8(3):1801–12. doi: https://doi.org/10.1007/s12517-014-1332-z.

[9] Ding Q, Chen W, Hong H. Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping, Geocarto Int 2016;32 (6):619-639, 10.1080/10106049.2016.1165294 District Survey Document Shimla, accessed on 03/05/2021 available on https://emerginghimachal.hp.gov.in/miningstone/survay_docs/shimla.pdf

[11] Donnarumma A, Revellino P, Grelle G, Guadagno FM. Slope angle as indicator parameter of landslide susceptibility in a geologically complex area. In: Margottini C, Canuti P, Sassa K, editors. Landslide Science and Practice. Berlin, Heidelberg: Springer; 2013. p. 425–33. doi: https://doi.org/10.1007/978-3-642-31325-7_56.

[12] Fan W, Wei X, Cao Y, Zheng B. Landslide susceptibility assessment using the certainty factor and analytic hierarchy process. J Mountain Sci 2017;14 (5):906–25. doi: https://doi.org/10.1007/s11629-016-4068-2.

[13] Guzzetti F, Mondini AC, Kardinali M, Fiorucci F, Santangelo M, Chang K. Landslide inventory maps: new solution to old problems. Earth-Sci Rev 2012;112:42–66. doi: https://doi.org/10.1016/j.earscirev.2012.02.001.

[14] Hasekioğulları GD, Ercanoglu M. A new approach to use AHP in landslide susceptibility mapping: a case study at Yenice (Karabuk, NW Turkey). Nat Hazards 2012;63:1157–79. doi: https://doi.org/10.1007/s11069-012-0218-1.

[15] Ilia I, Tsangaratos P. Applying weight of evidence method and sensitivity analysis to produce a landslide susceptibility map. Landslides 2015;13 (2):379–97. doi: https://doi.org/10.1007/s10346-015-0576-3.

[16] Jazouli AE, Barakat A, Khelloud R. GIS-multicriteria evaluation using AHP for landslide susceptibility mapping in Oum Er Rbia high basin (Morocco). Geoenviron Disasters 2009;6:1–12. doi: https://doi.org/10.1186/s40677-019-0119-7.

[17] Khamhong P, Yingviwatanapong C, Ransikarbum K. December). Fuzzy Analytic Hierarchy Process (AHP)-based Criteria Analysis for 3D Printer Selection in Additive Manufacturing. Res, Invent, Innovat Cong (RI2C) 2019:1–5. doi: https://doi.org/10.1109/RI2C48728.2019.8999950.

[18] Kayastha P, Dhital MR, Smedt FD. Application of the analytical hierarchy process (AHP) for landslide susceptibility mapping: A case study from the Tinau watershed, West Nepal. Computers Geosci 2012;52:398–408. doi: https://doi.org/10.1016/j.cageo.2012.11.003.

[19] Kreuzer TM, Wilde M, Terhorst B, Damm B. A landslide inventory system as a base for automated process and risk analyses. Earth Sci Inf 2017;10:507–15. doi: https://doi.org/10.1007/s12145-017-0307-5.

[20] Kavzoglu T, Kutlug Sahin E, Colkesen I. An assessment of multivariate and bivariate approaches in landslide susceptibility mapping: a case study of Duzkoy district. Nat Hazards 2015;76(1):471–96. doi: https://doi.org/10.1007/s11069-014-1506-8.

[21] Kumar R, Anbalagan R. Landslide susceptibility mapping using analytic hierarchy process (AHP) in Tehri Reservoir Rim Region, Uttrakhand. J Geol Soc India 2016;87:1–16. doi: https://doi.org/10.1007/s12594-016-0395-8.

[22] Kumar R, Brookfield ME. Sedimentary environments of the Simla Group (Upper Precambrian), Lesser Himalaya, and their palaeotectonic significance. Sed Geol 1987;52(1–2):27–43. doi: https://doi.org/10.1016/0037-0738(87)90015-7.

[23] Lee S, Min K. Statistical analysis of landslide susceptibility at Yongin. Korea Environ Geol 2001;40(9):1095–113. doi: https://doi.org/10.1007/s002540100310.

[24] Mattivi P, Franci F, Lambertini A, Bitelli G. TWI computation: a comparison of different open source GISs, Open Geospatial Data, Software and Standards 2019;4(6). 10.1186/s40965-019-0066-y Map of India downloaded from https://d-maps.com/, accessed on 02/06/2021

[26] Milevski I, Dragićević S, Zorn M. Statistical and expert-based landslide susceptibility modeling on a national scale applied to North Macedonia. Open Geosci 2019;11(1):750–64. doi: https://doi.org/10.1515/geo-2019-0059.

[27] Myronidis D, Papageorgiou C, Theophanous S. Landslide susceptibility mapping based on landslide history and analytic hierarchy process (AHP). Nat Hazards 2015;81:245–63. doi: https://doi.org/10.1007/s11069-015-2075-1.

[28] Ohlmacher GC. Plan curvature and landslide probability in regions dominated by earth flows and earth slides. Eng Geol 2007;91(1–2):117–34. doi: https://doi.org/10.1016/j.enggeo.2007.01.005.
[29] Panchal S, Shrivastava AK. Application of analytic hierarchy process in landslide susceptibility mapping at regional scale in GIS environment. J Statist Manage Syst 2020;23(2):199–206. doi: https://doi.org/10.1080/09720510.2020.1724620.
[30] Panchal S, Shrivastava AK. A Comparative Study of Frequency Ratio, Shannon's Entropy and Analytic Hierarchy Process (AHP) Models for Landslide Susceptibility Assessment. ISPRS Int J Geo-Inf 2021;10(9):1–20. doi: https://doi.org/10.3390/ijgi10090603.
[31] Pardeshi SD, Autade SE, Pardeshi SS. Landslide hazard assessment: recent trends and techniques. SpringerPlus 2013;2(523):1–11. doi: https://doi.org/10.1186/2193-1801-2-523.
[32] Pathak D. Knowledge based landslide susceptibility mapping in the Himalayas. Geoenviron Disasters 2016;3(8):1–11. doi: https://doi.org/10.1186/s40677-016-0042-0.
[33] Park S, Choi C, Kim B, Kim J. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. Environ Earth Sci 2013:1443–1464. 10.1007/s12665-012-1842-5
[34] Polykretis C, Chalkias C. Comparison and evaluation of landslide susceptibility maps obtained from weight of evidence, logistic regression, and artificial neural network models. Nat Hazards 2018;93(1):249–74. doi: https://doi.org/10.1007/s11069-018-3299-7.
[35] Pourghasemi HR, Mohammady M, Pradhan B. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin. Iran Catena 2012;97:71–84. doi: https://doi.org/10.1016/j.catena.2012.05.005.
[36] Pradhan B. Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. J Indian Soc Remote Sens 2010;38(2):301–20. doi: https://doi.org/10.1007/s12524-010-0020-z.
[37] Ram P, Gupta V, Devi M, Vishwakarma N. Landslide susceptibility mapping using bivariate statistical method for the hilly township of Mussoorie and its surrounding areas, Uttarakhand Himalaya. J Earth Syst Sci 2020;129(1):1–18. doi: https://doi.org/10.1007/s12040-020-01428-7.
[38] Ransikarbum K, Mason SJ. Multiple-objective analysis of integrated relief supply and network restoration in humanitarian logistics operations. Int J Prod Res 2016;54(1):49–68. doi: https://doi.org/10.1080/00207543.2014.977458.
[39] Ransikarbum K, Mason SJ. Goal programming-based post-disaster decision making for integrated relief distribution and early-stage network restoration. Int J Prod Econ 2016;182:324–41. doi: https://doi.org/10.1016/j.ijpe.2016.08.030.
[40] Ransikarbum K, Pitakaso R, Kim N, Ma J. Multicriteria decision analysis framework for part orientation analysis in additive manufacturing. J Comput Des Eng 2021;8(4):1141–57. doi: https://doi.org/10.1093/jcde/qwab037.
[41] Regmi AD, Yoshida K, Pourghasemi HR, DhitaL MR, Pradhan B. Landslide susceptibility mapping along Bhalubang — Shiwapur area of mid-Western Nepal using frequency ratio and conditional probability models. J Mountain Sci 2014;11(5):1266–85. doi: https://doi.org/10.1007/s11629-013-2847-6.
[42] Reichnenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F. A review of statistically based landslide susceptibility models. Earth-Sci Rev 2018;180:60–91. doi: https://doi.org/10.1016/j.earscirev.2018.03.001.
[43] Sarkar S, Roy AK, Martha TR. Landslide susceptibility assessment using Information Value Method in parts of the Darjeeling Himalayas. J Geol Soc India 2013;82(4):351–62. doi: https://doi.org/10.1007/s12594-013-0162-z.
[44] Senouci R, Taibi N-E, Teodoro AC, Duarte L, Mansour H, Yahia Meddah R. GIS-Based Expert Knowledge for Landslide Susceptibility Mapping (LSM): Case of Mostaganem Coast District, West of Algeria. Sustainability 2021;13(2):1–21. doi: https://doi.org/10.3390/su13020630.
[45] Shahabi H, Ahmad BB, Khezri S. Evaluation and comparison of bivariate and multivariate statistical methods for landslide susceptibility mapping (case study: Zab basin). Arabian J Geosci 2013;6(10):3885–907. doi: https://doi.org/10.1007/s12517-012-0650-2.
[46] Siddique T, Pradhan SP, Vishal V, Mondal MEA, Singh TN. Stability assessment of Himalayan road cut slopes along National Highway 58. India Environ Earth Sci 2017;76(22). doi: https://doi.org/10.1007/s12665-017-7091-x.
[47] Singh K, Kumar V. Hazard assessment of landslide disaster using information value method and analytical hierarchy process in highly tectonic Chamba region in bosom of Himalaya. J Mountain Sci 2018;15(4):808–24. doi: https://doi.org/10.1007/s11629-017-4634-2.
[48] Shano L, Raghuvanshi TK, Meten M. Landslide susceptibility evaluation and hazard zonation techniques – a review. Geoenviron Disasters 2020;7(18):1–19. doi: https://doi.org/10.1186/s40677-020-00152-0.
[49] Saaty TL. The analytic hierarchy process. New York: McGraw-Hill; 1980.
[50] Saaty RW. The analytic hierarchy process—what it is and how it is used. Math Modell 1987;9:161–76. doi: https://doi.org/10.1016/0270-0255(87)90473-8.
[51] Tanyaş H, van Westen CJ, Persello C, Alvioli M. Rapid prediction of the magnitude scale of landslide events triggered by an earthquake. Landslides 2019;16:661–76. doi: https://doi.org/10.1007/s10346-019-01136-4.



Sandeep Panchal is currently pursuing PhD in Geotechnical Engineering at Delhi Technological University. His research interests are GIS applications, landslide mapping and slope stability.



Dr. Amit Kr. Shrivastava is Professor at department of civil engineering in Delhi Technological University, India. He has authored many research papers and developed scientific equipment.

Low Power Pre-charge Voltage Level and Low Swing Logic Based 8T SRAM Cell for High Speed CMOS Circuits

R. M. Rawat and V. Kumar

Abstract – This article describes the memory cell proposed design of a low-power 8T SRAM cell. Two PMOS transistors are employed in this proposed approach, one of which is connected to the node voltage Q and the other one is connected to the Qbar. In order to remove the swing voltage at the bit and bit bar lines. When the cell is in working mode, the swing voltage is minimized, which reduces the dynamic power dissipation. Pre-charge voltage level is a technique in which a PMOS transistor acts as a switch and an NMOS transistor acts as a resister, resulting in a series of leakage current reductions when the transistors move from sleep to active and vice versa. The findings of the comparison show that the read, write, and hold mode operation for the proposed 8T SRAM cell is superior to that of the 6T SRAM cell. This is frequently because greater static noise margins are obtained, ensuring good bit cell write ability. All simulation results of read and write on the proposed 8T SRAM cell and current models of 6T SRAM cell have been applied to 130 nm CMOS technology utilizing the Symica DE tool.

I. Introduction

Low-power advanced semiconductor memory cell design is a critical component of many digital systems, including personal and workstation computers, commercial and military satellite equipment, telecommunications, consumer technology, and so on. Because the battery powers high-speed portable devices such as notebook computers, laptops, smartphones, cellular phones, and so on, its demand is steadily increasing. For a good reaction, high-speed transportable devices require primary memory. SRAM is used for this function since it is faster and does not require regular replenishment. The most serious issues with large speed SRAM cells involve dynamic power dissipation and leakage current, which diminish the backup power lifetime of portable electronics. As a result, an SRAM cell design with low static and dynamic power dissipation is required [1].

"For numerous years, there has been a focus on low-power architecture and style innovation on increasing the speed and operation frequency of digital systems. The development of complementary Metal Oxide Semiconductor (CMOS) technology enables integrated circuits to have a higher density and greater performance (ICs). When an IC's density, and thus its power consumption, grows, temperature management becomes more challenging [2]". Every bit in an SRAM is recorded on two cross-coupled inverters comprised of four transistors [3]. The better performance demand, however, limits supply voltage scaling. The scalability of supply voltage alone is insufficient to keep the power density within bounds for power-sensitive applications [4]. Circuit approaches and system level strategies, in addition to supply voltage scaling, are required to achieve low power designs [5]. Because of the forced scaling of the gadgets, sub-threshold leakage rises, as well as other bad impacts, such as better Drain Induced Barrier Lowering (DIBL), threshold voltage (Vth) spiel off, lowered on current to off current ratio, and increased source to vacant resistance [6]. The severity of Vth on the channel length increases as Vth recites. If the channel length varies slightly, the device properties become unpredictable, resulting in large threshold voltage variations. As these gadgets are scaled, oxide thickness scaling and improved, non-uniform doping should be introduced to prevent these small channel lighting tricks [7]. Due to the thin oxide thickness, a large electric field is created, resulting in a large direct tunnelling current [8]. "A lower memory transparency Adaptive Body Bias (ABB) circuitry is expected to increase SRAM constancy and yield by balancing age and process fluctuations [9]. A threshold voltage sensor circuit with an analogue controller is part of the intended ABB circuit for power saving. The Multi-Threshold Complementary Metal Oxide Semiconductor (MTCMOS) technology offers low leakage and superior performance by utilizing high speed, Low Threshold Voltage (LVT) transistors throughout the active region and low leakage, High Threshold Voltage (HVT) transistors throughout the sleep mode, reducing the static power loss of the SRAM circuit [10][11]". The CMOS cell that is being developed consumes low power and has faster read and write rates. The cell is based on the notion of single-ended memory storage. This strategy is based on the lowering of static power. A symmetrical architecture based SRAM cell has also been predicted for greater stability [12]. In a concept [13], a 5T SRAM cell is anticipated to have quick performance, high density, and power efficiency. Two transistors separate the cell storage nodes from the read operations path to maintain the cell's information reliability. This topology improves the data stability. "It describes a power-efficient multiport multiport

R. M. Rawat and V. Kumar are with the Department of Computer Science and Engineering, Delhi Technological University, Delhi, Bawana Rd, Shahbad Daulatpur Village, Rohini, Delhi, 110042, India. E-mail: rammurtirawat@dtu.ac.in, vinod_k@dtu.ac.in

978-1-6654-4528-3/21/$31.00 ©2021 IEEE

2021 IEEE 32nd International Conference on Microelectronics (MIEL) | 978-1-6654-4528-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/MIEL52794.2021.9569198

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:56:26 UTC from IEEE Xplore. Restrictions apply.

SRAM featuring cross-point write word lines, shared write bit lines, and shared write row access transistors. The design has one write, one read and two write, two read multi-port SRAMs for registering file applications in Nano scale CMOS technology" [14]. "The cell has a cross-point write word line configuration, which reduces the risk of writing half select and increases the static noise margin (SNM). The write bit lines (WBLs) and write row access transistors are exchanged with neighbouring bit cells to minimize cell transistor consumption and space. Since the design halves the amount of WBL, there is less leakage and power usage. In contrast, for the reading stack, column based virtual ground direct is used to reduce read power usage. The low power consumption of the SRAM cell is attributed to its chained linked drivers, which are powered by bit lines, and read buffers, which give stack consequences" [15].

During this research, a novel low-power 8T SRAM cell was presented. To reduce leakage currents and static power dissipation throughout mode transition, a pre-charge voltage level approach is used. Two voltage sources are employed at the output layer to reduce the fluctuating voltage. As an outcome, the dynamic energy dissipation throughout switch movement is reduced. The efficiency parameters of the suggested SRAM cell are evaluated and compared to those of the competing SRAM designs.

The following is how the paper is structured: Section II goes over existing SRAM cells, whereas Section III goes through the circuit design and operation of the proposed unique 8T SRAM cell. Section IV covers the full examination of the suggested cell's features and comparisons with existent SRAM cells, and Section V brings the paper to a close.

II. Existing SRAM Cell Designs

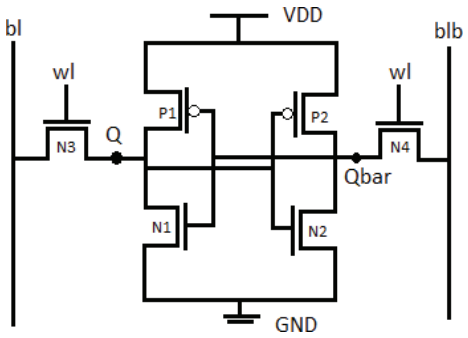


Fig. 1. Conventional 6T SRAM cell [16].

Figure 1 depicts a standard 6T SRAM cell. It has the same six transistors as an SR latch. 4 cross-coupled CMOS inverters (P1, P2 and N1, N2) and two NMOS transistors (N3 and N4) provide accessibility to the cell. The access transistors are switched on when word line (WL) is selected [16]. As an outcome, the supply voltage (Vdd) is generated via the pull-up transistor P1 of the node storing '1'. On the other hand, current passes from the bit bar line's pre-charged state to ground, draining the bit bar lines.

III. Proposed SRAM Cell

A pre-charge voltage level and minimal swing logic based low power 8T SRAM cell are implemented to achieve low static and dynamic power losses for read and write processing, as well as better stability, in order to solve the problems existing with standard SRAM cells and other established SRAM cells.

A. Operation

The presented design can be seen in Figure 2 Two PMOS transistors, M7 and M8, are joined to the cross-coupled node volts Q and Qbar, respectively, in the presented architecture. Throughout the writing and reading operations, PMOS transistors are turned on and off automatically. The pre-charge voltage level reduces the swing voltage throughout read and write operations in the suggested 8T SRAM cell [17]. The dynamic energy dissipation is reduced by lowering the swing voltage. The comparative results show that the suggested 8T cell performs better in read, write, and hold modes than the 6T cell.

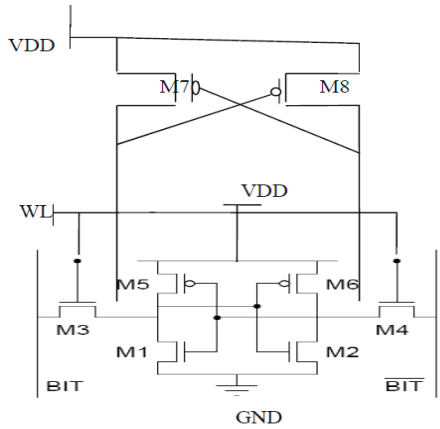


Fig. 2. Proposed 8T SRAM Cell [17].

IV. Results and Discussions

The system is defined by its use of 130 nm technology and a 1.2 volt voltage supply. The simulation is run in the SymicaDE tool. This study examines the suggested 8T SRAM cell in regards to write capability, read stability, and holding static noise margin. This value is then compared to a 6T SRAM cell that is standard.

A. Hold Stability

The most popular method for determining the cell's hold ability is to use the static noise margin (SNM).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:56:26 UTC from IEEE Xplore. Restrictions apply.

Whenever the SRAM cell is in the hold state, hold stability is determined. The word lines are turned off in the holding state, so the cell is completely separated from the bit lines. Before switching the content of the cell, SNM sets the maximum amount of noise that can be placed on the storage nodes. The hold static noise margins of 6T and suggested 8T SRAM cells are shown in Figures 3 and 4, respectively.

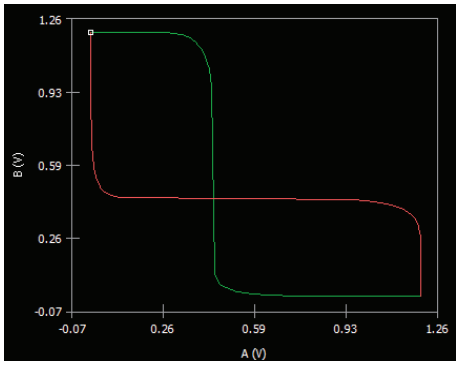


Fig. 3. HSNM 6T SRAM CELL.

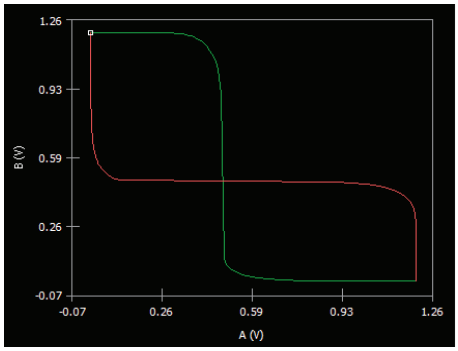


Fig. 4. HSNM PROPOSED 8T SRAM CELL.

B. Read Stability

In an SRAM, read static noise margin (RSNM) is used to determine read stability. Due to the remoteness of the storing nodes, the suggested 8T SRAM cell has a superior RSNM than a normal 6T SRAM cell. The read reliability of 6T and projected 8T SRAM Cells is depicted in Figures 5 and 6, correspondingly.

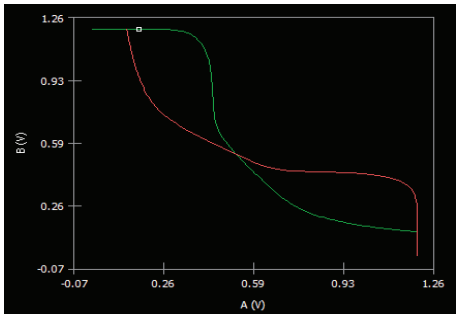


Fig. 5. RSNM 6T SRAM CELL.

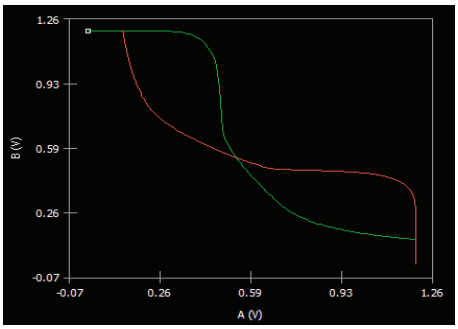


Fig. 6. RSNM PROPOSED 8T SRAM CELL.

TABLE I
COMPARISON OF HOLD SNM AND READ SNM

Cell	Read SNM (in Volts)	Hold SNM (in Volts)
6T SRAM	0.0711	0.125
Proposed 8T SRAM	0.1000	0.125

C. Write Stability

Write static noise margins are used to determine write stability (WSNM). Due to the remoteness of the storage nodes, the suggested 8T SRAM cell has a superior WSNM than a regular 6T SRAM cell. The write reliability of 6T SRAM and projected 8T SRAM cells is depicted in Figures 7 and 8, correspondingly.

The write trip level of a cell is a measurement of its ability to write. It demonstrates how challenging it is for the cell's storage nodes. When Q and Qbar switch their content, the bit line voltage is swept from 0 to Vdd, and the cell switching is registered. The Write trip point is defined as the value of bitline voltages at the intersection of internal storage nodes Q and Qbar.

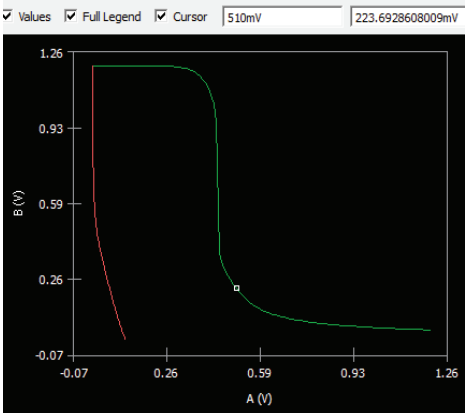


Fig. 7. 6T SRAM Write Trip Point.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:56:26 UTC from IEEE Xplore. Restrictions apply.

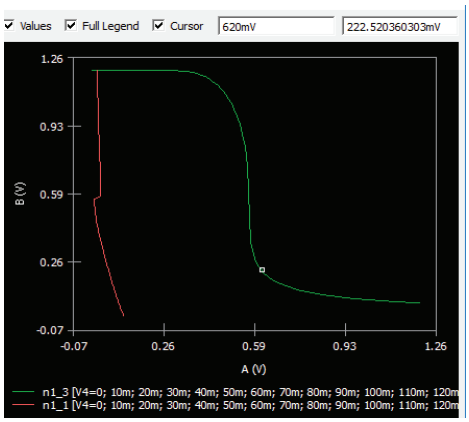


Fig. 9. Proposed 8T SRAM Write Trip Point.

TABLE II
COMPARISON FOR WRITE TRIP POINT OF 6T SRAM AND 8T SRAM CELLS

Cell	Write Trip Point
6T SRAM	675 mv
Proposed 8T SARM	900 mv

V. CONCLUSION

A unique low power 8T Cell and low swing oriented 8T SRAM Cell is employed in this study, which consumes lesser dynamic power throughout write/read operations due to lower swing voltages supplied by the Node voltages Q and Qbar. As a result, the suggested 8T SRAM cell uses reduced static power. Low leakage current and Node voltages offer better stability. According to the simulation, the suggested 8T SRAM cell consumes less dynamic power, has better stability, and has less current leakage throughout mode transitions to standard 6T SRAM cells. Albeit the size increases in compared to 6T SRAM cells, the reduced power consumption and higher stability can overcome the disadvantages. The large increases in results gained by the use of pre-charge voltage levels and low swing logic would be relevant to future limited energy memory designs such as laptop, cellphones, and programmable logic device, among others.

REFERENCES

[1] P. Upadhyay, S. Ghosh, R. Kar, D. Mandal, and S. P. Ghoshal, "Low static and dynamic power MTCMOS based 12T SRAM cell for high speed memory systems", in *Proc. 11th IEEE international joint conference on computer science and software engineering (JCSSE-14)*, 2014, pp. 212–7.

[2] G. Pasandi and S. M. Fekhraie, "A 256-kb near-threshold SRAM with 1 k cells per bit line and enhanced write and read operations", *IEEE Transactions on VLSI Systems*, 2015, vol. 23, no. 11, pp. 1063–8210.

[3] K. Gavaskar and S. Priya, (2013). "Design of efficient low power stable 4-bit memory cell", *International Journal of Computer Applications*, 2013, vol. 84, no. 1, pp. 0975–8887.

[4] K. Gavaskar, U. S. Ragupathy, "An efficient design and comparative analysis of low power memory cell structures", in *Proc. IEEE international conference on green computing, communication and electrical engineering (ICGCCEE'14)*, 2014, pp. 489–493.

[5] S. Borkar, "Design challenges of technology scaling", *IEEE Micro*, 1999, vol. 19, no. 4, p. 23.

[6] J. Brews, *High speed semiconductor devices* , pp. 139–210, 1990, New York: Wiley.

[7] K. Roy and S. C. Prasad, (2000). *Low power CMOS VLSI circuit design*, 2000, pp. 27–29, New York: Wiley.

[8] Y. Taur and T. H. Ning, *Fundamentals of modern VLSI devices*, 1998, pp. 285–286. New York: Cambridge University Press.

[9] H. Mostafa, M. Anis, and M. Elmasry, "Adaptive body bias for reducing the impacts of NBTI and process variations on 6T SRAM cells", *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2011, vol. 58, no. 12, pp. 2859–2871.

[10] J. Kao, A. Chandrakasan, and D. Antoniadis, "Transistor sizing issues and tool for multi threshold cmos technology", in *Proc. IEEE 34th design automation conference (DAC)*, 1997, pp. 409–14.

[11] J. Kao, S. Narenda, and A. Chandrakasan, "MTCMOS hierarchical sizing based on mutual exclusive discharge patterns", in *Proc. IEEE 35th design automation conference (DAC)*, 1998, pp. 495–500.

[12] R. Saeidi, M. Sharifkhani, and K. Hajsadeghi, "A sub-threshold symmetric SRAM cell with high read stability", *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2014, vol. 61, no. 1, pp. 26–30.

[13] R. Gupta and S. S. Gill, "A novel low leakage and high density 5T CMOS SRAM cell in 45 nm technology", in *Proc. Conference on recent advances in engineering and computational sciences*, 2014, pp. 1–6.

[14] D. P. Wang, H. J. Lin, C. T. Chuang, and W. Hwang, "Low-power multiport SRAM with cross-point write word-lines, shared write bit-lines, and shared write row access transistors", *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2014, vol. 61, no. 3, pp. 188–192.

[15] A. Islam and H. Mohd, "Variability aware low leakage reliable SRAM cell design technique", *Journal of Microelectronics Reliability*, 2012, vol. 52, pp. 1247–1252.

[16] T. H. Kim, J. Liu, J. Keane, C. H. Kim, "Circuit techniques for ultra-low power subthreshold SRAMs", In IEEE international symposium on circuits and systems (ISCAS), 2008, pp. 2574–77.

[17] R. M. Rawat, "A Novel Low power and Swing Restoration SRAM Logic Circuit Technique", 2018, *International Journal of Engineering and Technical Research*, vol. 7, no. 02.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:56:26 UTC from IEEE Xplore. Restrictions apply.

Wireless Personal Communications
https://doi.org/10.1007/s11277-021-09277-3

Modeling and Simulation of Molecular Communication Based Nanonetwork Using Finite Shaped Spherical Receiver

Garima Singh[1] · **Gurjit Kaur**[1]

Accepted: 19 October 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract
In this paper, we have presented an analytical diffusion-based molecular communication system for nanonetwork with a finite spherical receiver in a three dimensional fluid channel. Use of finite shaped spherical absorbing receiver is one of the driving forces behind the study of this molecular communication system model. We have derived the mathematical expression to calculate the exact number of which got absorbed by the spherical receiver molecules expected at any time duration. For this, we have formulated a mathematical expression to calculate life expectancy of molecules propagating through Brownian motion. We also calculated the total number of time slots for which the molecules can survive. To enhance the practical feasibility, the molecules of previous slot are also being considered for a particular time slot. In this proposed system, the transmitted information is encoded by the number of molecules and modulated via *M*-ary modulation schemes. Since the diffusion coefficient of fluid medium and modulation schemes are crucial parameters for the channel capacity, so an exhaustive analysis for the channel capacity has been done using maximum likelihood detection at the finite spherical receiver and simulated results are presented accordingly. Importantly, our proposed system will act like a benchmark to setup a molecular communication based nanonetwork for finite shape spherical receiver.

Keywords Nanotechnology · Nanonetworks · Brownian motion · Molecular Communication · Capacity · Diffusion · Hitting rate · Mutual information

1 Introduction

In the current scenario, nanomachines are recognized as the most essential functional unit, facilitating the miniaturization and fabrication of nano scale devices at one to hundreds of nanometers e.g. bio-chip integrated mobile phones, and can perform many nano-level tasks like sensing, data storing and computing. In order to compute more complex tasks in

✉ Garima Singh
garimasinghmavi@gmail.com

Gurjit Kaur
gurjitkaur@dtu.ac.in

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

a distributed manner these nanomachines can be interconnected to form a *nanonetworks* [1]. These nanotworks are envisaged to expand the capabilities and applications of single nano-machines, both in terms of complexity and range of operation. Regardless of the final application, traditional communication paradigms requires a profound revision in order to meet the requirements of these new nano-scenarios. The electromagnetic waves based traditional communication system, having an either wireless or wired links, might not be applicable directly as it is practically unfeasible to wire a large quantity of them in nanonetworks. Even the electromagnetic transceivers of the traditional communication systems faces the constraint of the size, current complexity, availability of spectrum and antenna size thus making them uneasy to get integrated into the nanonetworks. Even the conventional communication systems are considered inappropriate for target health solution, diagnostic applications, communication among biological-nanomachines, drug delivery system in body. On the other hand, molecular communcation based nanonetwork emerged has vast applications as a drug carrier among bio-nanomachines which could be implanted in the human body.

The advancements in miniaturization by using biological materials have necessitated the development of a nano-scale communication network that is bio-inspired and can be used at nano-scale environment where traditional schemes fails. Molecular communication (MC) presents a bio-inspired solution for nano-scale networks where two nano-devices can be connected via fluid medium and uses molecules as the information carrier [2, 3]. Additionally, this technique has potential applications in fields like industries, bio-medical, military and would be helpful in communication inside networks of tunnel, pipelines, or unpredictable underwater environment. This communication is completely a new concept and would potentially enable many new applications in bio-nanotechnology related to medical science and health care where it can be used for targeted delivery of medicines inside body where conventional communication system fails.

In molecular communication, information is encoded and decoded in the form of molecules, rather than electrons or electromagnetic waves and these molecules could be proteins, ions, or DNAs. In a fluid medium, random bombardment by the fluid's molecules results in propagation through Brownian motion. Brownian motion is considered physically realistic for nanodevices, because dimensions of these nanodevices are broadly of the same magnitude as of individual molecules. Different properties are used to modulate the information which need to be sent like concentration [4], time [5], type [6] and number [7] of transmitted molecules. Also, there are other ways to transmit the information molecules from the transmitter to the receiver like diffusion flow [8, 9], biologically engineered bacteria and molecular motors [10] etc.. In the diffusion-based molecular communication, the molecules propagate through their spontaneous diffusion in a fluid medium and solely follows the laws of diffusion stated by Wiener process. The example of such diffusion-based architecture is air, water and calcium signaling among cells [11]. In [12], amplitude modulation scheme in terms of information rate and capacity has been analyzed for diffusion based molecular communication. An expression for the capacity with achievable information rate for a concentration-modulated time-slotted molecular communication channel has been obtained in [14, 15]. In [16], the author has encoded the transmitted information by the release time of molecules called as timing based molecular channels and has done the symbol error probability (SEP) and channel capacity analysis. For this analysis he has taken the positive drift between transmitter and receiver. For the same, a theoretical foundation for inverse Gaussian (IG) molecular channel is developed in [17] and its performance has been analyzed in terms of capacity bound and SEP. In [18] the authors have consider a pair of nanomachines communicating by means of on-off keying (OOK)

transmission protocol in a three-dimensional ideal (i.e. free) diffusion-based unbounded propagation environment. Mathematical analysis for diffusion driven, timing based molecular communication channel has been done in [19] and first passage time for information molecules through truncated L*é*vy distribution has been obtained. Life expectancy of molecules while traveling in fluid medium has been discussed in[20]. In [21] authors have calculated the collision probability of two information molecules while travelling in fluid before reaching at the receiver.

To the best of author's knowledge, very limited research has been done to address the modeling and analysis of diffusion-based molecular communication channel in nanonetworks. Aforementioned research have used the receiver which have infinite absorbing surface which means that a molecule either arrives at the receiver or dies out. However, in practical scenario, the receiver should have finite absorbing surface which means molecule might cross the receiver without being absorbed. In this research paper we have proposed a molecular based communication system by considering a spherical shaped receiver with finite radius and surface area. The choice of spherical shape for the receiver is motivated by the fact that membrane bilayers in molecules form sphere shape naturally [22, 23], so we focus on this shape. These membrane bilayers inherently have molecule receptors as a part of functioning nanomachine like in a cell. None of existing research in literature analyze the molecular communication channel for these issues of practical interest such as finite receiver surface. In such scenario only those molecules will be counted which would actually hit the receiver. Unlike existing work on MC, In our proposed system the molecule is not bounded by the arrival time. So, a time-slotted channel of slot duration T_s is considered. We have proposed the model in which previous slot molecules having information are also been considered till they are alive in the fluid medium. To do this, a mathematical expression has been derived to calculate the life expectancy of current as well as previous slot molecule. Since, the enhanced channel capacity is an important requirement for any molecular communication system, so there is a need to do a rigorous analysis of the other design parameters that have a great impact on system capacity i.e. diffusion coefficient of fluid medium as well as modulation schemes. In our research work, diffusion process is based on Brownian motion, which is due to interactions among the fluid particles. Further, a generalized M-ary modulation scheme is considered in this paper, where M representing the amplitude levels number.

Despite the aforementioned challenges, the objective of this paper is to address all these issues of molecular communication. The main contribution of paper is

- For the first time diffusion-based MC system have been modeled and simulated with a finite shaped spherical receiver for nanonetworks.
- We derive the exact expression for exponentially decaying life expectancy of the molecules in a three dimensional (3D) fluid environment. In which current as well as previous slot molecules have also been considered.
- To increase the practical feasibility of the system, we have derived a mathematical expression to calculate the first hitting probability for a spherical shaped absorbing receiver. For this, we have considered the finite lifetime molecules that can be absorbed in a given time slot.
- Based on the above, the released molecules are modulated through M-ary modulation scheme and its closed form expressions for mutual information and channel capacity are obtained using maximum likelihood detection at the finite surface receiver.

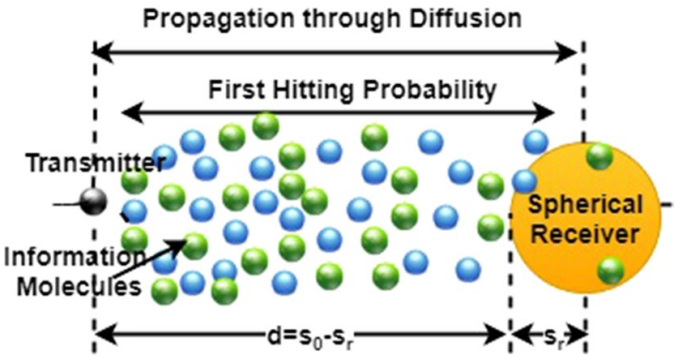


Fig. 1 Representation of molecular communication system model for spherical receiver

The remainder paper is arranged as follows, Sect. 2 discusses the system model. The mathematical model for hitting rate to the spherical absorbing receiver is developed in Sect. 3. In Sect. 4, the mutual information and capacity of spherical receiver system are calculated and numerical results and conclusion of the paper is provided in Sect. 5.

2 Modeling of Spherical Shaped Absorbing Receiver

We consider a molecular communication based channel where the information molecules propagate by pure diffusion in a 3-D environment of finite expanse in all directions expressed in spherical coordinate system. We consider a point source transmitter which is capable of emitting a large number of molecules. The receiver considered in our analysis is a sphere of finite shape having perfectly absorbing surface, which means that it will absorb and count all the molecules that hit it's surface. Furthermore, we assume that transmitter and receiver are perfectly synchronized in time via an external signal as discussed in [24]. The proposed system model for molecular communication is shown in Fig. 1. We assume that transmitter is a point source, located at the origin and which is s_0 distance apart from the center of the spherical receiver. The radius of considered spherical receiver is s_r. The distance between transmitter and nearest point on the spherical receiver's surface is $d = s_0 - s_r$. The transmitter releases the information molecules into 3-D fluid media, which follow an independent trajectory and takes a random propagation time while traveling to the receiver. In proposed diffusion-based MC system, the information-carrying molecules requires no additional infrastructure for transportation in fluid medium and relies on free diffusion only. Additionally, various environmental effects can affect the signaling molecules such as chemical reactions or flow, which may prevent the molecules from reaching the receiver in the current time slot and enabling them to reach in subsequent time slots. Thus, a natural assumption regarding the information molecules is that it has finite lifetime. This finite lifetime, $h(\tau)$ can be modelled as an exponential decay

$$h(\tau) = \alpha e^{-\alpha\tau}, \qquad \tau > 0\,, \tag{1}$$

where *degradation parameter* is represented by α and τ represents the life expectancy of molecules. For this exponential decaying model, the information molecule will decay to $l\%$ of its initial state in t_c time. After this molecule will no longer be recognized by the receiver and will terminate as useless. This t_c can be expressed in terms of α and l as

$$t_c = \frac{1}{\alpha} ln\left(\frac{100}{l}\right). \tag{2}$$

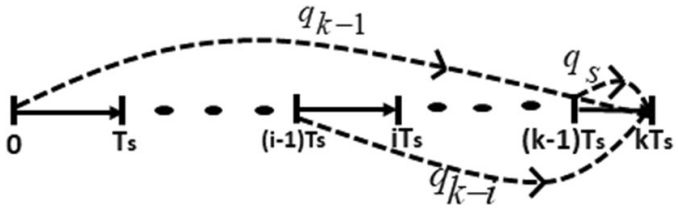


Fig. 2 Time-slotted channel for kth slot

In the next section, we characterize this molecular communication channel in terms of the number of molecules which arrives at the receiver within a given time slot.

3 Channel Characterization

A receiver having a finite spherical surface area will have a smaller number of molecular arrivals within a time duration as compared to an infinite plane receiver. To analyze this type of channel, we first need to characterize it in terms of the count of molecules reaching the receiver within a given time slot. For such MC channels the number of molecules $h_{hit}(\Theta, t/s_0)$, hitting the spherical receiver per unit time is given below where the transmitter is at a distance s_0 from the receiver's center

$$h_{hit}(\Theta, t/s_0) = \frac{s_r}{s_0}\frac{1}{\sqrt{4\pi Dt}}\frac{(s_0 - s_r)}{t}\exp\left(-\frac{(s_0 - s_r)^2}{4Dt}\right) \tag{3}$$

where D represents the diffusion coefficient and Θ is the symbol used to represent the spherical receiver.

Through integrating the $h_{hit}(\Theta, t/s_0)$ over the time interval [0, t] we get the count of molecules received and then absorbed by the receiver until time t, $H_{hit}(\Theta, t/s_0)$ as

$$H_{hit}(\Theta, t/s_0) = \int_0^t h_{hit}(\Theta, t'/s_0)\mathrm{d}t', = \frac{s_r}{s_0} erfc\left[\frac{s_0 - s_r}{\sqrt{4Dt}}\right] \tag{4}$$

Thus, the number of molecules absorbed by the receiver until time t is parameterized by the diffusion coefficient D of the information molecules in the fluid medium and the transmitter-receiver distance s_0 as shown by (4). Therefore, the expected number of molecules that are hitting the receiver, Θ and absorbed by it in the interval $[t, t+\Delta t]$ and is given by

$$\mathbb{E}\left[N_r^{\Theta}(t, t+\Delta t)\right] = N_t^{\Theta}\left[H_{hit}(\Theta, t+\Delta t/s_0) - H_{hit}(\Theta, t/s_0)\right]. \tag{5}$$

where N_t^{Θ} and N_r^{Θ} are the transmitted and received number of molecules, respectively.

The considered channel is partitioned into time slots of duration T_s with *ith* slot being the time duration $[(i-1)T_s, iT_s], i \in [1, 2, \ldots]$. At the starting of every time slot, a fixed number of molecules are emitted by the transmitter in to the fluid medium. Molecules which are not able to reach the receiver in current slot will stay in medium according to their life expectancy and will appear as stray molecules in the successive slots and cause an inter-symbol interference (ISI) as shown in Fig. 2. Molecules from the last r time slots will contribute to this ISI, where r is given as

$$r = \left\lfloor \frac{l}{T_s} \right\rfloor, \tag{6}$$

where $\lfloor \cdot \rfloor$ represents the floor function.

The count of molecules received in current slot are denoted by N_0. For the kth slot, N_i represents the count of stray molecules coming from the ith previous slot. Therefore, the total number of molecules received by the spherical receiver in kth slot is indicated as

$$N_k = N_0 + N_1 + N_2 + N_3 + \cdots + N_{k-1}, \tag{7}$$

and from (5) we will be able to conclude that the probability of such molecules received by the receiver in a given time interval $\left[kT_s, (k+1)T_s\right]$ is given by

$$p_k = \left[H_{hit}(\Theta, (k+1)T_s/s_0) - H_{hit}(\Theta, kT_s/s_0)\right]. \tag{8}$$

We assume that molecules will not interfere with each other while traveling and follow an independent trajectory. Then, assuming the Gaussian distribution for each of the terms on the right hand side of (7), N_k is also Gaussian distributed with mean and variance given as

$$\begin{aligned} \mu_{N_k} =& \mu_{N_0} + \mu_{N_1} + \mu_{N_2} + \cdots + \mu_{N_{k-1}}, \\ \sigma^2_{N_k} =& \sigma^2_{N_0} + \sigma^2_{N_1} + \sigma^2_{N_2} + \cdots + \sigma^2_{N_{k-1}}. \end{aligned} \tag{9}$$

The variance $\sigma^2_{N_k}$ is responsible for the randomness, caused by the fluid diffusion and can be cited as the Brownian Noise [25, 26] while the stray molecule components μ_{N_i} and $\sigma^2_{N_i}$ in (9), $1 \leq i \leq k-1$ can be referred as residual noise, for the ith previous slot [25, 26].

4 Amplitude Modulation Scheme for Spherical Receiver

The input symbol alphabet is given by $\mathbb{X}$ and has the cardinality M. Let α_i is the a priori probability of sending $x_i \in \mathbb{X}$. Let H_i represent the hypothesis that the x_i symbol is transmitted, and let the number of molecules released to transmit x_i be β_i, where $\beta_0 < \beta_1 < \cdots < \beta_{M-1}$. The conditional probability density function (PDF) of received molecules i.e. N_k given the hypothesis H_i is true and denoted as $f(N_k/H_i)$. If β_i molecules are transmitted, the variable $N_i, 1 \leq i \leq k-1$, follows a binomial distribution having β_i and p_k as the parameters , when the hypothesis H_i is true. Then we approximate the distribution of N_0 for large β_i by Gaussian distribution with mean μ_{N_0}

$$\mu_{N_0} = \beta_i p_k, \tag{10}$$

and variance $\sigma^2_{N_0}$

$$\sigma^2_{N_0} = \beta_i p_k \left(1 - p_k\right). \tag{11}$$

In our proposed model M-ary modulation is used to make the interference independent of the transmitted symbols, x_i. We can average out the number of molecules transmitted from previous r slots arriving in current slot i.e. N_i to obtain mean μ_{N_i}

$$\mu_{N_i} = \sum_{k=0}^{r-1} \left(\sum_{i=0}^{M-1} \alpha_i \beta_i p_k \right), \tag{12}$$

and variance $\sigma^2_{N_i}$ as

$$\sigma^2_{N_i} = \sum_{k=0}^{r-1} \left(\mathbb{E}[N_i^2] - \left(\sum_{i=0}^{M-1} \alpha_i \beta_i p_k \right)^2 \right), \tag{13}$$

with

$$\begin{aligned} \mathbb{E}[N_i^2] &= \sum_{k=0}^{r-1} \left(\sum_{i=0}^{M-1} \alpha_i \left(\beta_i^2 p_k^2 + \beta_i p_k \left(1 - p_k \right) \right) \right) \\ &= \sum_{k=0}^{r-1} \left(\sum_{i=0}^{M-1} \alpha_i \beta_i p_k \left(1 + \left(\beta_i - 1 \right) p_k \right) \right), \end{aligned} \tag{14}$$

where $\mathbb{E}[.]$ represents the expectation operator.

Using (7), the kth slot received molecular count, N_k, is expressed as random variables, each following the independent Gaussian distribution. Hence, we can consider Gaussian approximations for the conditional PDF of received molecules $f(N_k/\beta_i)$, where, through (9), the respective mean μ_i and variance σ_i^2 can be represented as follows

$$\begin{aligned} \mu_i &= \mu_{N_0} + \mu_{N_i} \\ \sigma_i^2 &= \sigma^2_{N_0} + \sigma^2_{N_i}, \end{aligned} \tag{15}$$

from the above calculation μ_i and σ_i^2 are mathematically same as μ_{N_k} and $\sigma^2_{N_k}$, respectively. Then maximum likelihood (ML) decision rule is used at the receiver to detect the kth slot symbol, $\hat{x_k}$, i.e.,

$$\hat{x_k} = arg \max_{x_i} f(N_k/H_i)\,. \tag{16}$$

The Maximum likelihood decision rule diminishes to pairwise comparisons of conditional PDFs [27], and the decision threshold $\eta_1, \ldots, \eta_M - 1$ can be calculated through

$$\begin{aligned} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(\eta_i - \mu_i)^2}{2\sigma_i^2} \right) = \frac{1}{\sqrt{2\pi\sigma_{i-1}^2}} \exp\left(-\frac{(\eta_i - \mu_{i-1})^2}{2\sigma_{i-1}^2} \right), \\ i = 1, ..., M-1\,. \end{aligned} \tag{17}$$

so the threshold can be computed as

$$\begin{aligned} \eta_i = \frac{\sigma_i \sigma_{i-1} \sqrt{\left(\mu_i - \mu_{i-1} \right)^2 + 2\left(\sigma_i^2 - \sigma_{i-1}^2 \right) \ln L_i} + K_i}{\sigma_i^2 - \sigma_{i-1}^2}, \\ i = 1, \ldots, M-1\,. \end{aligned} \tag{18}$$

where

$$L_i = \frac{\sigma_i}{\sigma_{i-1}}, K_i = \sigma_i^2 \mu_{i-1} - \sigma_{i-1}^2 \mu_i \,. \tag{19}$$

The above mentioned threshold η_i satisfies the condition i.e, $\mu_{i-1} < \eta_i < \mu_i$ for $i = 1, \ldots, M-1$. If the number of received molecules N_k comply with $N_k < \eta_1$, then "0" is the received symbol, and to get $\}\}M-1\varepsilon$ received symbols the $N_k > \eta_{M-1}$, and, if $\eta_i < N_k < \eta_{i+1}$ then received symbol will be "i" for $i = 1, \ldots, M-2$. With the help of all these calculation the channel capacity is calculated and analyzed.

5 Bounds on Capacity

In this section, the channel capacity of molecular communication using spherical receiver for different modulation scheme discussed in the Sect. 3 is evaluated, also the impact of other parameters on channel capacity will be discussed here.

As discussed in the above section, the count of molecules that are transmitted and received in the kth slot are represented by β_i and N_k, respectively, and their conditional probability $Pr(N_k = y_j|\beta_i = x_i)$ is given by $f(y|x)$. Where, $f(y|x)$ is same as $f(N_k|\beta_i)$ with the a priori probability $Pr(\mathbb{X}_k = x_i)$ denoted as α_i. The mutual information between N_k and β_i is represented by

$$I(N_k : \beta_i) = \sum_y \sum_x \alpha_i f(y|x) \log_2 \frac{f(y|x)}{\sum_y \alpha_i f(y|x)} \,. \tag{20}$$

where

$$f(y|x) = Q\left(\frac{\eta_i - \mu_i}{\sigma_i}\right) - Q\left(\frac{\eta_{i+1} - \mu_i}{\sigma_i}\right), \quad i = 0, \ldots, M-1 \,. \tag{21}$$

with μ_i and σ_i are given by (15), and $\eta_1, \ldots, \eta_{M-1}$ is given by (17), respectively, with $\eta_0 = -\infty, \eta_M = \infty$.

This above stated mutual information (20) is maximized over all possible a priori probabilities to get the channel capacity C_s, i.e,

$$C_s = \max_{\{\alpha_i\}} I(N_k : \beta_i)/T_s \, symbols/slots \,, \tag{22}$$

where α_i stands for the set of a priori probabilities.

The input symbols for this scheme are given by $\mathbb{X} = \{0, 1 \ldots, M-1\}$. Hence, the mutual information $I(N_k : \beta_i)$ between input β_i and output N_k for kth slot of M-ary modulated scheme is given as equation given below

$$I(N_k : \beta_i) = \sum_{y=0}^{M-1} \sum_{x=0}^{M-1} \alpha_i f(y|x) \log_2 \frac{f(y|x)}{\sum_{y=0}^{M-1} \alpha_i f(y|x)} \,. \tag{23}$$

To maximize this mutual information Lagrange multiplier is used with respect to the constraint $\alpha_0 + \alpha_1 + \cdots + \alpha_{M-1} = 1$. For this case, the Lagrange function is expressed as

$$\wedge\left(\alpha_0,\ldots,\alpha_{M-1},\lambda\right)=I\left(N_k:\beta_i\right)+\lambda\left(\alpha_0+\cdots+\alpha_{M-1}-1\right), \tag{24}$$

where Lagrange multiplier is denoted by λ. This function is partially derivated with respect to $\alpha_i, 0\le i\le M-1$ and equated to zero to get the following set of equations

$$\sum_{y=0}^{M-1}\left[\sum_{x=0}^{M-1}\alpha_i\frac{\partial f(y|x)}{\partial\alpha_i}\log_2\frac{f(y|x)}{\sum_{x=0}^{M-1}\alpha_i f(y|x)}+f(y|i)\log_2\frac{f(y|i)}{\sum_{m=0}^{M-1}\alpha_i f(y|x)}\right], \tag{25}$$

where $f\left(N_k|\beta_i\right)$, μ_i and σ_i are also the dependent functions of α_i, i.e.,

$$\begin{aligned}\frac{\partial f(y|x)}{\partial\alpha_i}=\frac{1}{\sqrt{2\pi\sigma_i^2}}\frac{\partial\sigma_i}{\alpha_i}\Bigg[&\exp-\frac{(\eta_y-\mu_i)^2}{2\sigma_i^2}\left(\frac{\partial\eta_y}{\partial\alpha_i}-\frac{\partial\mu_i}{\partial\alpha_i}\right)\\&-\exp-\frac{(\eta_{y+1}-\mu_i)^2}{2\sigma_i^2}\left(\frac{\partial\eta_{y+1}}{\partial\alpha_i}-\frac{\partial\mu_i}{\partial\alpha_i}\right)\Bigg],\end{aligned} \tag{26}$$

with

$$\frac{\partial\mu_i}{\partial\alpha_i}=\beta_i\sum_{i=0}^{k-1}p_i, \tag{27}$$

$$\frac{\partial\sigma_i}{\partial\alpha_i}=\frac{\beta_i}{2\sigma_i}\beta_i\sum_{i=0}^{k-1}p_i\left(2+\left(\beta_i-1\right)p_i-2\mu_{N_i}\right), \tag{28}$$

$$\begin{aligned}\frac{\partial\eta_i}{\partial\alpha_i}=\frac{1}{\sigma_i^2-\sigma_{i-1}^2}\Bigg[&2\sigma_i\frac{\partial\eta_i}{\partial\alpha_i}\left(\mu_{i-1}-\mu_i\right)+\left(\sigma_i^2-\sigma_{i-1}^2\right)\frac{\partial\mu_i}{\partial\alpha_i}\\&+\left(\sigma_i\frac{\partial\sigma_{i-1}}{\partial\alpha_i}+\frac{\partial\sigma_i}{\partial\alpha_i}\right)\sqrt{\left(\mu_i-\mu_{i-1}\right)^2+2\left(\sigma_i^2-\sigma_{i-1}^2\right)\ln k_i}.\\&+\frac{\sigma_i\sigma_{i-1}\left(\frac{2}{k_i}\frac{\partial k_i}{\partial\alpha_i}\left(\sigma_i^2-\sigma_{i-1}^2\right)\right)}{2\sqrt{\left(\mu_i-\mu_{i-1}\right)^2+2\ln k_i\left(\sigma_i^2-\sigma_{i-1}^2\right)}}\Bigg],\end{aligned} \tag{29}$$

and

$$\frac{\partial k_i}{\partial\alpha_i}=\frac{1}{\sigma_{i-1}}\frac{\partial\sigma_i}{\partial\alpha_i}-\frac{\sigma_i}{\sigma_{i-1}}\frac{\partial\sigma_{i-1}}{\partial\alpha_i}. \tag{30}$$

The optimum value of the constraint used in Eq. (24) is obtained by solving the Eq. (25) mathematically and the Lagrange multiplier $\left(\bar{\alpha_0},\bar{\alpha_1},\cdots,\bar{\alpha_{M-1}},\bar{\lambda}\right)$. The maximum values of the mutual information (20) and Lagrange function (24) are achieved at the optimum value.

$$\begin{aligned}=\sum_{i=0}^{M-1}\bar{\alpha_i}\Bigg[\sum_{y=0}^{M-1}\sum_{x=0}^{M-1}&-\bar{\alpha_x}\frac{\partial f(y|x)}{\partial\alpha_i}\log_2\frac{f(y|x)}{\sum_{x=0}^{M-1}\bar{\alpha_x}f(y|x)}\\&+1-\bar{\lambda}\Bigg].\end{aligned} \tag{31}$$

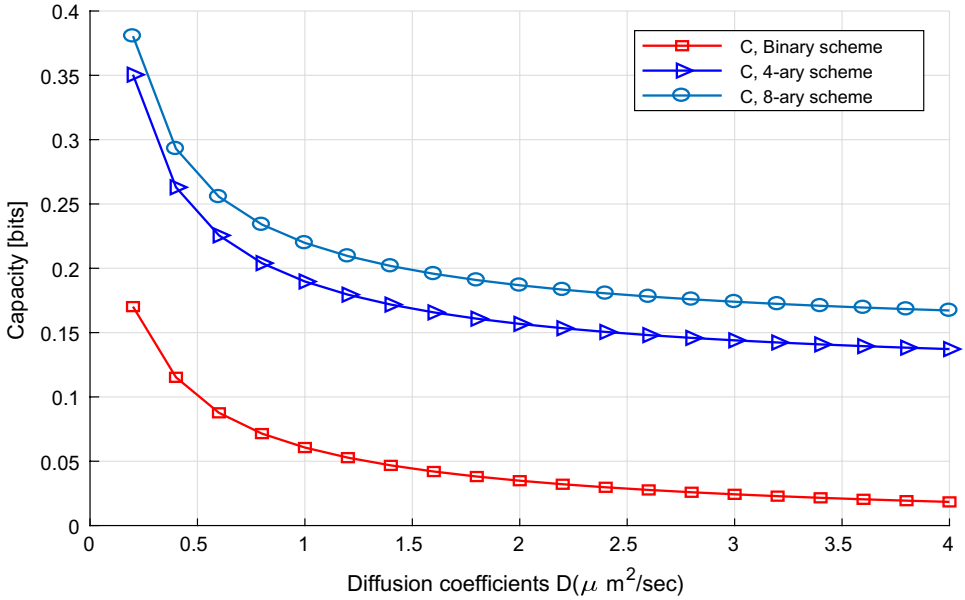


Fig. 3 Spherical receiver capacity variation with diffusion coefficient D when $\alpha_i = 0.1$ and $\beta_i = 1000N$

Above stated Eq. (31) denotes is the capacity of the kth slot which is the maximum value of mutual information achieved by using the above obtained optimum values.

6 Numerical Results and Discussion

This section contains the numerical evaluation effect of different parameters on channel capacity. The behavior of channel capacity for M-ary modulation scheme is also evaluated to figure out how channel capacity can be increased with amplitude modulation. Further, a three-way correlation has been derived for different design parameters like diffusion, radius of spherical receiver and number of transmitted molecules which gives a complete analysis of channel capacity.

Figure 3 is showing the variation of channel capacity with diffusion coefficient D. At constant number of transmitted molecules $\beta_i = [1000, 2000, 3000, 4000]$ the channel capacity first decreases sharply with increase in diffusion coefficient D until $D \leqslant 1 \times 10^{-10}$ m^2/s and the gradually for $D \geqslant 1.5 \times 10^{-10}$ m^2/s. Increase in the value of diffusion coefficient will add the randomness in the system which in turn decreases the channel capacity. It is also seen that channel capacity is better for M-ary ($M = 4, 8$) with value of 0.35 and 0.4 *bits/slot* than the binary scheme having value 0.19 *bits/slot*.

To analyze the effect of variation in the number of molecules transmitted β_i for different values of a priori probability $\alpha_i = 0.1, 0.2$ and diffusion coefficient $D = 1.5 \times 10^{-10}$ m^2/s on channel capacity of spherical receiver for both binary and M-ary ($M = 4, 8$) modulation scheme Fig. 4 is shown. The channel capacity is increasing gradually and then sharply at $\beta_i \geqslant 3000$. As the β_i increases, the count of signalling molecules increase but simultaneously it will also increases the number of stray molecules which will not let the capacity to increase further by increasing the β_i. It also shows that M-ary modulation is showing higher capacity than binary scheme for the higher value of a priori probability.

In Fig. 5, the trend of channel capacity is seen with the variation of spherical receiver radius S_r for different values of diffusion constant $D = 0.1$ mm^2/s and 0.2 mm^2/s for M-ary modulation schemes. As the radius of spherical receiver increases the molecules will get a

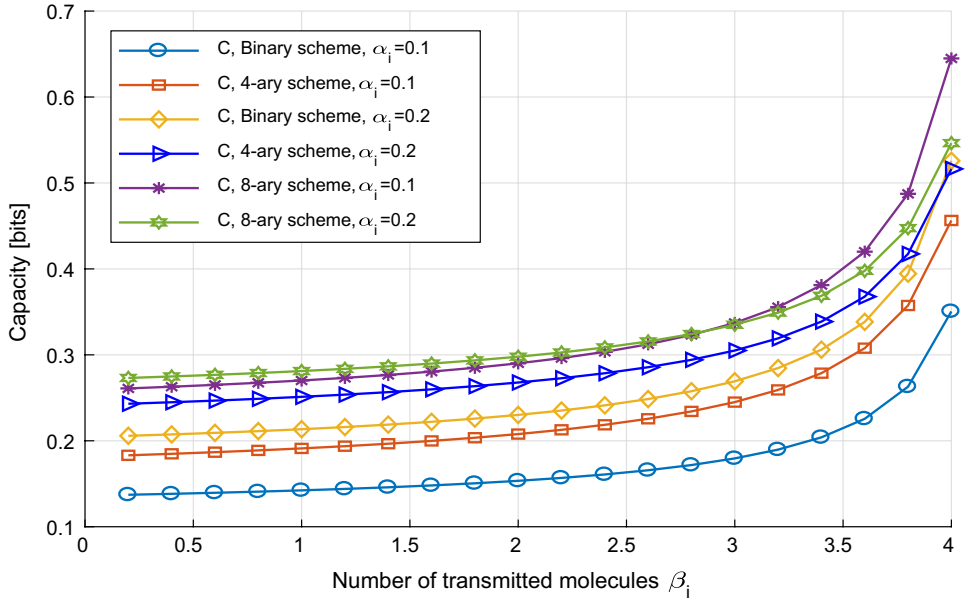


Fig. 4 Variation of spherical receiver capacity with Number of transmitted molecules β_i when $D = 1 \times 10^{-10}$ m^2/s

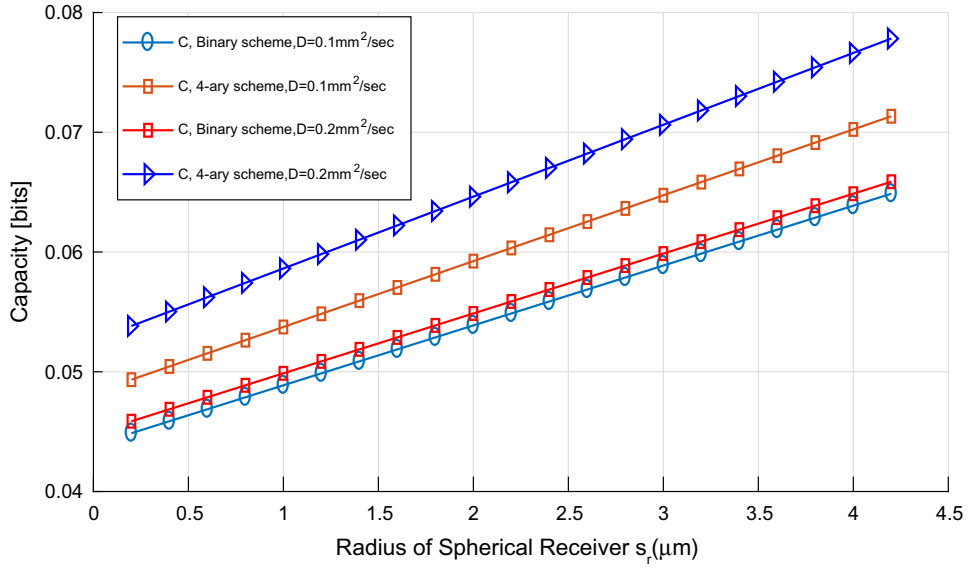


Fig. 5 Variation of capacity with radius of spherical receiver when $\alpha_i = 0.1$

larger surface area to hit the receiver, which in turn will increase the channel capacity. It's been observed that channel capacity is better for M-ary ($M = 4$) than the binary scheme even for smaller radius devices.

In Fig. 6, the trend of channel capacity is seen with the variation of noise variance, σ^2 for different values of diffusion constant $D = 0.1$ mm^2/s and 0.2 mm^2/s for M-ary modulation schemes. Result shows that channel capacity increases w.r.t. noise variance as diffusion increases, this means diffusion supports communication as if there is no diffusion, molecules will always be stationary and never reaches at the receiver. The channel capacity variates minutely from 0.1*bits* to .045*bits* with noise variance.

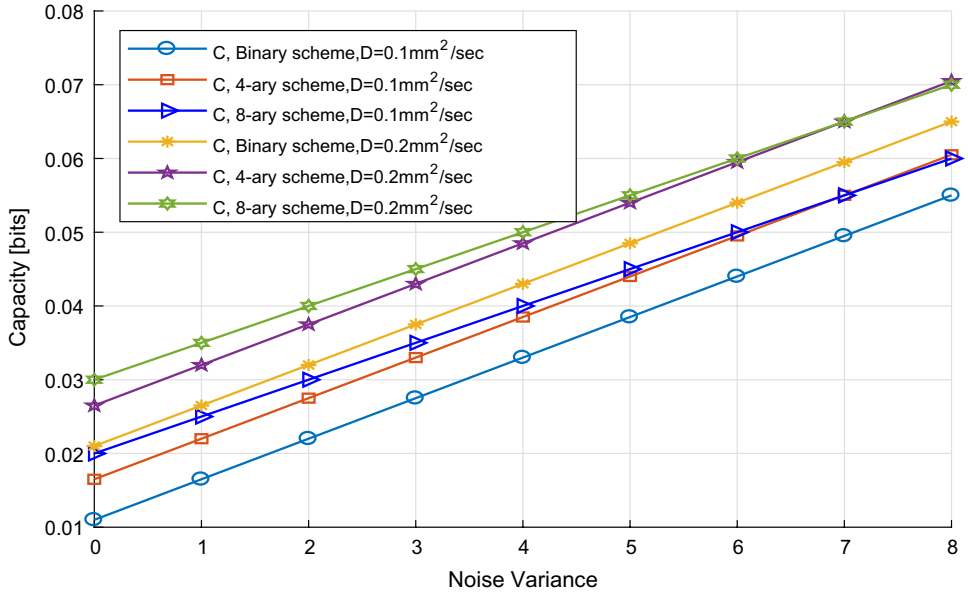


Fig. 6 Variation of spherical receiver capacity with noise variance

7 Conclusion

In this research work, a purely diffusion based molecular communication channel for nanonetworks having finite spherical shape receiver is modeled and simulated. To increase the practical feasibility of the system, we have derived a mathematical expression to calculate the first hitting probability for a spherical shaped absorbing receiver. For this, we have considered the finite lifetime molecules that can be absorbed in a given time slot. We have considered a generalized M-ary modulation scheme for molecular communication between two nano-devices placed in a drift free fluid medium. According to author's knowledge, for the first time a finite spherical shape receiver is considered for a diffusion based molecular channel. Further, an expression for channel capacity is calculated by taking all possible sources of irregularities into account.

The derived mathematical expressions for the capacity provides an absolute end-to-end evaluation of molecular propagation for various medium parameters like diffusion, radius of receiver and number of molecules transmitted. Numerical results revels a interesting two-way correlation between the values of parameter and the performance of derived system. On the whole, the optimum values of parameters are obtained to set the stage for establishing a required reliable molecular communication for finite shape receiver system in nanonetworks.

This research can help the technology providers of nano-devices to evaluate the optimum values of different design parameters like diffusion, radius of receiver and number of transmitted molecules to obtain desired channel capacity. It can be concluded that maximum channel capacity can be achieved by increasing the level of amplitude modulation even for nano-device of scale 3–5 μm and the effect of diffusion on channel capacity can also be stabilized by this. If the diffusion is low then a desired value of channel capacity can also be obtained even for binary amplitude modulation scheme.

Author Contributions Author agree with the contribution of co-author.

Funding No funding available.

Data availability No data and materials required in this paper.

Declarations

Conflict of interest authors have no conflict of interest.

Ethical approval authors agree with the journal ethical policy.

Consent for publication Authors have all the consent for publication.

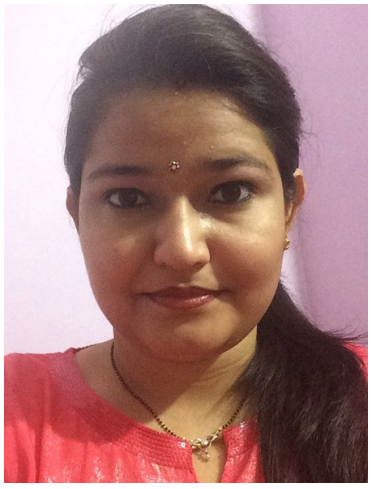
Code avalability There is no code associated.

References

1. Llatser, I., Demiray, D., Cabellos-Aparicio, A., Altilar, D. T., & Alarcón, E. (2014). N3Sim: Simulation framework for diffusion-based molecular communication nanonetworks. *Simulation Modelling Practice and Theory, 42,* 210–222.
2. Nakano, T., Suda, T., Moore, M., Egashira, R., Enomoto, A., & Arima, K. (2005). Molecular communication for nanomachines using intercellular calcium signaling. In *5th IEEE conference on nanotechnology*, pp. 478–481, IEEE.
3. Farsad, N., Yilmaz, H. B., Eckford, A., Chae, C.-B., & Guo, W. (2016). A comprehensive survey of recent advancements in molecular communication. *IEEE Communications Surveys and Tutorials, 18*(3), 1887–1919.
4. Kuran, M. S., Yilmaz, H. B., Tugcu, T., & Akyildiz, I. F. (2012). Interference effects on modulation techniques in diffusion based nanonetworks. *Nano Communication Networks, 3*(1), 65–73.
5. Kadloor, S., Adve, R. S., & Eckford, A. W. (2012). Molecular communication using Brownian motion with drift. *IEEE Transactions on Nanobioscience, 11*(2), 89–99.
6. Kim, N. R., & Chae, C. B. (2013). Novel modulation techniques using isomers as messenger molecules for nano communication networks via diffusion. *IEEE Journal on Selected Areas in Communication, 31*(12), 847–856.
7. Kuran, M. S., Yilmaz, H. B., Tugcu, T., & Akyildiz, I. F. (2011). Modulation techniques for communication via diffusion in nanonetworks. In *Proceedings of IEEE international conference communication*, Kyoto, Japan, pp. 1–5.
8. Mahfuz, M. U., Makrakis, D., & Mouftah, H. T. (2010). On the characterization of binary concentration-encoded molecular communication in nanonetworks. *Nano Communication Networks, 1*(4), 289–300.
9. Bicen, A., & Akyildiz, I. (2013). System-theoretic analysis and least-squares design of microfluidic channels for flow-induced molecular communication. *IEEE Transactions on Signal Processing, 61*(20), 5000–5013.
10. Hiyama, S., & Moritani, Y. (2010). Molecular communication: Harnessing biochemical materials to engineer biomimetic communication systems. *Nano Communication Network, 1*(1), 20–30.
11. Nakano, T., Suda, T., Moore, M., Egashira, R., Enomoto, A., & Arima, K. (2005). Molecular communication for nanomachines using intercellular calcium signaling. In: *Fifth IEEE conference on nanotechnology*, Vol. 2, pp. 478–481.
12. Singhal, A., Mallik, R. K., & Lall, B. (2015). Performance analysis of amplitude modulation schemes for diffusion-based molecular communication. *IEEE Transactions on Wireless Communications, 14*(10), 5681–5691.
13. Eckford, A. W. (2007). Nanoscale communication with Brownian motion. In *Processing Conference Information Science System, Baltimore, MD, USA*, (pp. 160–165).
14. Einolghozati, A., Sardari, M., Beirami, A., & Fekri, F. (2011). Capacity of discrete molecular diffusion channels. In *Proceedings of IEEE international symposium on information theory (ISIT)*, St. Petersburg, Russia, pp. 723–727.
15. Pandey, N., Mallik, R. K., & Lall, B. (2019). Performance analysis of diffusive molecular timing channels. *IET Communications, 13*(18), 3059–3067.
16. Srinivas, K. V., Eckford, A. W., & Adve, R. S. (2012). Molecular communication in fluid media: The additive Inverse Gaussian noise channel. *IEEE Transactions on Information Theory, 58*(7), 4678–4692.

17. Li, H., Moser, S. M., & Guo, D. (2014). Capacity of the memoryless additive inverse Gaussian noise channel. *IEEE Journal on Selected Areas in Communications, 32*(12), 2315–2329.
18. Mahfuz, M. U., Makrakis, D., & Mouftah, H. T. (2014). Strength-based optimum signal detection in concentration-encoded pulse-transmitted OOK molecular communication with stochastic ligand-receptor binding. *Simulation Modelling Practice and Theory, 42,* 189–209.
19. Pandey, N., Mallik, R. K., & Lall, B. (2017). Truncated Lévy statistics for diffusion based molecular communication. In: *IEEE global communications conference*, pp. 1–6.
20. Nakano, T., Okaie, Y., & Liu, J. Q. (2012). Channel model and capacity analysis of molecular communication with Brownian motion. *IEEE Communications Letters, 16*(6), 797–800.
21. Pandey, N., Joshi, S., & Mallik, R. K. (2021). Characterizing the probability of collision between information particles in molecular communications. *IEEE Wireless Communications Letters, 10*(6), 1252–1255.
22. Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., et al. (2010). *Essential cell biology*. New York: Garland Science.
23. Yilmaz, H. B., Heren, A. C., & Tugcu, T. (2014). Three-dimensional channel characteristics for molecular communications with an absorbing receiver. *IEEE Communications Letters, 18*(6), 929–932.
24. Moore, M. J., Nakano, T., Enomoto, A., & Suda, T. (2012). Measuring distance from single spike feedback signals in molecular communication. *IEEE Transactions on Signal Processing, 60*(7), 3576–3587.
25. ShahMohammadian, H., Messier, G. G., & Magierowski, S. (2013). Blind synchronization in diffusion-based molecular communication channels. *IEEE Communication Letter, 17*(11), 2156–2159.
26. ShahMohammadian, H., Messier, G. G., & Magierowski, S. (2012). Optimum receiver for molecule shift keying modulation in diffusion-based molecular communication channels. *Nano Communication Network, 3*(3), 183–195.
27. Trees, H. L. V. (2001). *Detection, estimation, and modulation theory, Part I*. Newark: Wiley.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Garima Singh The author Garima Singh is research scholar at Delhi Technological University, Delhi. She obtained her BE degree in Electronics and Communication Engineering from Sharda University in 2013 and ME (Electronics and Communication) from Jaypee Institute of Information and Technology in 2015. She has been the topper throughout her academic carrier and has the distinction of receiving an A+ for her ME thesis. She has spent over 5 years towards research and teaching. Her professional experience and research are in the area of Cognitive Radio, Cooperative Communication, Green Smart Technology, Internet of Things, Optical and Wireless Communication System and Networks. She has several publications in SCI indexed International Journals and conferences including book chapters in leading international press like IGI, CRC and Elsevier.



Dr. Gurjit Kaur a bright scholar, gold medalist throughout including B.Tech and M.Tech and president awardee has spent over 15 years of her academic career towards research and teaching in the field of Electronics and Communication. She is having distinction of receiving a Gold medal by former President of India A P J Abdul Kalam for being overall topper of the Punjab Technical University, Jalandhar by securing 82% marks. She has been a topper throughout her academic career. She earned her Ph.D. degree from Panjab University, Chandigarh in 2010 and her M. Tech from PEC University of Technology, Chandigarh in 2003 with distinction. Presently Dr Kaur is working as an Associate Professor at Delhi Technological University, Delhi India. Prior to that she was working as an Assistant Professor in the School of Information and Communication Technology, Gautam Buddha University, Greater Noida, India. Her research interests include Optical CDMA, Wireless Communication system, high-speed interconnect and IOT. Her name has been listed in Marquis Who's Who in Science and Technology, USA. She has published more than 70 papers in journal and conferences, authored one book on Optical Communication, and one book chapter on WiMax of Florida Atlantic University, published by CRC Press. Besides this, she has presented her research work as short courses/tutorials in many national and international conferences. She served as a reviewer of various journals like IEEE transactions on communication etc. Dr Kaur is recipient of Prof. Indira Parikh 50 Women in Education Leader Award in world Education Congress, Mumbai, India in 2017. She also received Bharat Vikas Award by Institute of Self Reliance in National Seminar on Diversity of Cultural and Social Environment at Bhubneswar, Odisha, in 2017. She worked as an convener for two international conferences i.e. International ICIAICT 2012 which was organized by CSI, Noida Chapter and International conference EPPICTM 2012 which was organized in collaboration with MTMI, USA, University of Maryland Eastern Shore, USA and Frostburg State University, USA at School of Information and Communication Technology, Gautam Buddha University, Greater Noida, India.

Journal of Engg. Research ICARI Special Issue pp. 121-130 **DOI : 10.36909/jer.ICARI.15291**

Modelling Of Wind Turbine Magnet for Magnetic Abrasive Finishing and Magnetic Field Assisted Abrasive Flow Machining Process

Anant Bhardwaj, Krovvidi Srinivas* and Rajiv Chaudhary

Department of Mechanical Engineering, Delhi Technological University, Shahabad Daulatpur, Main Bawana Road Delhi, India
** Corresponding Author : srinivaskrovvidi@dtu.ac.in*

ABSTRACT

Magnetic Abrasive Finishing is a significant process for finishing up to the micro-level. However, with the advancement of technology and hybrids like Viscoelastic magnetic abrasive Finishing and Magnetic abrasive Flow machining, it has become a nano finishing process. To improve the finishing process, the researchers have made a Model and tested the feasibility of the wind turbine magnet in Finishing. The Maxwell simulations were done for the cylindrical Specimen of Brass, Steel Aluminum. The simulations results were in accordance with the fact that the proposed wind turbine magnet may be used for the simulations.

Keywords: Magnetic field assisted abrasive flow machining; Viscoelastic magnetic abrasive finishing.

INTRODUCTION

The magnet used in the wind turbine is powerful .The research on magnetic abrasive Finishing is mostly confined to the electromagnet or the particular shape magnet. The research's main objective is to open a new area in the finishing process by magnet where wind magnets play a vital role. Magnetic abrasive Finishing is an effective process. Many researchers have worked on it and (Wang et al. 2004) used a magnetic abrasive finishing process on the tube's inner surface, which produces a highly finished inner surface of the critical application tube. The researchers calculated the stress on the abrasive grain and are analyzed it along with the grinding mechanism. The researchers used the magnetic abrasive Machining process for calculating the Finishing on the austenitic stainless steel tubing. After the successful experimentation, the researchers observed that a big size magnetic abrasive particle has a more significant normal force than smaller particles. Also, the normal force is proportional to the exciting current. The normal force is the essential factor for deciding the MRR compared with other factors like exciting current, grid size etc. (Rohit et al. 2017) developed a new magnetic abrasive. The researcher developed the magnetic Particle by mixing abrasive, ferromagnetic powder and a particular type of adhesive bonding. For comparing the created magnetic Particle, the researchers made another two particles, one by sintering and the other by simple mixing techniques. When the researcher compared magnetic abrasive particles by using Brass as the workpiece. After reaching the results, it is found that the magnetic Particle made by mixing silicon carbide and abrasive yields 18 per cent of the surface improvement. In contrast, the abrasive made by adhesive bond iron-sic the percentage improvement is 42 per cent, and for the Particle made by sintered iron-Sic, the surface roughness observed was 49 per cent. (Kang et al. 2017) Summarized the machining process and the magnetic particles and reviewed their machining process and the abrasive magnetic finishing processes. The researcher concluded that Sintered Magnetic Abrasive have greater life compared to the mixed MAPs. Moreover, Sintered Magnetic Abrasive Particles has an edge over mixed merely MAPs when MRR is taken into account. Also,

the surface finish for the Sintered magnetic abrasive is more than that of the sample size. The researchers' other key observation was the wet abrasive give better finishing results compared to the dry abrasive. (Muhammad et al. 2016) Conducted there to experiment on the tube's inner surface finish. The researcher used the electrolysis method combined with magnetic abrasive Finishing, making it suitable for effective machining. The researcher claimed that the process of Electrolysis is effective for the machining process. Still, the main disadvantage is the formation of the pit residual and the formation of oxidation level on the surface, resulting in the degradation of the character's quality. However, magnetic abrasive Finishing over the surface is an effective method of removing the oxidation layer over the surface and effectively finishing. (Srikande et. al. 2018) Compared the magnetic abrasive flow machining with the conventional abrasive flow machining process to Perform the Finishing operations on dog clutch with internal splines. The researcher compared the finishing process by taking five parameters: Machining time (per work piece), Setup Time, Surface Finish, Productivity, and accuracy. According to the researchers, the machining time for the MAFM process is 3-minute compare to the machining time of 5 minutes for the conventional one. The surface finish value obtained for the MAFM process is up to 3-5 microns, whereas, for the conventional process, it is 0.05 to 1 micron. The material removal rate for the MAFM is also high, and it went up to 0.020 to 0.030 mm, and for the conventional machining process, it was 0.008 to 0.01 mm. Finally, the researchers concluded that the productivity is higher for the MAFM process than the conventional AFM process. (Chen et al. 2010) Prepared a magnetic abrasive by the process of sintering and pressing. The researchers conducted experiments on the developed abrasives. They concluded that the pressing and sintering abrasive has a good finishing performance, Effective improvement in surface quality, long service life, adaptability, and improvement in the surface finish. (Judal et al. 2016) studied CMAF, which is a cylindrical Magnetic abrasive finishing process. The researchers developed the FEM model and simulated the process, and verified it by the literature present. The researchers highlighted the advantage of the cylindrical Magnetic abrasive finishing process over the Conventional Magnetic Abrasive Finishing. (Srinivas et al. 2019) studied the permanent magnet's various shape and selected four among them. The researchers model the 4 cases on the Ansys Maxwell and found that the Arc magnet shows the other magnet's best results. The distribution of magnetic field intensity over the workpiece's surface was better in this case. The value of Magnetic Field intensity was highest in this case. In another investigation, (Srinivas et.al.2019) took the arc magnet commercially available in the market and studied the magnet's effect by arranging it at a different pole. the researchers took two arc magnet types with two central angles, with one having an angle of 60 degrees and the other having an angle of 90.The research's important inferences were that the magnet's central angle is one of the crucial parameters in the finishing effect of the workpiece; among all the cases, the one with a 90-degree angle has the maximum magnetic felid intensity. In the further application, (Srinivas et al. 2019) used ANSYS Fluent and tested a viscoelastic media and computed the value of pressure and the force exerted by the viscoelastic fluid on the workpiece to be machined.(Srinivas et al. 2019) used electromagnet for the process of viscoelastic magnet abrasive finishing process and plotted a curve representing the magnetic felid variation with the current, and concluded that the magnetic felid is practically proportional to the current as it represents an approximate straight line. thus MAF is Key parameters in the field of finishing.

METHODOLOGY

The wind magnet in the market was modelled on CREO Software and tested on the ANSYS MAXWELL 16 software. The result is tested and based on aluminium, Brass and Steel; the magnet was used for this purpose. A hollow work-piece is taken and is kept inside the hollow magnet, whose outer diameter is about 36 mm diameter. The material of the magnet is NdFe 35 and the hollow workpiece material was taken differently. So firstly, the analysis is done on the hollow workpiece. The workpiece taken was to check the distribution, so on CREO 3.0 software, so a hollow tube of outer diameter 30 mm and the thickness of 5 mm was made of height 20 mm was drawn. For the magnet, the magnet's assembly is made with a collection of 8 Pieces with 36 mm outer diameter and 32 mm inner diameter. The assembly was made on the CREO 3.0 Software and a copy of the assembly is made is sent to ANSYS Maxwell through an IGES File and is open on ANSYS MAXWELL for simulations. The figure 1(a) yellow portion shows the magnet and the blue part represent the workpiece. Figure 1(b) shows the actual magnet.

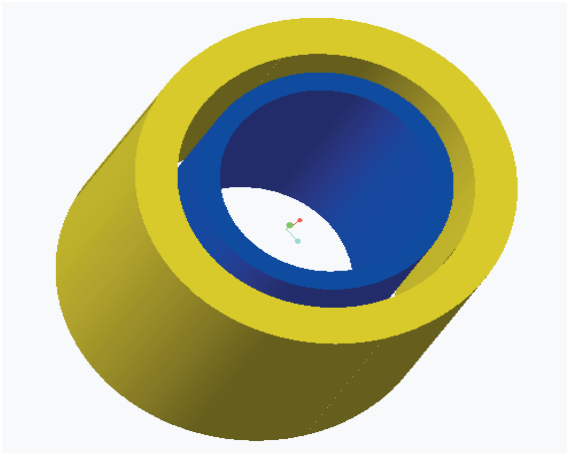


Figure 1. (a)
CREO 3.0 model for 36 diameter magnet

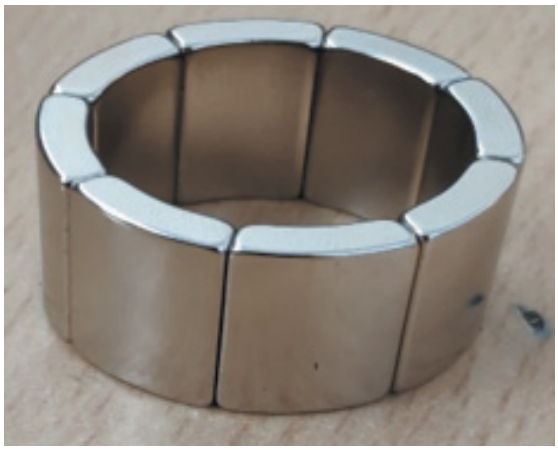


Figure 1. (b)
Actual 36 mm diameter magnet

SIMULATIONS

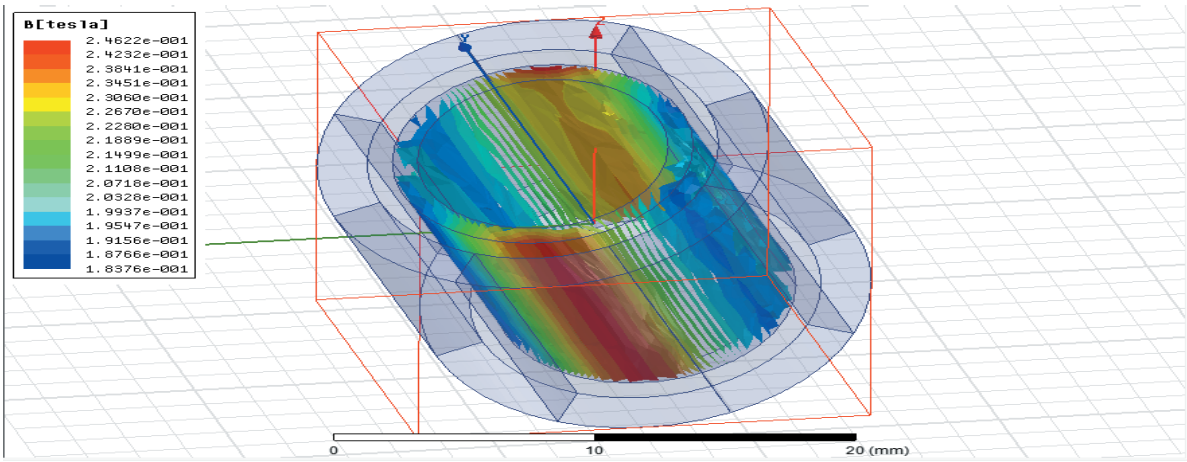


Figure 2. Magnetic Field distribution for aluminium workpiece

The above figure shows a magnet with an eight-piece of an outer diameter of 36 mm diameter and an inner diameter of 32 mm .theworkpiece is made up of a hollow tube of an outer diameter of 30 mm. A thickness of 5 mm was made. The material assign to the magnet is NDFE 35 and that of the workpiece is aluminium.The above figure shows the aluminium's magnetic field distribution with the maximum value of 0.2462 Tesla and the minimum value of 0.1837 Tesla. Aluminium is non-ferrous material. With this simulation, it is evident that the magnetic field's intensity on the surface would be less than the ferrous material. The result would be almost similar to Brass as they both share the same non-ferrous material property.

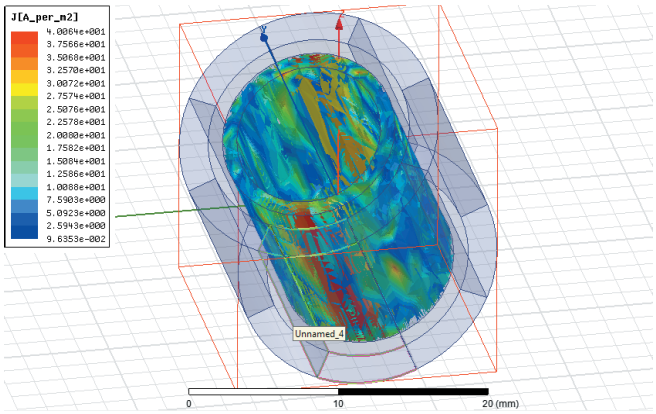


Figure 3. Current Density Distribution for aluminium work piece

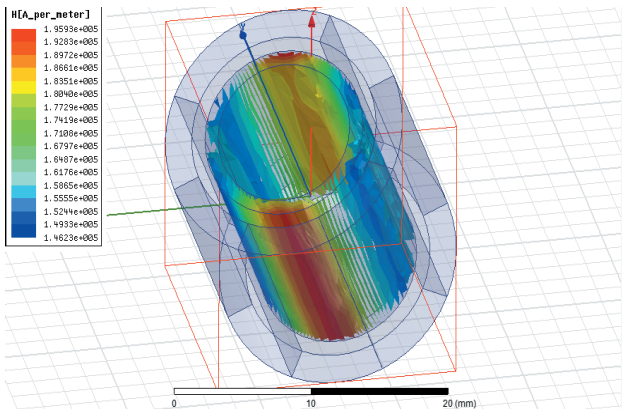


Figure 4. Magnet Field Strength Distribution for aluminium work piece

The above figure shows a magnet with eight pieces of an outer diameter of 36 mm diameter and an inner diameter of 32 mm .theworkpiece is made up of a hollow tube of an outer diameter of 30 mm, and a thickness of 5 mm was made. The material assign to the magnet is NDFE 35 and that of the workpiece is aluminium.The above figure shows the J's distribution on the aluminium with the maximum value of 40.064 (A/m2) and the minimum value of 0.09635 (A/m2). Aluminium is non-ferrous material. With this simulation, it is evident that the magnetic field's intensity on the surface would be less than the ferrous material. The result would be almost similar to Brass as they both share the same non-ferrous material property.

The above figure shows a magnet with eight pieces of an outer diameter of 36 mm diameter and an inner diameter of 32 mm .the workpiece is made up of a hollow tube of an outer diameter of 30 mm. A thickness of 5 mm was made. The material assign to the magnet is NDFE 35 and that of the workpiece is aluminium. The above figure shows the H's distribution on the aluminium with the maximum value of 19593 (A/m) and the minimum value of 14623 (A/m). Aluminium is a non-ferrous material. With this simulation, it is evident that the magnetic field's intensity on the surface would be less compared to the ferrous material. The result would be almost similar to Brass as they both share the same non-ferrous material property.

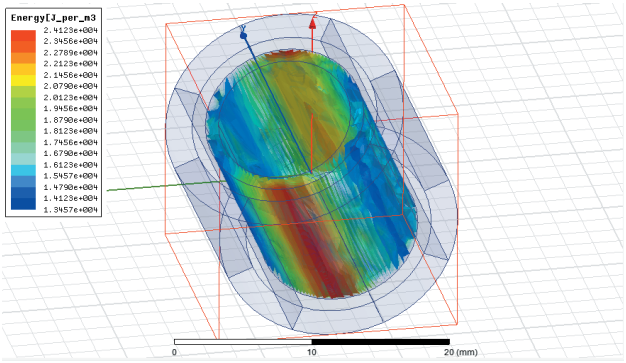


Figure 5. Energy Distribution for aluminium work piece

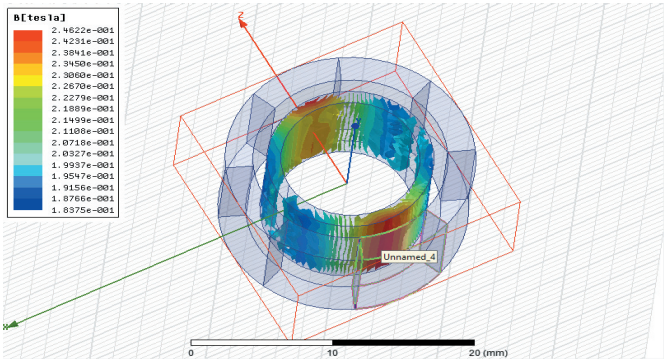


Figure 6. Magnetic field Distribution for Brass work piece

The above figure shows a magnet with eight pieces of an outer diameter of 36 mm diameter and an inner diameter of 32 mm .The workpiece is made up of a hollow tube of an outer diameter of 30 mm. A thickness of 5 mm was made. The material assign to the magnet is NDFE 35 and that of the workpiece is aluminium. The above figure shows the aluminum's energy distribution with the maximum value of 24123 (J/m3) and the minimum value of 13457 (J/m3). Aluminium is a non-ferrous material. With this simulation, it is evident that the magnetic field's intensity on the surface would be less compared to the ferrous material. The result would be almost similar to Brass as they both share the same non-ferrous material property.

The above figure shows a magnet with eight pieces of an outer diameter of 36 mm diameter and an inner diameter of 32 mm. The workpiece is made up of a hollow tube of an outer diameter of 30 mm. A thickness of 5 mm was made. The material assign to the magnet is NDFE 35 and that of a workpiece is aluminium. The above figure shows the Magnetic Field's distribution on the Brass with the maximum value of 0.2462 Tesla and the minimum value of 0.18375 Tesla. Brass is a non-ferrous material. With this simulation, it is evident that the magnetic field's intensity on the surface would be less than the ferrous material. The result would be almost similar to Brass as they both share the same non-ferrous material property.

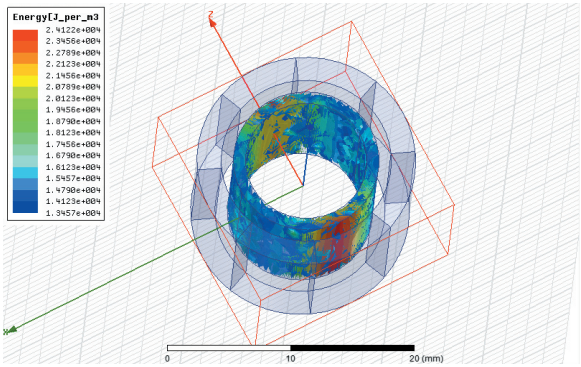


Figure 7. Energy Distribution for Brass work piece

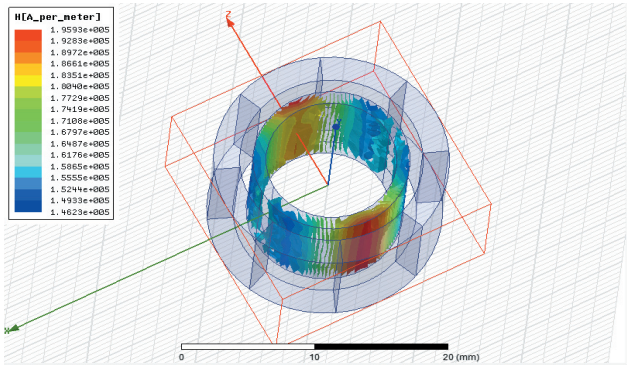


Figure 8. Magnetic Field Strength Distribution for Brass work piece

The above figure shows a magnet with eight pieces of an outer diameter of 36 mm diameter and an inner diameter of 32 mm. The workpiece is made up of a hollow tube of an outer diameter of 30 mm. A thickness of 5 mm was made. The material assign to a magnet is NDFE 35 and that of a workpiece is aluminium. The above figure shows the energy distribution on the Brass with the maximum value of 24122 (J/m3) and the minimum value of 13457 (J/m3). Brass is a non-ferrous material. With this simulation, it is evident that the magnetic field's intensity on the surface would be less compared to the ferrous material. The result would be almost similar to Brass as they both share the same non-ferrous material property.

The above figure shows a magnet with eight pieces of an outer diameter of 36 mm diameter and an inner diameter of 32 mm .The workpiece is made up of a hollow tube of an outer diameter of 30 mm. A thickness of 5 mm was made. The material assign to a magnet is NDFE 35 and that of a workpiece is aluminium. The above figure shows the H's distribution on the Brass with the maximum value of 19593 (A/m) and the minimum value of 14623 (A/m). Brass is a non-ferrous material. With this simulation, it is evident that the magnetic field's intensity on the surface would be less compared to the ferrous material. The result would be almost similar to Brass as they both share the same non-ferrous material property.

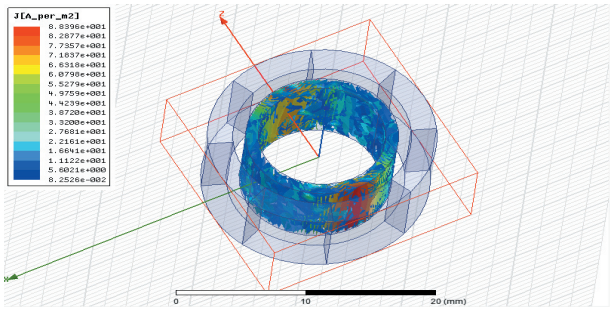


Figure 9. Current Density Distribution for Brass workpiece

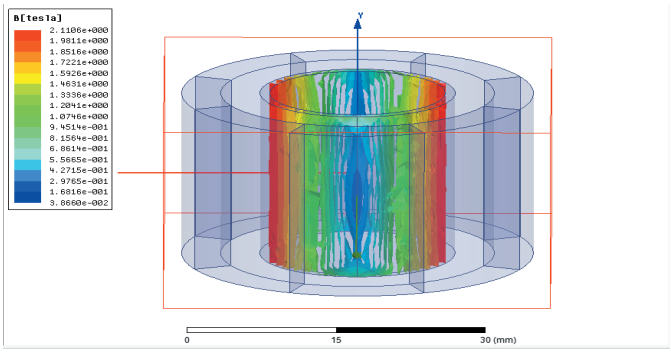


Figure 10. Magnetic Field Distribution for Steel work piece

The above figure shows a magnet with eight pieces of an outer diameter of 36 mm diameter and an inner diameter of 32 mm. The workpiece is made up of a hollow tube of an outer diameter of 30 mm. A thickness of 5 mm was made. The material assign to the magnet is NDFE 35 and that of the workpiece is aluminium. The above figure shows the J on the Brass's distribution with the maximum value of 88.396 (J/m2) and the minimum value of 0.00825626 (J/m2). Brass is a non-ferrous material. With this simulation, it is evident that the magnetic field's intensity on the surface would be less compared to the ferrous material. The result would be almost similar to Brass as they both share the same non-ferrous material property.

The above figure shows a magnet with eight pieces of an outer diameter of 36 mm diameter and an inner diameter of 32 mm .The workpiece is made up of a hollow tube of an outer diameter of 30 mm. A thickness of 5 mm was made. The material assign to the magnet is NDFE 35 and that of the workpiece is aluminium. The above figure shows the Magnetic Field's distribution on the Steel with the maximum value of 2.11 Tesla and the minimum value of 0.03886 Tesla. Steel and Steel's workpiece is a ferrous material that supports a larger magnetic field.

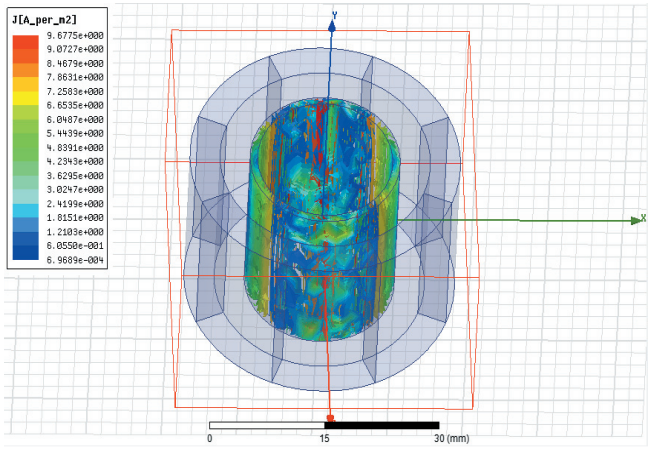


Figure 11. Current Density Distribution for Steel workpiece

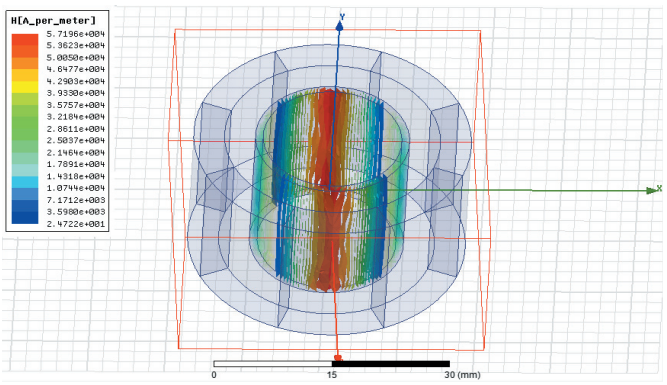


Figure 12. Magnet Field Strength Distribution for Steel workpiece

The above figure shows a magnet with eight pieces of an outer diameter of 36 mm diameter and an inner diameter of 32 mm .The workpiece is made up of a hollow tube of an outer diameter of 30 mm. A thickness of 5 mm was made. The material assign to the magnet is NDFE 35 and that of the workpiece is aluminium. The above figure shows the J on the Steel's distribution with the maximum value of 9.67 (A/m2) and the minimum value of 0.00069 (J/m2). Steel and Steel's workpiece is a ferrous material that supports a larger magnetic field.

The above figure shows a magnet with an eight-piece of an outer diameter of 36 mm diameter and an inner diameter of 32 mm. The workpiece is made up of a hollow tube of an outer diameter of 30 mm and a thickness of 5 mm was made. The material assign to a magnet is NDFE 35 and that of the workpiece is aluminium. The above figure shows the H on the Steel's distribution with the maximum value of 57196 (A/m) and the minimum value of 24.722 (J/m). The workpiece is made up of Steel, and Steel is a ferrous material that supports a larger magnetic field.

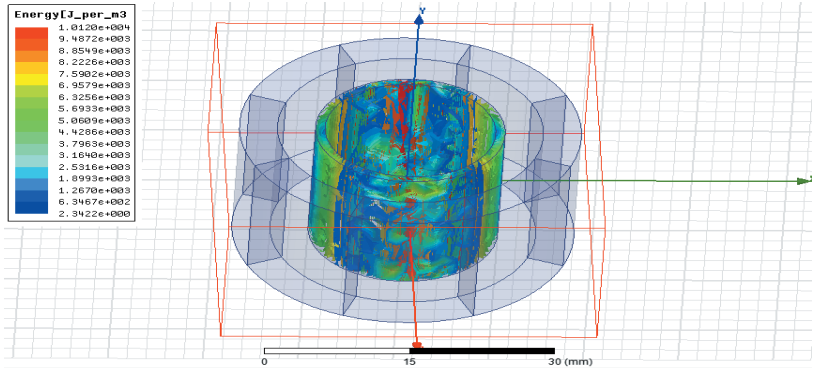


Figure 13. Energy Distribution for Steel workpiece

The above figure shows a magnet with eight pieces of an outer diameter of 36 mm diameter and an inner diameter of 32 mm. The workpiece is made up of a hollow tube of an outer diameter of 30 mm. A thickness of 5 mm was made. The material assign to a magnet is NDFE 35 and that of a workpiece is aluminium. The above figure shows Steel's energy distribution with the maximum value of 10120 (J/m3) and the minimum value of 203422 (J/m3). Brass is a non-ferrous material. With this simulation, it is evident that the magnetic field's intensity on the surface would be less compared to the ferrous material. The result would be almost similar to Brass as they both share the same non-ferrous material property.

RESULTS AND DISCUSSION

After the successful simulations on ANSYS MAXWELL, to compare the magnetic property produced by the wind turbine magnet on the internal gear made up of different material, The result obtained from simulations was tabulated. It was grouped into four sub-groups: Magnetic flux density, Magnetic field strength, current Density, and magnetic energy. Table 1 shows the values of magnetic flux density of Aluminium, Brass and Steel.

Table 1 Magnetic Flux Density of Aluminium, Brass and Steel

Material	Maximum Value	Minimum Value	Average Value
Aluminium	0.2462	0.1837	0.21495
Brass	0.246	0.1875	0.21675
Steel	2.1106	0.0386	1.0746

Figure 14 shows the Comparison of the Magnetic Flux Density of Aluminium, Brass and Steel in the graphical form, which was made using the above values.

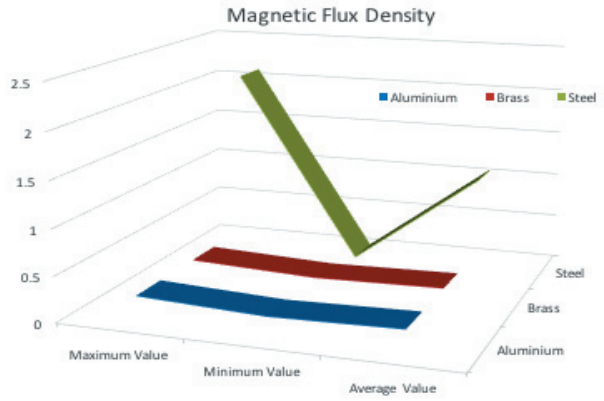


Figure 14 Comparison of the Magnetic Flux Density of Aluminium, Brass and Steel

It is evident that the value of magnetic flux density for the Steel is far better than the Brass and Aluminium as it is a ferromagnetic material and as the magnetic flux density increases, the effect of brush formation over the surface is more or, in another word the interaction of abrasive and surface rises which results in the effective machining process. Table 2 shows the value of Aluminium's Current Density, Brass and Steel obtained after the simulation.

Table 2 Comparison of Magnetic Flux Density of Aluminium, Brass and Steel

Material	Maximum Value	Minimum Value	Average Value
Aluminium	4.0064	0.0963	2.05135
Brass	8.8396	0.0825	4.46105
Steel	9.6775	0.0069	4.8422

Figure 15 shows the Comparison of the Magnetic Flux Density of Aluminium, Brass and Steel in the graphical form, which was made using the above values.

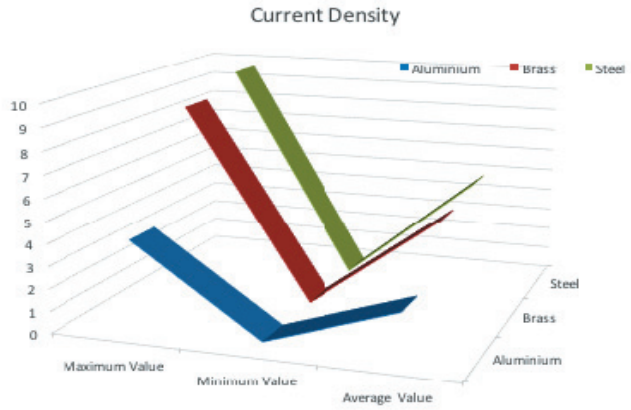


Figure 15 Comparison of the Current Density of Aluminium, Brass and Steel

After comparing the values of current Density from the simulation, it is evident that aluminium possesses the lowest value of current Density, Then Brass and Steel have the highest value. Thus making Steel the eligible candidate for the machining when electromagnets are used. Table 3 shows the value of the Magnetic Field Strength of Aluminium, Brass and Steel obtained after the simulation.

Table 3. Magnetic Field Strength of Aluminium, Brass and Steel

Material	Maximum Value	Minimum Value	Average Value
Aluminium	19539	14263	16901
Brass	19593	14632	17112.5
Steel	57196	2.4722	28599.2361

Figure 16 shows the Comparison of the Magnetic Field Strength of Aluminium, Brass and Steel in the graphical form, which was made using the above values.

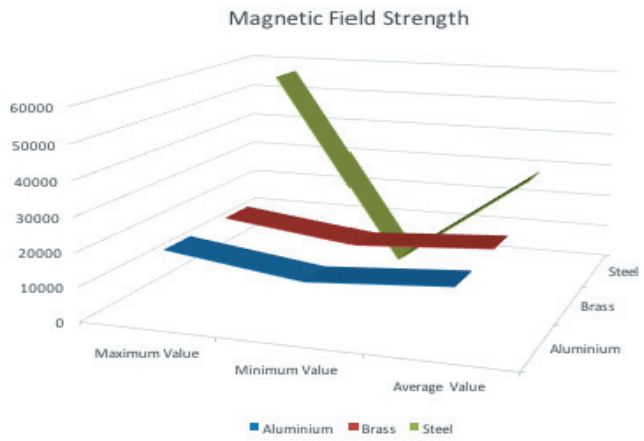


Figure 16. Comparison of the Magnetic Feild Strength of Aluminium, Brass and Steel

After comparing the Magnetic Feild Strength values from the simulation, it is evident that Brass possesses the highest value of Magnetic Feild Strength, followed by aluminium and Steel have the most negligible value. Thus making Brass the eligible candidate for the machining when electromagnets and Permanent Magnet are used. Table 4 shows the value of the Magnetic Energy of Aluminium, Brass and Steel obtained after the simulation.

Table 4. Magnetic Energy of Aluminium, Brass and Steel

Material	**Maximum Value**	**Minimum Value**	**Average Value**
Aluminium	24132	13475	18803.5
Brass	24122	13457	18789.5
Steel	10120	2.3422	5061.1711

CONCLUSION

After going through the simulations and the graphs, the researchers made the following conclusions firstly Viscoelastic Magnetic Abrasive finishing Process is the most prominent method for finishing internal gears. From this investigation it can be concluded that ANSYS MAXWELL is one of the efficient tools for Magnetic based simulations. The Wind Turbine magnet of 36 mm Outer Diameter magnet is most suitable for viscoelastic magnetic abrasive Finishing than the other commercially available magnet. It can be observed that when the above magnet is used, the distribution of the field is maximum, making it suitable for the MAF process. In terms of magnetic flux density, Steel's workpiece has the edge over the other two materials. Brass has the edge over the other two materials in terms of magnetic field strength and magnetic energy. In terms of current Density, Steel leads the other two materials.

REFERENCES

Wnag, Y., Hu, D.J. 2004. Mechanical Cutting Model of Magnetic Abrasive Particles and Analysis of experimental Results. DOI 10.4028/Scientific.net/KEM.274-276.451.

Rampal, Rohit., Goyal, Tarun. 2017. Comparing the techniques for Magnetic Abrasives by Investigating the Surface Finish IJAER, ISSN 0973-4562 Volume 12, Number 19 pp.8972-8975.

Kang, Paramvir.Singh., Singal, L.C., Gill, Rajwinder. Singh. 2017. Effectiveness of sintered Abrasive in Magnetic Abrasive Finishing. IJAMTE, ISSN NO: 2249-7455.

Muhamad, M.R., Zou, Y., Sugiyama, H. 2016. Investigation of the finishing characteristics in an internal tube finishing process by magnetic abrasive finishing combined with electrolysis, Taylor and Francis, DOI 10.1080/00202967.2016.1162400.

Shrikhande, Nikhil. S., Deshkar, Amey. A., Awasthi, Saurabh. S., Meshram, Akshay V., Dhoble, P. Ashwini., Jumle, Patric. A., Kapgate, Bhubneshwar. M. 2018. Magnetic Field assisted Abrasive Flow machine-based Surface Super Finishing, (IJRASET), 6(4).

Chen, H.L., Li, W.H., Yang, S. Q., Yang, S.C. 2010**.** Research of Magnetic Abrasive Prepared by hot Pressing sintering process, IEEE 978-1-4244-5046-6.

Judal, K.B., (2016) FEM Based Modelling of Cylindrical Magnetic Abrasive Finishing (C-MAF) Process using unbounded Magnetic Abrasives, Asia Pacific Journal, ISBN 978-81-930411-8-5.

Srinivas, K., Murtaza, Q., Aggarwal, A.K. 2019. Effect of Shape of Magnet on the Machining of Workpiece, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, 8(2): 2011 .

Srinivas, K., Murtaza, Q., Aggarwal, A.K. 2019. effect of permanent magnetic pole orientation on-field strength in viscoelastic magnetic abrasive finishing process International Journal of Mechanical and Production Engineering Research and Development. 9, Issue 5, Oct 2019, 43–52.

Srinivas, K., Murtaza, Q., Aggarwal, A.K. 2019. Modelling Of Viscoelastic Fluid In Finishing Processes. International journal of scientific & technology research. 8(9).

Srinivas, K., Murtaza, Q., Aggarwal, A.K. 2019. Variation of flow parameters and Magnetic flux in Viscoelastic Magnetic Abrasive Finishing Process, International Journal of Applied Engineering Research. 14(8): 1940-1946.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion

Modified Current Control for Tracking Global Peak under fast changing Partial Shading Conditions

Raghavendra Rao P, Vignesh Kumar V, *Senior Member, IEEE,* Balasubramanian Venkatesaperumal, *Senior Member, IEEE* and Vanjari Venkata Ramana, *Member, IEEE*

***Abstract*— The power – voltage (*P-V*) characteristics of photovoltaic (PV) systems exhibit multiple power peaks under partially shaded conditions. Several global maximum power point tracking (GMPPT) algorithms in the literature recognize the irradiance change, only after the convergence of operating point to global peak, or use additional hardware to call GMPPT subroutine at definite time intervals to detect any insolation change, and thus track the global peak. However, during fast changing partial shading conditions, these methods are less effective, as they do not detect any irradiance change during the tracking phase of any shading pattern. This paper proposes a novel modified current control approach that uses current as a parameter to detect the insolation change during the tracking phase and track the global peak under fast changing partial shading conditions without any additional hardware. The proposed technique improves the tracking efficiency by as much as 39%, thus proving to be effective under fast-changing partial shading conditions. The superior tracking performance of the proposed algorithm over the existing techniques in terms of its tracking efficiency, dynamic tracking capability, tracking speed, and convergence to the global peak is demonstrated with extensive simulations using MATLAB/Simulink and experimental results.**

***Index Terms*— Current Control, Maximum power point tracking, Photovoltaic systems and partial shading conditions.**

I. INTRODUCTION

The maximum power point tracking (MPPT) process is a major challenge in the operation of photovoltaic (PV) power generation systems. This is attributed to the non-linear nature and dependency of PV system characteristics on environmental conditions [1]. Generally, PV systems consist of PV arrays in which several PV modules are connected in series-parallel to meet the total demand. The MPPT process becomes further complicated, when such PV arrays are subjected to partially shaded conditions due to surrounding buildings, moving clouds, human shade, etc., [2], [3]. Under these situations, the power-voltage (*P-V*) characteristics exhibit multiple power peaks. The peak pertaining to the highest power value is called global maximum power point (GMPP) and all other peaks are local maximum power points (LMPPs) [2].

Traditionally, MPPT is performed using methods such as perturb and observe (P&O) [4], incremental conductance (INC) [5], and hill climbing (HC) [6]. It is worth mentioning that these techniques are well-established, time-tested, and effective under uniform insolation conditions [7]. While in the presence of mismatching scenarios, these approaches by themselves are incapable of tracking GMPP and result in under-utilization of the PV systems [2].

In the past years, several MPPT techniques have been reported in the literature for tracking GMPP under mismatching conditions [8]-[9]. These techniques vary in the number of sensors required, complexity, tracking speed, cost of the system, tracking efficiency, and steady-state oscillations. In general, the methods for mitigating the effect of mismatching conditions in PV systems can be grouped into four broad categories namely, modified GMPPT techniques [10]-[11], array re-configuration methods [12], PV system architectures [13], and deployment of different converter topologies [14]. The work presented in [9] shows that the last three strategies involve complex control, require additional hardware, and have higher switching loss compared to modified GMPPT techniques. Further, modified GMPPT techniques are found to be promising with guaranteed convergence to GMPP and have minimal dependence on the system, thereby becoming the appropriate method for MPPT under mismatching conditions [15]. This has led to the development of many modified MPPT methods [16]-[31]. In [16], the authors have determined the perturbation step size by the Fibonacci sequence to identify GMPP. This method is almost similar to the P&O technique and may sometimes converge to the local maxima. A reduced voltage scanning path GMPPT algorithm is used in [17] to reduce the convergence time. This algorithm seems to be effective in most cases, but it is complex resulting in a significant increase in the computational burden of the overall MPPT system. The artificial neural network (ANN) based GMPPT technique is proposed in [18]. Though ANN is effective in dealing with nonlinear characteristics of PV system, a requirement of large data for training is the major constraint in using this method.

Further, in the literature, many works [19]-[26] have employed metaheuristic optimization techniques for GMPP tracking because of their ability to handle multimodal objective functions. These works introduced modifications in the classical structure of metaheuristic algorithms and shown improved tracking performance with changing shading patterns and varying load conditions. However, these algorithms suffer from drawbacks such as prolonged oscillations in the tracking curve, parameter tuning, comparatively lesser tracking speed, and implementation complexity.

Besides the aforementioned methods, there are other modified approaches proposed for global peak detection in PV systems. In [27], an improved proportional integral (PI) controller that directly identifies the GMPP instead of searching

0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion

each LMPP is proposed. This method resulted in improved tracking speed, but it is not analyzed for changing insolation conditions. In another work reported in [28], a single current sensor-based GMPPT method, that sets the upper and lower limits for duty cycle and judicially skips the portion of duty cycle not pertaining to the global peak is proposed. The GMPPT algorithms in [29]-[31] introduce voltage and current boundaries in the tracking phase to reduce the *P-V* curve scanning window thereby resulting in improved tacking speed.

Usually, most of the modified GMPPT techniques work in two stages. In the first stage, a global search is performed to identify the best operating point and in the next stage, a local search is initiated to converge the operating point to GMPP. Though these modified GMPPT methods reported in the literature have higher tracking speed, their convergence time depend predominantly on the location of the operating point prior to the search and the PV array configuration [30]. Therefore, it is vital to confine the operating point to the possible closest proximity of GMPP in every step as the search progresses. Also, it is important to note that, most of the works [19]-[31] in the literature have discussed the effectiveness of GMPPT algorithms in detecting the insolation change, "after the steady-state is reached". But, the GMPPT algorithms must witness the fast change in insolation levels due to birds flying over the panels, leaves falling on the panels, fast-moving clouds, bird droppings, soiling on panels, etc., even during the tracking period. If they fail to detect the fast-changing insolation under these conditions, they may result in premature convergence or may get trapped in LMPP. This results in unnecessary energy loss and under-utilization of the PV panels. Hence, while the two stages of tracking the GMPP are ongoing, the MPP tracker should also witness the variations in the shading profile during any stage and track the new GMPP corresponding to the new pattern. Such detection of fast-changing insolation conditions in the tracking phase is not addressed in the literature.

In order to alleviate the aforementioned problem, this paper proposes a new modified current control algorithm for MPPT under non-uniform insolation conditions. This approach attempts to introduce a few modifications in the steps of the existing current control technique algorithm [31] to have inherent dynamic tracking capability during the initial stages of the search. Then, P&O is cascaded in the second stage to successfully track the GMPP. In the new method, the operating point is progressed to GMPP by making step changes in the PV current. At every step, the changes in shading pattern is observed directly from the current value without using any additional equations in the algorithm, or sensors in the hardware. This approach incorporates checking for pattern change as an essential component in the tracking phase which leads to guaranteed convergence to GMPP. Further, the modified current control also enhances the tracking speed as it quickly detects the best operating point in the vicinity of GMPP under transient conditions. The rest of the article is organized as follows.

Section II explains the PV system used in this work for testing the MPPT algorithms. Section III describes the development of the proposed algorithm for tracking GMPP under fast-changing partially shaded conditions. Section IV discusses the findings from simulations and experiments that are carried out to authenticate the performance of the proposed technique followed by section V which summarizes the conclusions of the proposed work.

TABLE I
SPECIFICATIONS OF PV MODULE

Parameter	Notation	Value
Open Circuit Voltage	V_{oc}	23.3V
Short Circuit Current	I_{sc}	2.68A
Voltage at MPP	V_{mpp}	17.09V
Current at MPP	I_{mpp}	2.41A
Maximum Power	P_m	41.2W
Number of cells in a module	N_s	36

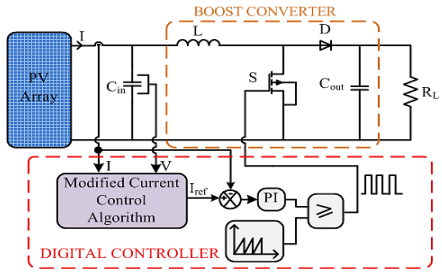


Fig. 1. MPPT employed PV system.

II. SYSTEM DESCRIPTION

The PV system used to verify the proposed MPPT algorithm is shown in Fig. 1. It consists of a boost converter that interfaces the PV array with resistive load. The PV array used in this work consists of different numbers of PV modules connected in series and parallel. The specifications of a single PV module at standard test conditions (STC) are listed in Table I. The boost converter operates at 25 kHz. This closed loop system consists of the MPPT controller that senses the panel voltage (V) and current (I) and determines the reference current (I_{ref}). This generated I_{ref} is further compared with actual current (I), and is fed to the PI control block to generate the gating signal of appropriate duty ratio.

III. DEVELOPMENT OF THE MODIFIED CURRENT CONTROL ALGORITHM FOR MPPT

This section briefs about conventional current control technique followed by, the key parameter for detecting the irradiance change during the tracking phase of proposed technique, the proposed methodology and the failure case of current control technique.

A. *Brief review of current control algorithm*

The current control MPPT algorithm [31] involves two phases, namely backward phase, where the current perturbation is carried out by dividing the PV current by 0.9 [32], and forward phase, where the PV current is multiplied by 0.9 [32]. The initiation of any of these phases is based on (1).

$$V < V_{min}, \quad \text{where} \tag{1}$$

$$V_{min} = \frac{P_{max}}{I_{mpp@STC}} \tag{2}$$

where P_{max} is the maximum power, whose initial value, is the

0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion

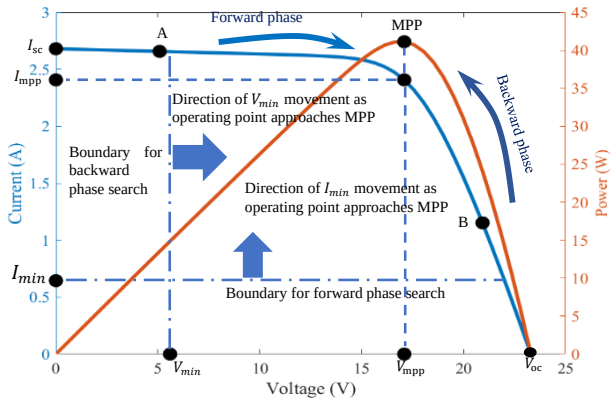


Fig. 2. Illustration of current control algorithm based MPPT.

power calculated at the first sampling instant, and its value gets updated in each perturbation if the present power is greater than the stored value of P_{max}. $I_{mpp@STC}$ is the current at maximum power point (MPP) at STC and V_{min} is the minimum voltage below which there is no chance of occurrence of global peak [31]. The forward phase gets initiated if (1) satisfies, else backward phase gets initiated. A detailed explanation of this is presented in [31]. To illustrate the current control technique, the *P-V* and *I-V* characteristics of a single module shown in Fig. 2, obtained from the MATLAB environment are considered [33]. If the initial operating point on the *I-V* curve is at point A, as indicated in Fig. 2, then the algorithm makes its movement in the forward direction, as (1) is satisfied. Perturbation continues in the forward phase by multiplying the current at point A with 0.9. This process continues till the termination criteria in (3) is satisfied.

$$I < I_{min}, \quad \text{where} \tag{3}$$

$$I_{min} = \frac{P_{max}}{0.9 \times V_{oc_array}} \tag{4}$$

where, I is the current at the present operating point, I_{min} represents the minimum current below which there is no chance of occurrence of global peak [31]. V_{oc_array} is the open circuit voltage of the PV array at STC. It can be observed from Fig. 2 that, as the operating current decreases during the forward phase, I_{min} is moved up along the current axis with the increase of P_{max} thus reducing the scanning window to the right side of GMPP.

Similarly, if the initial operating point is at B, as depicted in Fig. 2, the backward phase gets initiated as (1) is not satisfied. MPP tracking is carried out by dividing the current at point B by 0.9. The value of V_{min} increases along the voltage axis with the increase in the value of P_{max} thus reducing the scanning window to the left side of GMPP. The perturbation continues till the termination criteria (1) is satisfied. With regard to the aforementioned facts, in every iteration of the current control algorithm, the reference current of the MPPT controller is navigated towards MPP location as decided by (5).

$$I_{ref} = \begin{cases} I \times 0.9, & in\ forward\ phase \\ I/_{0.9}, & in\ backward\ phase \end{cases} \tag{5}$$

It can be noted in Fig. 2 that, (2) and (4) act as the limiting conditions thereby restricting the search window in the *I-V* curve of the PV system. Also, (2) and (4) ensure that the GMPP will not exist outside their limits.

B. Criteria for detecting the change in shading pattern

Whenever there is a change in solar insolation level, it is observed that the change in current is considerably large. The main advantage of the current control algorithm is the small perturbation size of the current, determined by (5). Hence, it is inferred that, while GMPP tracking is being carried out, the change in shading pattern can be easily identified directly by comparing the current values. In addition, this can be judicially incorporated to operate simultaneously with the tracking phase so that the flow of the algorithm is not affected. This is the key change that is introduced in the proposed modified current control algorithm to improve the tracking performance. If the difference in present current (I) and the previous sampled current (I_2) is greater than the critical current (I_{crit}) as in (6), it indicates that there is a change in the shading pattern. I_{crit} is set at 0.11 A (current is multiplied or divided by a factor of 0.9) as the step size of the current between any two perturbations in the forward or backward phase cannot exceed this value. If there is a change in insolation according to (6), then the maximum power P_{max} and the initial operating current (I_1), are reinitialized according to the new pattern and the algorithm is proceeded without restarting the search.

$$|I - I_2| > I_{crit} \tag{6}$$

C. Description of the Proposed Algorithm

The tracking of GMPP using the proposed modified current control algorithm is explained by subjecting the PV system shown in Fig. 1 to two different shading conditions. The two *P-V* characteristics generated in MATLAB are named as Pattern 1 and Pattern 2 and are depicted in Fig. 3(a). The Pattern 1 consists of three power peaks with a GMPP of 145.5W whereas Pattern 2 has GMPP at 114 W. The corresponding *I-V* curves of Patterns 1 and 2 are given in Fig. 3(b).

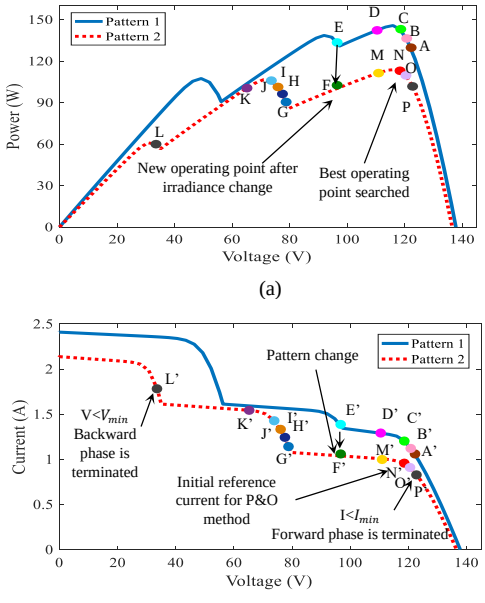


Fig. 3. Operating points during GMPPT in proposed algorithm (a) *P-V* curve (b) *I-V* curve.

0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion

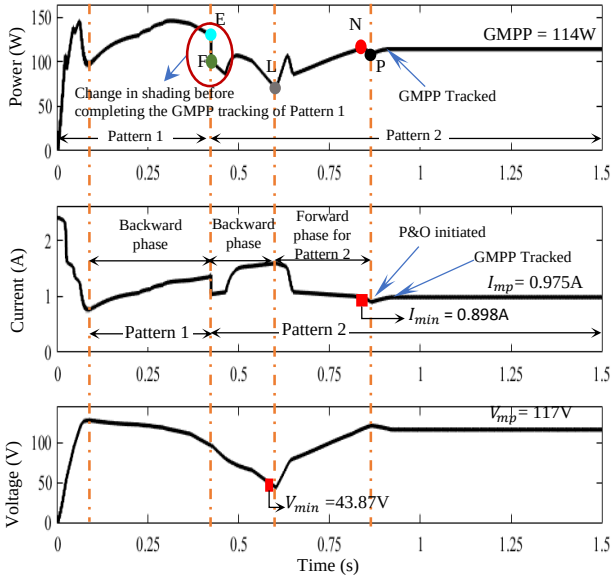


Fig. 4. Illustration of GMPPT process using proposed algorithm with pattern change during tracking phase.

The modified current control algorithm is now employed to track the GMPP with Pattern 1. The initial operating point is denoted as A' in Fig. 3(b) and it corresponds to point A in the *P-V* curve of Fig. 3(a). The *(I, V, P)* corresponding to this point is (1.06 A, 122.16 V, 129.62 W). The computed values of V_{min} and I_{min} using (2) and (4) are 53.78 V and 1.03 A respectively. As the operating voltage is greater than V_{min} (1), GMPP tracking will proceed in the backward direction. The new reference current for the next movement is set according to (5) by the MPPT controller. The backward phase is carried out by increasing the current, and the operating point moves through B', C', D', and E' as shown in Fig. 3(b). At point E', *(I, V, P)* is (1.38 A, 96.60 V, 133.59 W) and V_{min} is updated as 59.30 V.

Now, while the backward tracking phase is progressing to locate GMPP of Pattern 1, the shading profile is changed to Pattern 2. This results in a significant change in the PV current and is observed by (6). At this instant of search, the initial operating current is reinitialized to the value corresponding to point F'. Also, the P_{max} value used for computing V_{min} and I_{min} is reset to zero, which resembles the starting step of the algorithm. This enables the algorithm to act on a new shading profile, without restarting the search process thereby enhancing the tracking speed of the proposed GMPPT. To be precise, the algorithm will continue to proceed in the backward phase as shown in Fig. 3(b) from point F' to L', but with P_{max} and initial current parameters re-initialized according to Pattern 2. At G', the operating current is 1.14 A, and power at point G noted from Fig. 3(a) is 90.61 W which is considered as P_{max}. Since P_{max} value is updated at G' from Pattern 2, V_{min} and I_{min} will also get changed as per Pattern 2 coordinates. From Fig. 3(b), it is identified that at point K' the *(I, V, P)* values are found to be (1.54 A, 65.22 V, 100.58 W). The corresponding (V_{min}, I_{min}) values are computed as (43.87 V, 0.84 A). When the algorithm makes the next step to the point L' which has the reference value of current as 1.67 A, it detects that condition (1) is satisfied and the backward phase is terminated. The algorithm will now, tend to execute the forward phase by jumping to point F' and checking the condition (3). As the value of current at this point (1.0614 A) is greater than I_{min} (0.84 A), the forward phase will be initiated. The search is proceeded to the next operating point M' where *(I, V, P)* is (1 A, 110.96 V, 111.54 W). The computed values of V_{min} and I_{min} using (2) and (4) are 46.28 V and 0.88 A respectively and the search continues till point P'. At point P', the forward phase is terminated as the current at this point is 0.81 A which is less than I_{min}. From Fig. 3(a), it can be seen that the best point in the search is point N which is having maximum power compared to all other points and the current corresponding to N' is set as the initial reference current for the P&O method. The P&O algorithm will then successfully track the GMPP of Pattern 2 that is 114.09 W. The above explained instants can be best viewed with the help of GMPPT tracking curves obtained for the proposed algorithm that is illustrated in Fig. 4.

With shaded modules in the PV system, GMPP can be spotted by employing one complete forward search and backward search. After the completion of this phase, the best operating point will be at the vicinity of GMPP and then the P&O subroutine is called to converge to GMPP. For better understanding, the sequence of events in the proposed technique is provided in the flow chart shown in Fig. 5. The steps involved in the modified current control algorithm for global peak detection are summarized below:

Step 1: Set the initial value of current and operate the boost converter with a duty ratio corresponding to this current.

Step 2: Measure the PV voltage, current, and power at the initial operating point.

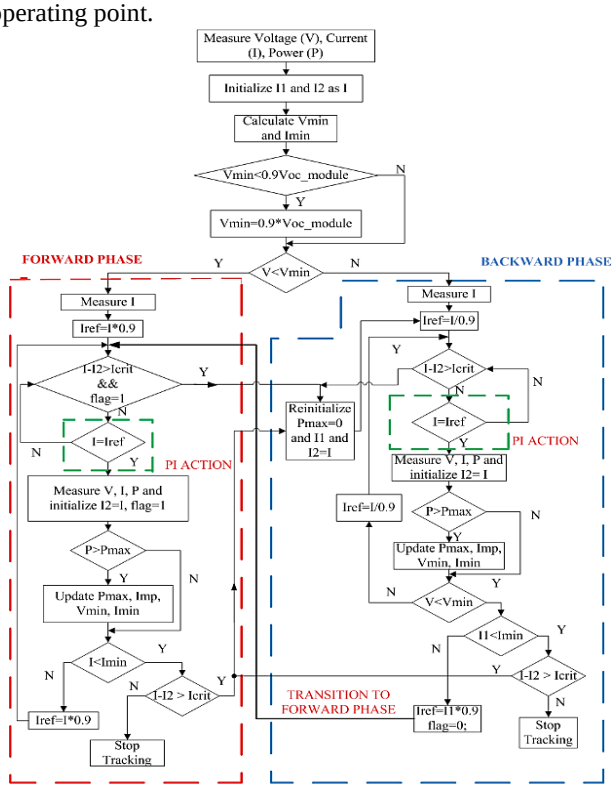


Fig. 5. The flow chart of proposed algorithm for MPPT.

0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion

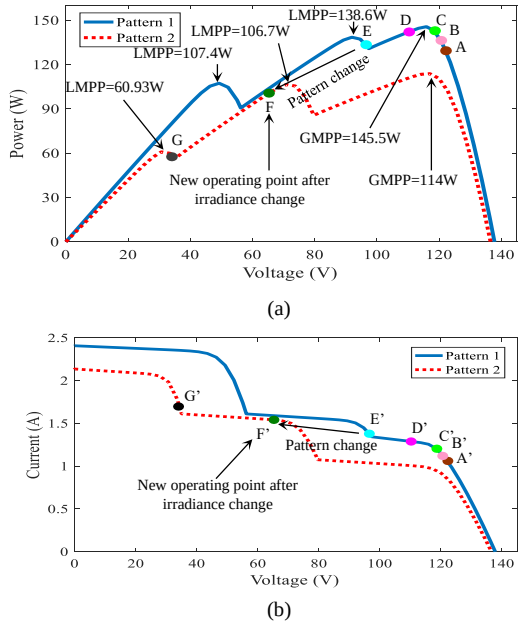


Fig. 6. Operating points during GMPPT tracking in conventional current control method (a) *P-V* curve (b) *I-V* curve.

Step 3: Compute V_{min} and I_{min} and check for condition given in (1) and (3).
Step 4: Initiate forward or backward search according to the decision taken at step 3.
Step 5: Check for shading pattern change using (6) and update the initial conditions and P_{max} value.
Step 6: Call P&O subroutine, if forward and backward search is completed, or else go to step 3.

D. Failure case of GMPP tracking with an existing current control technique.

The inability of the existing current control technique [31] in tracking GMPP for the fast-changing irradiation conditions will now be demonstrated by considering the same Pattern 1 and 2 used in the preceding section. The current control algorithm [31] is now employed to track the GMPP with Pattern 1. The search will start from the backward phase with the initial point as A' and proceed to E' where the shading profile is changed to Pattern 2 as shown in Fig. 6(b). As condition (6) is not used, the existing technique fails to identify the change in insolation level and shifts the operating point to F in Pattern 2. At F', (*I, V, P*) values are obtained from Fig. 6(a) and (b) as (1.54 A, 65.22 V, 100.43 W). As P_{max} attained so far remains as 142.92 W that is corresponding to point D, V_{min} will not be updated, and is equal to 59.3 V. Same is the case for I_{min} which is equal to 1.13 A. This makes the algorithm continue in the backward phase up to point G' in the *I-V* curve. Next, the attempt is made to search in the forward direction but as condition (3) is reached, it is not executed. Now, the algorithm looks for the best operating point based on the power value it has tracked so far, which is found to be point D'. The current control algorithm decides the best operating point by considering Pattern 1 but actually, PV panel is subjected to Pattern 2. The P&O method starts to operate with this sub-optimal operating point on Pattern 2 and finally algorithm will settle down to LMPP of 106.7 W whereas the proposed algorithm tracks the GMPP of 114.09 W. This confirms that the conventional current control algorithm will mislead the P&O method, if there is a change in shading profile during the tracking period of GMPP. Thus, the existing current control method will become handicapped under such situations and will not guarantee convergence to GMPP all the time.

TABLE II
INSOLATION LEVEL FOR 6 X 1 PV ARRAY CONFIGURATION

Patterns	Insolation level in W/m²					
	Module 1	Module 2	Module 3	Module 4	Module 5	Module 6
Pattern 3	1000	1000	1000	1000	1000	1000
Pattern 4	1000	900	900	800	800	500
Pattern 5	900	900	900	500	500	500
Pattern 6	1000	1000	900	900	700	700
Pattern 7	900	800	700	600	400	200

IV. RESULTS AND DISCUSSIONS

This section presents the simulation and experimental results of the proposed GMPPT method. Further, to prove the effectiveness of the proposed technique, a detailed comparison of its performance with the current control [31] and the enhanced adaptive P&O [11] method is given.

A. Simulation results of 6X1 configuration

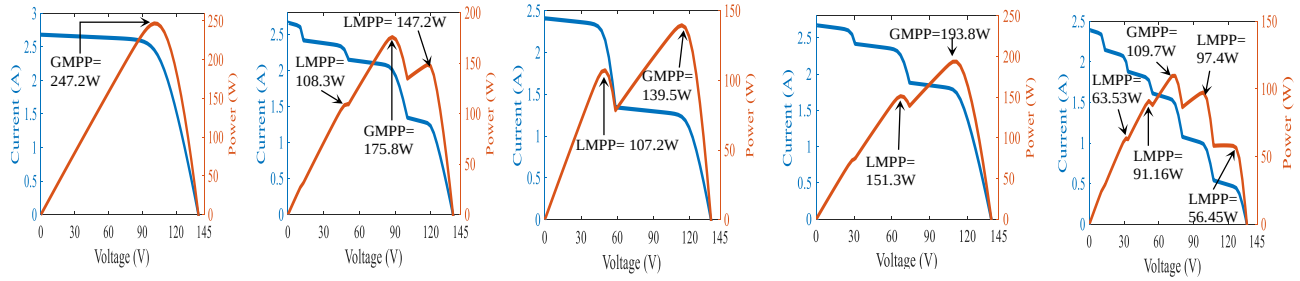
The proposed technique as explained in Section III is now employed to track the GMPP for a 6 x 1 PV configuration. This configuration is subjected to five different shading patterns and the corresponding *I-V* and *P-V* curves are shown in Fig. 7. These curves are named as Pattern 3, Pattern 4, Pattern 5, Pattern 6, and Pattern 7. The solar insolation level of each PV module for generating these patterns is shown in Table II and the GMPP of these patterns is represented in the matrix form as Patterns [3 4 5 6 7] = [247.2W 175.8W 139.5W 193.8W 109.7W]. The PV system shown in Fig. 1 is simulated with the new approach for two different cases and the results obtained are discussed in the following sections.

1) Case 1: Dynamic tracking for the reduced Insolation Level

Consider that the GMPP tracking of Pattern 4 (existing) is ongoing and the irradiance changes to Pattern 5 (new pattern). It can be observed from Fig. 7, that the GMPP value of Pattern 5 is lesser than Pattern 4. Such instances of pattern change are referred to as reduced insolation level in this work. Now, the shading Patterns 3 to 7 are applied sequentially to check for the confirmed convergence of the algorithm to GMPP, and the tracking curves obtained from MATLAB are delineated in Fig. 8(a). It can be seen from Fig. 8(a) that the proposed technique is capable of tracking MPP of 247 W for Pattern 3 in 0.66s. At the time, $t = 1$s, the partial shading condition with Pattern 4 is introduced. As observed from Fig. 8(a), the algorithm detects the change in insolation and re-initiates the search process. Now, to prove the effectiveness of the proposed approach during the tracking phase, Pattern 5 is employed at $t = 1.05$s; even before the algorithm converges to the GMPP of Pattern 4. It is evident from Fig. 8(a) that the modified current control technique has successfully tracked the GMPP of 139.3 W corresponding to Pattern 5 with a tracking time of 0.57s. This rapid tracking capability of the proposed technique is tested again at $t = 2$s, by repeating the same procedure for Pattern 6

0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion



(a) (b) (c) (d) (e)

Fig. 7. *I-V* and *P-V* curves for 6 x 1 configuration (a) Pattern 3 (b) Pattern 4 (c) Pattern 5 (d) Pattern 6 and (e) Pattern 7.

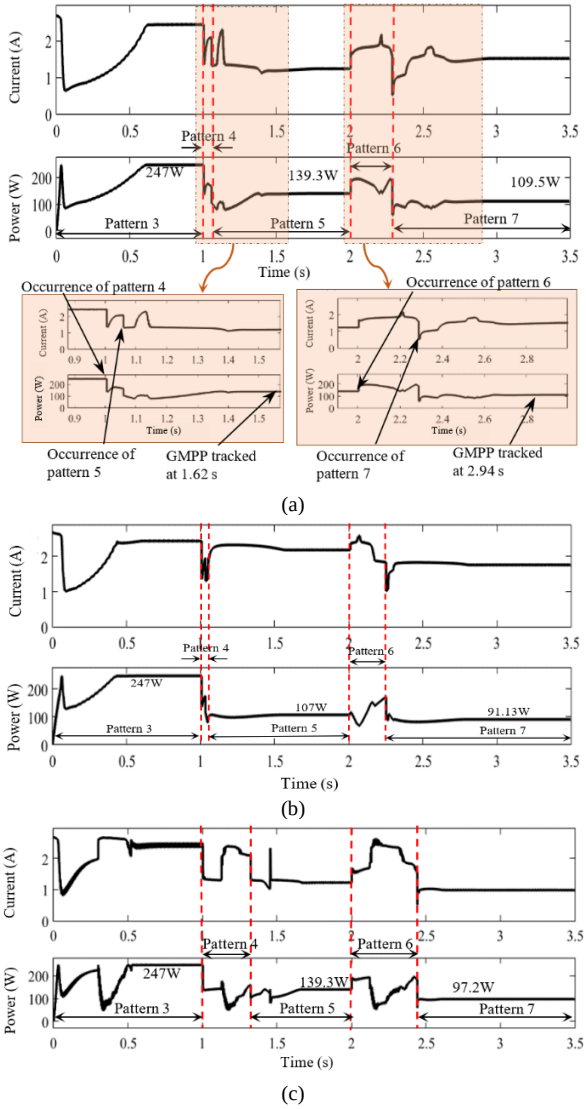


Fig. 8. Computed GMPPT curves for 6 x 1 configuration: Case 1 (a) proposed technique (b) conventional current control method (c) enhanced P&O method.

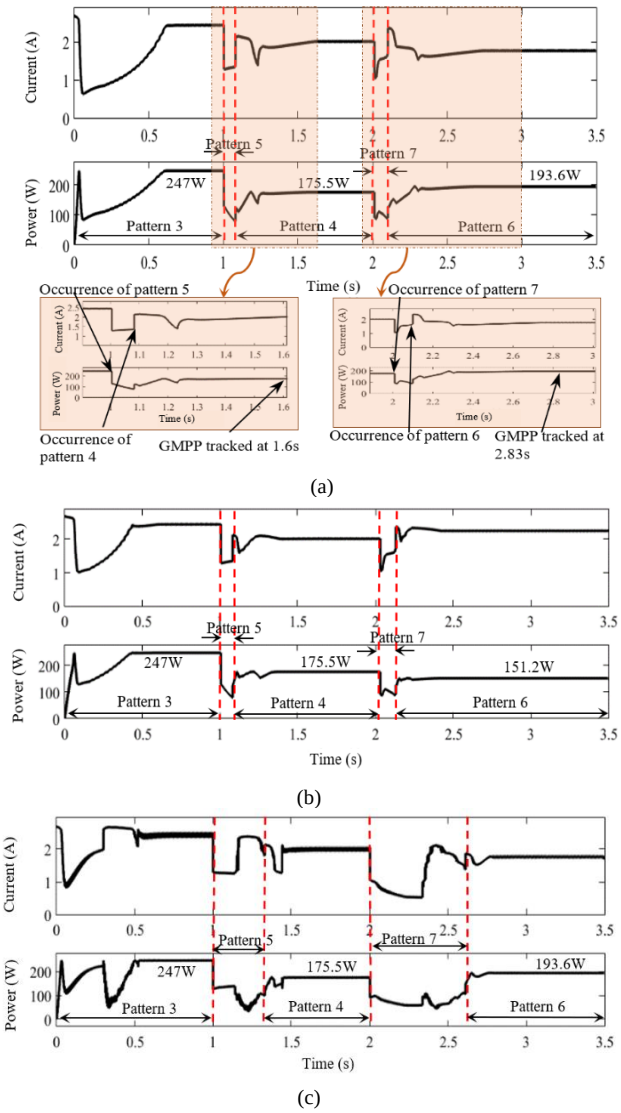


Fig. 9. Computed GMPPT curves for 6 x 1 configuration: Case 2 (a) proposed technique (b) conventional current control method (c) enhanced P&O method.

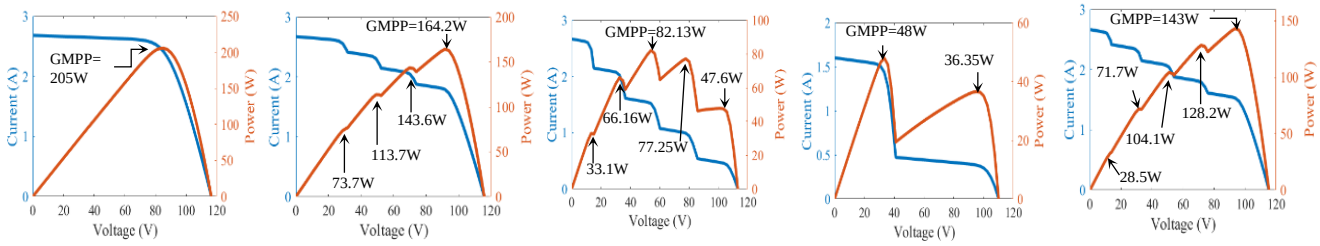
and Pattern 7. It is seen from Fig 8(a) that; proposed algorithm locates the GMPP of 109.5 W in 0.66s for Pattern 7. The significance of the proposed GMPPT technique is further endorsed by comparing its tracking curves with the current control technique [31] and enhanced P&O [11] method, whose tracking curves are shown in Fig. 8(b) and 8(c) respectively. It can be seen from Fig. 8(b) that the existing current control technique settles to a local peak of 107 W for Pattern 5 and 91.13W for Pattern 7. The observation from Fig. 8(c) shows that enhanced adaptive P&O tracks the GMPP for Pattern 3 and Pattern 5 whereas for Pattern 7, it gets trapped in a local peak of 97.2 W.

2) Case 2: Dynamic tracking with Increased Insolation level

This case refers to the situation wherein the GMPP of the upcoming pattern is greater than the existing one. That is while tracking the GMPP of Pattern 5 if the irradiance level changes to Pattern 4, it is referred to as increased insolation level. To

0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion



(a) (b) (c) (d) (e)

Fig. 10. *I-V* and *P-V* curves for 5 x 1 configuration (a) Pattern 8 (b) Pattern 9 (c) Pattern 10 (d) Pattern 11 and (e) Pattern 12.

TABLE III
INSOLATION LEVEL FOR 5 X 1 PV ARRAY CONFIGURATION

Patterns	Insolation level in W/m^2				
	Module 1	Module 2	Module 3	Module 4	Module 5
Pattern 8	1000	1000	1000	1000	1000
Pattern 9	1000	1000	900	800	700
Pattern 10	1000	800	600	400	200
Pattern 11	600	600	175	175	175
Pattern 12	1000	900	800	700	600

confirm the veracity of the proposed algorithm under this kind of situation, the shading patterns are made to occur in the following order viz. Pattern 3, Pattern 5, Pattern 4, Pattern 7, and Pattern 6. The GMPPT curves obtained are shown in Fig. 9(a). As expected, for Pattern 3, the algorithm settles to MPP in 0.66s. Then, Pattern 5 is applied at $t = 1$s, and while tracking its GMPP, the shading condition is changed to Pattern 4. Fig. 9(a) shows that the proposed method successfully tracks the GMPP of Pattern 4 by discarding the Pattern 5. A similar behavior is seen for the proposed technique at $t = 2$s for the sequence of Pattern 7 and Pattern 6. This ensures that (2) and (4) are updated effectively for all the possible partial shading conditions.

The results obtained for the current control technique shown in Fig. 9(b) indicate that for Pattern 4, it is capable of tracking GMPP and however, for Pattern 6 it settled to LMPP. But, the enhanced P&O method as seen from Fig. 9(c), in this case, is capable of tracking the GMPP for Pattern 4 and Pattern 6. This is because the operating point during the change in shading condition falls in the vicinity of GMPP of the existing pattern which may not occur for all the shading profiles. The extensive simulations carried out indicate that, though the existing current control method converges to GMPP under rapidly varying insolation conditions, it fails to track true GMPP. Also, the observations on Fig. 8(c) (case 1) and 9(c) (case 2) reveal that the enhanced P&O is not reliable under the presence of fast-changing partial shading conditions.

B. Simulation results of 5 X 1 configuration

The ability of the proposed technique to track GMPP for different numbers of panel in series is now confirmed with a 5 x 1 configuration. For this simulation, five different shading patterns namely Pattern 8, Pattern 9, Pattern 10, Pattern 11, and Pattern 12 with insolation levels shown in Table III are used. Fig. 10 depicts the *I-V* and *P-V* curves of these patterns. The simulation was carried out in a similar manner explained for the 6 x 1 configuration and the tracking curves obtained for the three GMPPT methods are presented in Fig. 11. The timing instants of a pattern change and the GMPP convergence are clearly marked for all the methods in Fig. 11. It is worth mentioning that, at t = 2.5 s, Pattern 11 with GMPP at extreme

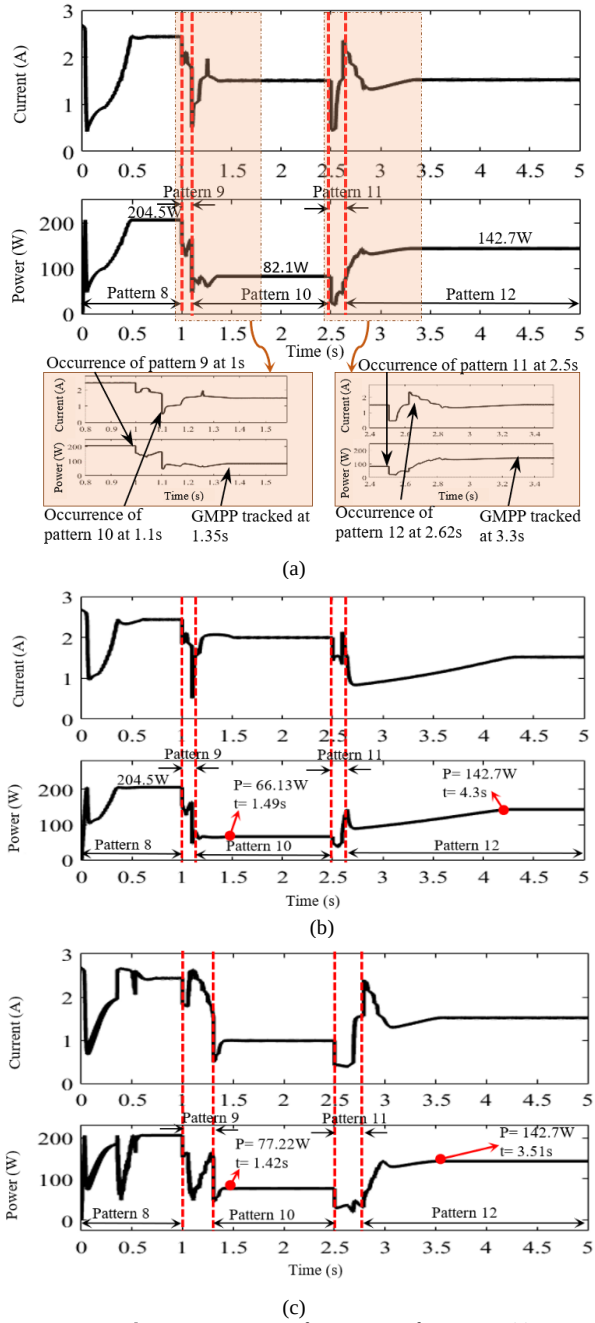
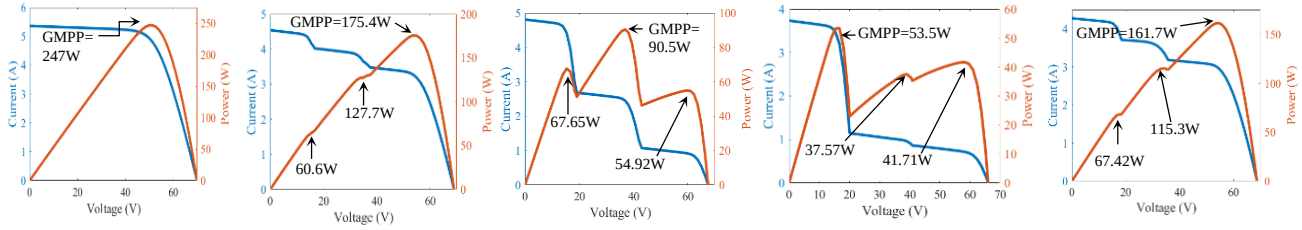


Fig. 11. Computed GMPPT curves for 5x1 configuration (a) proposed technique (b) conventional current control method (c) enhanced adaptive P&O method.

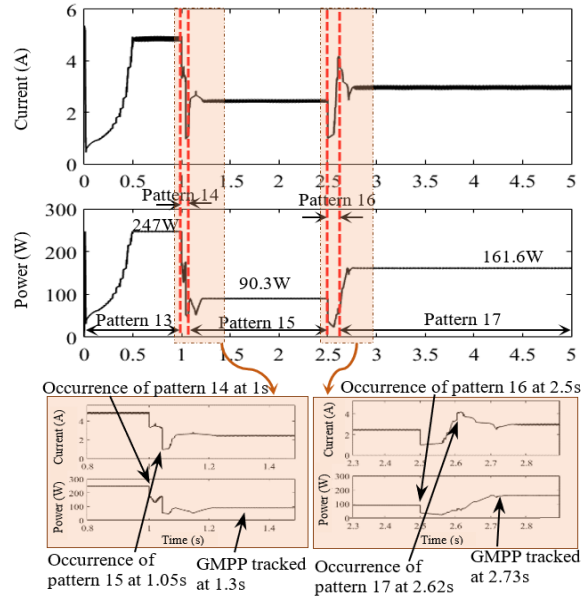
0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion

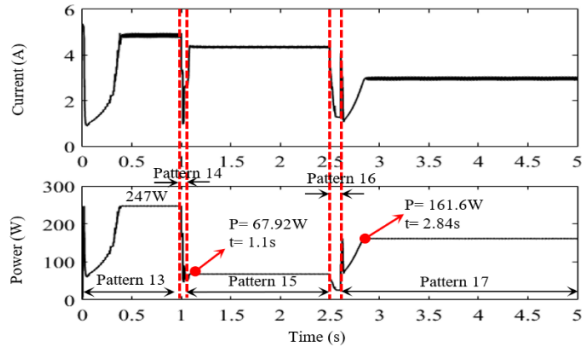


(a) (b) (c) (d) (e)

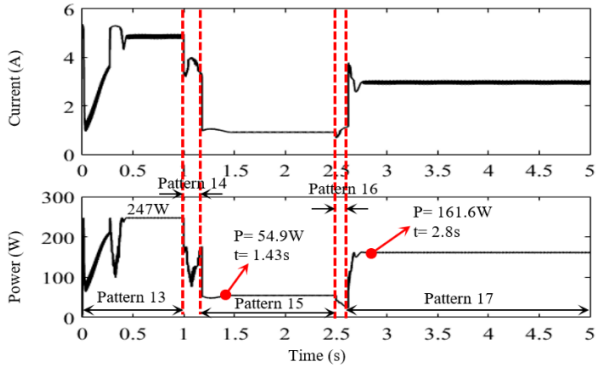
Fig. 12. *I-V* and *P-V* curves for 3 x 2 configuration (a) Pattern 13 (b) Pattern 14 (c) Pattern 15 (d) Pattern 16 and (e) Pattern 17.



(a)



(b)



(c)

Fig. 13. Computed GMPPT curves for 3 x 2 configuration (a) proposed algorithm (b) conventional current control method (c) enhanced adaptive P&O method.

TABLE IV
INSOLATION LEVEL FOR 3 X 2 PV ARRAY CONFIGURATION

Patterns	Insolation Level in W/m^2					
	String 1			String 2		
	Modules					
	1	2	3	4	5	6
Pattern 13	1000	1000	1000	1000	1000	1000
Pattern 14	1000	900	800	700	600	500
Pattern 15	1000	600	200	800	400	200
Pattern 16	600	175	150	800	250	175
Pattern 17	1000	1000	800	600	400	400

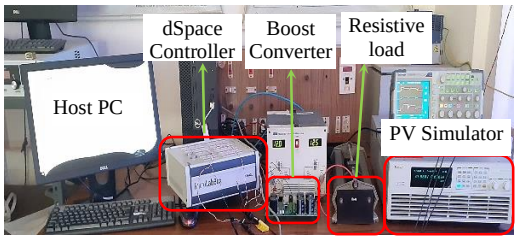


Fig. 14. Experimental setup developed to test GMPPT algorithm.

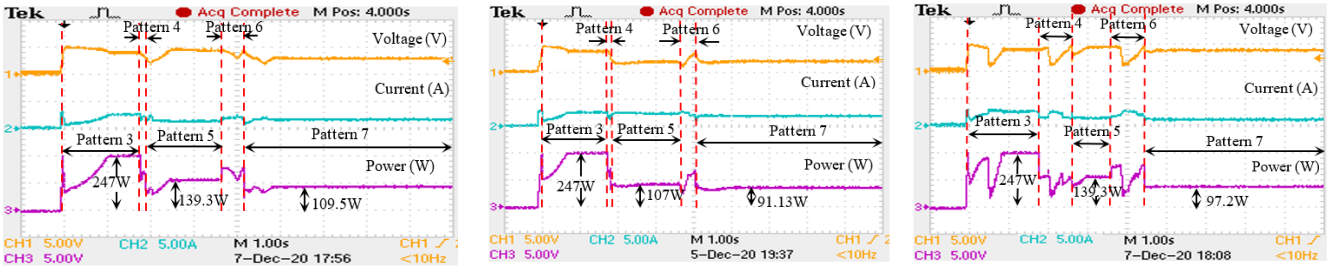
left end is employed. During the GMPP tracking of this pattern, the shading profile is changed to Pattern 12 with GMPP at extreme right end. The proposed technique shifts the operating point from the extreme left end to the right end of the *P-V* curve. Thus, the results provided in Fig. 11(a) further validates that, the proposed algorithm is capable of successfully tracking the GMPP over the complete range of the *P-V* curve. It can be seen in Fig. 11(b) that, for Pattern 10, the current control algorithm converges to a local peak of 66.13W, and for Pattern 12, it tracks the global peak of 142.7W in 1.63s. The enhanced adaptive P&O technique settles to global peak for Pattern 8 and Pattern 12, but it gets trapped at a local peak of 77.24 W for Pattern 10, which is observed from Fig. 11(c).

C. Simulation results of 3 X 2 configuration

The effectiveness of the proposed technique for PV array with parallel strings is now authenticated with 3 x 2 configuration. Like in the previous two configurations, five different shading patterns viz. Pattern 13, Pattern 14, Pattern 15, Pattern 16, and Pattern 17 are considered in the simulation. The insolation levels and PV array characteristics pertaining to these patterns are shown in Table IV and Fig. 12 respectively. The same procedure followed in the previous two configurations is employed in this simulation also and the results obtained for GMPP tracking with the three methods are presented in Fig. 13. It is evidenced from Fig. 13(a) that the proposed algorithm successfully tracks the GMPP for all the patterns. But the existing current control method, as seen from Fig. 13(b) fails to locate GMPP for Pattern 15, and the enhanced P&O method also settles at one of the local peaks of Pattern 15. Thus, the

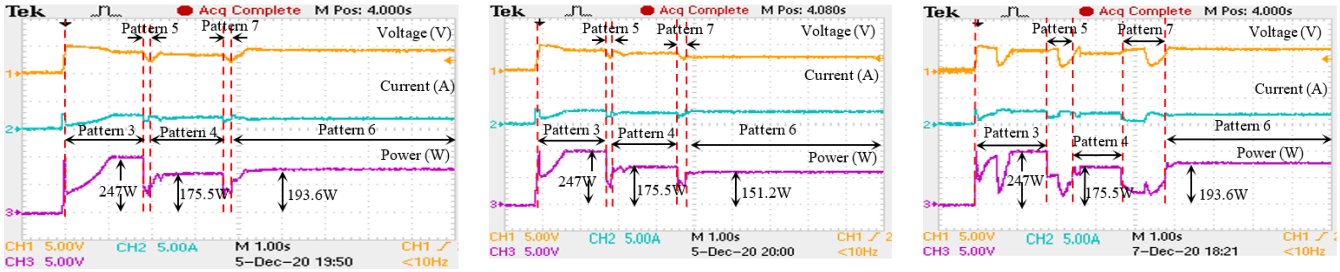
0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion



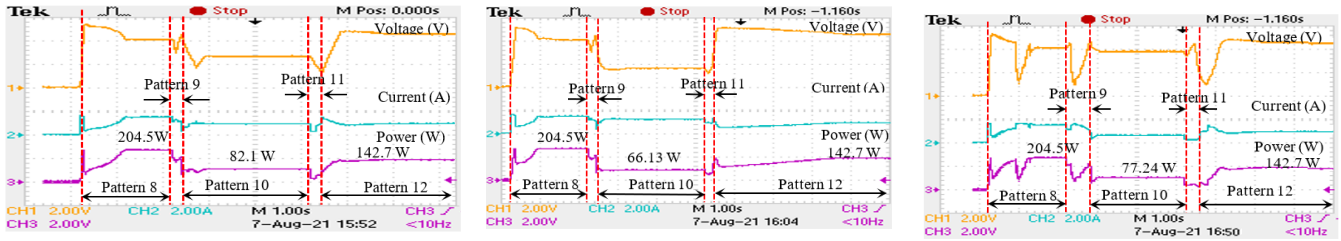
(a) (b) (c)

Fig. 15. Experimental results for 6x1 configuration: Case 1 (a) Proposed Technique (b) Current Control Algorithm (c) Enhanced Adaptive P&O. Scale: Voltage - 140V/div, Current -5A/div and Power-124W/div.



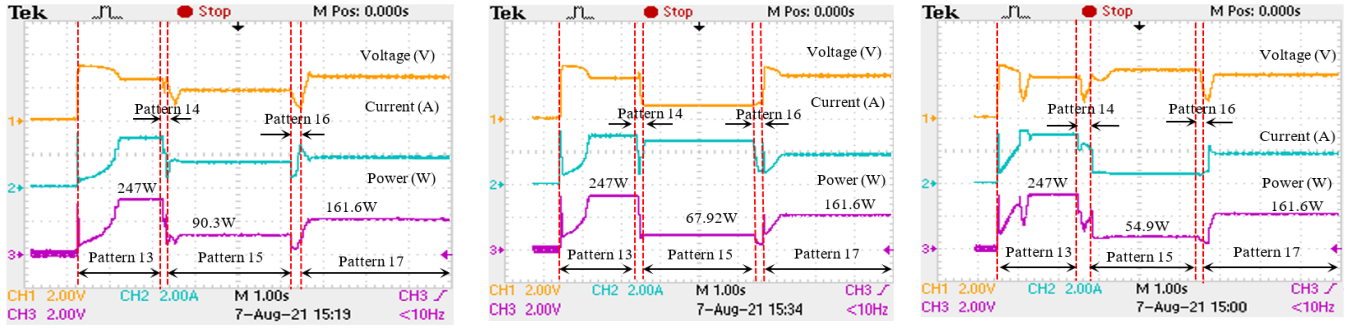
(a) (b) (c)

Fig. 16. Experimental results for 6x1 configuration: Case 2 (a) Proposed Technique (b) Current Control Algorithm (c) Enhanced Adaptive P & O Scale: Voltage - 140V/div Current -5A/div Power -124W/div.



(a) (b) (c)

Fig. 17. Experimental results for 5 x 1 configuration (a) Proposed Technique (b) Current Control Algorithm (c) Enhanced Adaptive P&O. Scale: Voltage 43V/div, Current -3A/div and Power-143W/div.



(a) (b) (c)

Fig. 18. Experimental results for 3 x 2 configuration (a) Proposed Technique (b) Current Control Algorithm (c) Enhanced Adaptive P&O. Scale: Voltage - 43V/div, Current -3A/div and Power-143W/div.

TABLE V
SPECIFICATIONS OF BOOST CONVERTER

Parameter	Value
Input Voltage	200V
Inductor (L)	1.5mH
Input Capacitor (C_{in})	300µF
Output Capacitor (C_{out})	300µF
Frequency of Switching (f)	25KHz

extensive simulations were carried out for different panels connected in series and parallel with an ample number of unique *P-V* curves. The results of these simulation studies shown in Figs. 8(a), 9(a), 11(a), and 13(a) indicate that the proposed method can locate the true GMPP quickly for all possible PV patterns and is also independent of PV array configurations. Hence, it can be concluded that it is preferable to use modified current control technique, as it gives good dynamic tracking capability and guaranteed convergence to GMPP with faster tracking speed.

D. Experimental results

The findings of simulation studies are now authenticated

0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion

TABLE VI
EXPERIMENTAL TRACKING PERFORMANCE OF MPPT METHODS FOR DIFFERENT PV ARRAY CONFIGURATIONS

PV Array Configuration	Patterns	MPP Tracked (W)			Tracking Efficiency (%)			Tracking Speed (s)		
		Proposed	[31]	[11]	Proposed	[31]	[11]	Proposed	[31]	[11]
6 x 1 (Case 1)	Pattern 3	247	247	247	99.81	99.91	99.91	1.12	0.9	1.14
	Pattern 5	139.3	107	139.3	99.85	76.7	99.85	0.64	0.74	0.74
	Pattern 7	109.5	91.13	97.2	99.81	83.07	88.6	0.76	1.04	0.34
6 x 1 (Case 2)	Pattern 3	247	247	247	99.91	99.91	99.91	1.14	0.92	1.1
	Pattern 4	175.5	175.5	175.5	99.82	99.85	99.82	0.58	0.72	0.36
	Pattern 6	193.6	151.2	193.6	99.89	88.6	99.89	0.92	0.52	0.7
5 x 1	Pattern 8	204.5	204.5	204.5	99.75	99.75	99.75	0.98	1.00	1.2
	Pattern 10	82.10	66.13	77.24	99.90	80.50	94.00	0.51	0.76	0.3
	Pattern 12	142.7	142.7	142.7	99.80	99.80	99.8	1.25	3.3	1.33
3 x 2	Pattern 13	247.0	247	247	99.90	99.90	99.9	0.98	0.98	0.92
	Pattern 15	90.30	67.92	54.9	99.77	75.04	60.66	0.40	0.19	0.58
	Pattern 17	161.6	161.9	161.9	99.81	99.81	99.81	0.38	0.66	0.23

with the experiments that were carried out in the laboratory setup shown in Fig. 14. For experimentation, the solar PV array simulator is used as a PV source and the boost converter is employed as an interface between the PV system and load. The specifications of the boost converter are listed in Table V. The MPPT algorithms are implemented in the dSPACE Microlab box controller using the Control Desk software. The PV array voltage and current are sensed using the analog to digital conversion (ADC) channels of the dSPACE controller box. These signals are acquired at the sampling rate of 100 milliseconds and are multiplied internally in the software to get the PV power. The values of PV array voltage, current, and power are given as input to the MPPT algorithm for computing the desired operating current. The PV simulator is configured with shading patterns pertaining to the 6 x1 configuration that is shown in Fig. 7. The experiment was conducted under the case 1 scenario as in simulation for the three MPPT methods and the recorded experimental tracking curves are given in Fig. 15. It can be seen from Fig. 15(a) that the proposed algorithm under case 1 scenario has tracked the GMPP in 1.12 s, 0.64 s, and 0.76 s for Pattern 3, Pattern 5 and Pattern 7 respectively. It is important to mention that, the difference in the tracking time from simulation results for the same shading profile is attributed to the non-idealities of the boost converter and the internal timing delays in sensing and ADC circuits. As in the simulation study, the existing current control technique [31] failed to track GMPP for Pattern 5 and Pattern 7 which is evident from Fig. 15(b). The experiment was repeated for the enhanced P&O [11] method and the results obtained as seen from Fig. 15(c) indicate that the enhanced P&O also fails to locate the GMPP. Now, the experiments are again repeated for all the three MPPT methods with the sequence of pattern used in case 2 of the simulation, and the measured tracking curves are depicted in Fig. 16. The inspection of Fig. 16(a) shows that the modified current control technique captures the GMPP for Pattern 3, Pattern 4 and Pattern 6 with tracking times of 1.14 s, 0.58 s, and 0.92 s respectively.

In a similar manner, the proposed technique is again experimentally tested for 5 x 1 and 3 x 2 configuration, and the results are compared with the existing current control method and the enhanced P&O method. The measured MPPT curves for 5 x 1 and 3 x 2 configurations are given in Fig. 17 and Fig. 18 respectively. The experimental results of the proposed algorithm in comparison with the existing methods are tabulated in Table VI. These results obtained from the detailed experimentation validate the success of the proposed technique in locating the true GMPP under fast-changing irradiance conditions during the tracking phase. But the existing techniques considered in this work for comparison could not track GMPP under fast-changing shading conditions. Thus, the experimental results obtained, convincingly verify the simulation findings.

TABLE VII
COMPARISON OF PROPOSED ALGORITHM BASED ON DIFFERENT PARAMETERS

Qualities	Proposed	[31]	[11]	[30]
Implementation Complexity	Medium	Medium	Medium	Medium
Sensors used	V, I	V, I	V, I	V, I
Tracking Speed	High	High	Medium	Medium
Tracking Accuracy	High	Medium	Medium	Medium
PV Array Dependency	Medium	Medium	High	Medium
GMPPT capability for fast irradiance change	High	Low	Low	Low
Convergence decision after GMPP is found	High	High	Medium	Medium

E. Performance evaluation of the proposed algorithm

The tracking performance of the proposed MPPT technique is evaluated based on three important characteristic features viz. tracking efficiency, tracking speed, and dynamic tracking capability, and is presented in Table VI. From this table, it can be observed that the tracking efficiency of the proposed technique is around 99.8 % for all the cases of simulation and experimentation. This confirms the guaranteed convergence to GMPP by the proposed technique irrespective of the shading profile. The convergence time to reach GMPP determines the quantum of energy extracted from the PV system and is expected to be low. It is apparent from Table VI that the convergence time is significantly low for the proposed method. Further, Figs 8(a), 9(a), 11(a), 13(a), 15(a), 16(a), 17(a), and 18(a) show that the dynamic tracking capability of the proposed technique is independent of the shading profile and the PV array configuration. Thus, the simulation and experimental findings manifest the substantially improved tracking performance. Additionally, the tracking performance of the modified current control technique with the existing methods on various other performance parameters is illustrated in Table VII. It can be seen that the proposed technique provides all the benefits

0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion

resulting in the efficient and economical MPPT controller that suites most of the applications.

V. Conclusion

This paper has introduced a novel GMPPT algorithm for PV systems under rapidly changing partial shading conditions. This method employs a modified current control technique in the formative stages of search and the traditional P&O method in the later stages. The veracity of the proposed method in tracking the global peak under dynamically varying shading conditions is presented for different shading profiles. The ability of the proposed algorithm in detecting the insolation change even during the tracking phase makes it a strong candidate for GMPPT tracking under fast-changing partially shaded conditions. Further, the effectiveness of the proposed algorithm was analyzed based on convergence to GMPP and tracking time. The superiority of the proposed method over the existing alternatives is demonstrated by the computed and measured results. The proposed technique increases the overall tracking efficiency by 39% for the cases presented, thus proving to be an effective technique under any fast-changing partial shading conditions. This manuscript reveals that limitations of the new dispensation in terms of implementation, hardware requirement, and control complexity are significantly low. There is further room for the extension of this work in analyzing the proposed method on the large-scale PV systems and other applications.

Acknowledgment

The authors thank Prof. Ashok K S Bhat, *Fellow, IEEE*, University of Victoria, Canada and Dr. Arjun Mudlapur, Director, APSARAM Technologies, for their valuable suggestions during the preparation of this manuscript.

References

[1] T. Eshram and P. L. Chapman, “Comparison of photovoltaic array maximum power point tracking techniques,” *IEEE Trans. Energy Convers.*, vol. 22, no. 2, pp. 439–450, Jun. 2007.

[2] H. Patel and V. Agarwal, “Maximum power point tracking scheme for PV systems operating under partially shaded conditions,” *IEEE Trans. Ind. Electron.*, vol. 55, no. 4, pp. 1689–1698, Apr. 2008.

[3] K. A. K. Niazi, Y. Yang and D. Sera, "Review of mismatch mitigation techniques for PV modules," in *IET Renewable Power Generation*, vol. 13, no. 12, pp. 2035-2050, Jul. 2019.

[4] M. Killi and S. Samanta, “Modified perturb and observe MPPT algorithm for drift avoidance in photovoltaic systems,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 9, pp. 5549–5559, Aug. 2015.

[5] G.-C. Hsieh, H.-I. Hsieh, C.-Y. Tsai, and C.-H. Wang, “Photovoltaic power-increment-aided incremental-conductance MPPT with two-phased tracking,” *IEEE Trans. Power Electron.*, vol. 28, no. 6, pp. 2895–2911, Jun. 2013.

[6] Koutroulis, Eftichios, Kostas Kalaitzakis and Nicholas C. Voulgaris, “Development of microcontroller based photovoltaic maximum power point tracking control system” *IEEE Trans. Power Electron.*, vol. 16, no. 1, pp. 46-54, Jan. 2001.

[7] M. A. G. de Brito, L. Galotto, L. P. Sampio, G. de Azevedo e Melo, and V. A. Canesin, “Evaluation of the main MPPT techniques for photovoltaic applications,” *IEEE Trans. Ind. Electron.*, vol. 60, no. 3, pp. 1156–1167, Mar. 2013.

[8] J. Ahmed and Z. Salam, "A critical evaluation on maximum power point tracking methods for partial shading in PV systems", *Renewable Sustain. Energy Rev.*, vol. 47, pp. 933-953, 2015.

[9] A. Bidram, A. Davoudi and R. S. Balog, "Control and Circuit Techniques to Mitigate Partial Shading Effects in Photovoltaic Arrays," in *IEEE Journal of Photovoltaics*, vol. 2, no. 4, pp. 532-546, Oct. 2012.

[10] Mahdi Jedari and S Hamid Fathi, “A new approach for photovoltaic arrays modelling and maximum power point estimation in real operating conditions”, *IEEE Trans. Ind. Electron.*, 2017.

[11] Jubaer Ahmed and Zainal Salam, “An enhanced adaptive P&O MPPT for fast and efficient tracking under varying environmental conditions”, *IEEE Trans. Sustainable Energy*, 2017.

[12] G. Velasco-Quesada, F. Guinjoan-Gispert, R. Pique-Lopez, M. Roman-Lumbreras and A. Conesa-Roca, "Electrical PV array reconfiguration strategy for energy extraction improvement in grid-connected PV systems", *IEEE Trans. Ind. Electron*., vol. 56, no. 11, pp. 4319-4331, Nov. 2009.

[13] R. Gules, J. De Pellegrin Pacheco, H. L. Hey and J. Imhoff, "A maximum power point tracking system with parallel connection for PV stand-alone applications", *IEEE Trans. Ind. Electron*., vol. 55, no. 7, pp. 2674-2683, Jul. 2008.

[14] S. V. Dhople, A. Davoudi, A. D. Dominguez Garcia and P. L. Chapman, "A unified approach to reliability assessment of multiphase DC–DC converters in photovoltaic energy conversion systems", *IEEE Trans. Power Electron*., vol. 27, no. 2, pp. 739-751, Feb. 2012.

[15] M. Kermadi and E. M. Berkouk, "Artificial intelligence-based maximum power point tracking controllers for photovoltaic systems: Comparative study", *Renewable Sustain. Energy Rev*., vol. 69, pp. 369-386, 2017.

[16] S. Malathy and R. Ramaprabha, "Maximum power point tracking algorithm of SPVA under inhomogeneous irradiation conditions: A modified Fibonacci search based approach", in *IEEE 12th International Conference on Power Electronics and Drive Systems* (PEDS), Honolulu, HI, USA, 2017, pp. 487-492, 2017.

[17] M. Boztepe, F. Guinjoan, G. Velasco-Quesada, S. Silvestre, A. Chouder and E. Karatepe, "Global MPPT Scheme for Photovoltaic String Inverters Based on Restricted Voltage Window Search Algorithm", *IEEE Trans. Ind. Electron.*, vol. 61, no. 7, pp. 3302-3312, July 2014.

[18] Subiyanto S, Azah Mohamed and M A Hannan, “Intelligent maximum power point tracking for PV system using Hopfield neural network optimized fuzzy logic controller”, *Energy and Buildings*, vol. 51, pp. 29-38, 2012.

[19] M. Seyedmahmoudian, T. K. Soon, B. Horan, A. Ghandhari, S. Mekhilef and A. Stojcevski, "New ARMO-based MPPT Technique to Minimize Tracking Time and Fluctuation at Output of PV Systems under Rapidly Changing Shading Conditions," in *IEEE Tran. Ind. Informatics*, Jan. 2019. doi: 10.1109/TII.2019.2895066.

[20] K. Sundareswaran, V. Vigneshkumar, P. Sankar, S. P. Simon, P. Srinivasa Rao Nayak and S. Palani, "Development of an Improved P&O Algorithm Assisted Through a Colony of Foraging Ants for MPPT in PV System", *IEEE Tran. Ind. Informatics*, vol. 12, no. 1, pp. 187-200, Feb. 2016.

[21] R. B. A. Koad, A. F. Zobaa, and A. El-Shahat, “A novel MPPT algorithm based on particle swarm optimization for photovoltaic systems,” *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 468–476, Apr. 2017.

[22] Imran Parvez, Immad Shams, Saad Mekhilef, Adil Sarwar, Mohd Tariq and Based Alamri, “Most Valuable Player Algorithm based Maximum Power Point Tracking for a Partially Shaded PV Generation System”, *IEEE Trans. Sustainable Energy*, 2021.

[23] Houssam Deboucha, Saad Mekhilef, Sofia Belaid and Amar Guichi, “Modified deterministic Jaya(DM-Jaya)-based MPPT algorithm under partially shaded conditions for PV system”, *IET Power Electronics*, vol. 13, no. 19, pp. 4625-4632, Feb. 2021.

[24] Dalila Fares, Mohamed Fathi, Immad Shams and Saad Mekhilef, “A novel global MPPT technique based on squirrel search algorithm for PV module under partial shading conditions”, *Energy Conversion and Mgt., Elsevier*, vol. 230, Jan. 2021.

[25] Immad Shams, Saad Mekhilef and Kok Soon Tey, “Maximum Power Point Tracking using Modified Butterfly Optimization Algorithm for Partial Shading, Uniform Shading and Fast Varying Load Conditions”, *IEEE Trans. Power Electron.*, vol. 36, no. 5, pp. 5569-5581, 2020.

[26] Saad Motahhir, Aissa Chouder, Aboubakr El Hammoumi, Abou Soufiane Benyoucef, et.al., “Optimal Energy Harvesting From a Multistrings PV Generator based Artificial Bee Colony Algorithm”, *IEEE Systems Journal*, 2020.

[27] Haider Islam, Saad Mekhilef et.al., “Improved Proportional-Integral Coordinated MPPT Controller, with Fast Tracking Speed for Grid-Tied PV Systems under Partially shaded conditions”, *Sustainability, MDPI*, vol. 13, no. 2, 2021.

0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2021.3122550, IEEE Transactions on Energy Conversion

[28] Guohua Zhou, Qiang Bi, Qingxin Tian, Minrui Leng and Guodong Xu, "Single Sensor Based Global Maximum Power Point Tracking Algorithm for PV System with Partial Shading Condition", *IEEE Trans. Ind. Electron.*, 2021.
[29] A. M. S. Furtado, F. Bradaschia, M. C. Cavalcanti, and L. R. Limongi, "A reduced voltage range global maximum power point tracking algorithm for photovoltaic systems under partial shading conditions," *IEEE Trans. Ind. Elect*., vol. 65, no. 4, pp. 3252-3262, Apr. 2018.
[30] Shungang Xu, Yuan Gao, Guohua Zhou and Guihua Mao, "A Global Maximum Power Point Tracking Algorithm for Photovoltaic Systems Under Partially Shaded Conditions Using Modified Power Trapezium Method," *IEEE Trans. Ind. Electron.*, vol. 68, no. 1, pp. 370–380, 2020.
[31] Vanjari Venkata Ramana, Arjun M, Roopa V, B Venkatesaperumal and Sukumar Mishra, "Global peak tracking of photovoltaic array under mismatching conditions using current control", *IEEE Trans. Energy Conv*. vol. 34, no.1, March 2019.
[32] M.A. Masoum, H. Dehbonei, and E. F. Fuchs, "Theoretical and experimental analysis of photovoltaic systems with voltage and current based maximum power-point tracking", *IEEE Trans. Energy Conv*. vol. 17, no. 4, pp. 514-522, Dec. 2002.
[33] M. G. Villalva, J. R. Gazoli, and E. F. Ruppert, "Comprehensive approach to modeling and simulation of photovoltaic arrays," *IEEE Trans. Power Electron*., vol. 25, no. 5, pp. 1198–1208, May 2009.



Raghavendra Rao P received the B.E. degree in Electrical and Electronics engineering and M.Tech in Microelectronics and Control Systems from NMAM Institute of Technology, Nitte, Karnataka, India in 2012 and 2015 respectively. He is currently working towards the Ph.D degree with the Department of Electrical and Electronics engineering, National Institute of Technology, Surathkal, India. His research interests include application of power electronics in renewable energy systems and power electronic converters control.



Vignesh Kumar Vethanayagam (S'16, M'18, SM'21) received the B.E. degree in electrical and electronics engineering, and the M.E. degree in power systems from Thiagarajar College of Engineering, Anna University, Tamil Nadu, India, in 2009 and 2011, respectively and the Ph.D. degree in electrical and electronics engineering from the National Institute of Technology, Tiruchinapalli, Tamil Nadu, India in 2018. He is currently an Assistant Professor in the Department of Electrical and Electronics Engineering, National Institute of Technology, Surathkal, India. His research interests include optimization techniques, fuzzy logic systems, power electronics application to renewable energy systems and electric vehicles.



Balasubramanian Venkatesaperumal (M'15, SM'18) received the B.E. degree in electrical engineering from Madras University, Chennai, India, in 1999, the M.E. degree in power electronics and drives from Bharathidasan University, Tiruchirappalli, India, in 2001, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology Delhi, New Delhi, India, in 2007. He has worked as a Technical Leader with GE, Schneider electric, Power Electronic R&D Division, SunEdison, Bangalore, India. He is currently serving as Professor in the Department of Electrical and Electronics Engineering, National Institute of Technology, Surathkal, India. His research interests include applications of power electronics in health care, solar generation, and brushless generation.



Vanjari Venkata Ramana (M'21) received the M.Tech (Research) and Ph.D. degrees from National Institute of Technology Karnataka, Surathkal, India in 2014 and 2019 respectively. He is currently serving as an Assistant Professor in the Department of Electrical Engineering, Delhi Technological University, Delhi. His research interests include Switch Mode Power Converters, Control of Power Electronic Converters, Photovoltaic Power Conversion, Fast Chargers and Battery Management Systems.

0885-8969 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:59:05 UTC from IEEE Xplore. Restrictions apply.

Multi-Band Polarization Insensitive Ultra-Thin THz Metamaterial Absorber for Imaging and EMI Shielding Applications

Gaurav Saxena[1], Yogita Khanna[2*], Y K Awasthi[3], Priyanka Jain[4]

[1]Electronics & Communication Engineering, Galgotia College of Engineering and Technology, Greater Noida, India
[2]Electronics & Communication Engineering, Manav Rachna University, Faridabad HR, India
[3]Electronics & Communication Engineering, Manav Rachna International Institute of Research and Studies, Faridabad HR, India
[4]Department of Electronics & Communication Engineering, Delhi Technological University, Delhi, India

Corresponding author: Yogita Khanna (e-mail: yogita@mru.edu.in)

ABSTRACT In this article, a multi-band polarization-insensitive ultrathin (0.036ʎ) single layer metamaterial absorber is designed for THz imaging and EMI shielding. A unique oval-shaped structure is compact with three circular ring-shaped resonators is proposed with a unit cell dimension of 36×36×19.6μm^3. The absorbance of the proposed multiband MMA is above 90% for three bands i.e. 98.57%, 90% and 99.85% at 5.58, 7.98-8.84, 11.45THz frequency respectively. Return loss is nearly the same for the changing incident and polarization angle. Therefore, this metamaterial absorber with a wide range of polarization insensitivity is found and it is also suitable for various applications such as quantum RADAR Imaging, energy harvesting, and optoelectronic devices.

INDEX TERMS MMA (Metamaterial Absorber), Triple-band, Polarization-Insensitive, EMC (Electromagnetic Compatibility), Terahertz (THz), EMI (Electromagnetic Interference)

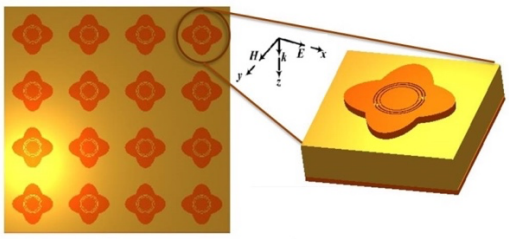
I. INTRODUCTION

In recent progresses, metamaterials have entered in various practical applications such as sensing, cloaking, stealth or absorber, and communication [1-5]. Among these applications absorbers have been found feasible and applicable in a microwave, terahertz as well as the infrared band for applications like terahertz and infrared imaging, perfect absorbers, wave manipulation, energy harvesting, chemical detection and, radioactive cooling. Moreover, it is also used for single-band, dual-band, multiband and wideband applications. Metamaterials in terahertz range are popular due to bandwidth scalability, high security, and strong penetration [6-7, 17, 18]. Metamaterial absorbers are generally arranged as a three-layered structure comprising of a ground plane as a bottom layer, a dielectric as middle, and a resonant periodic structure as a top layer. In comparison with like Salisbury and Jaumann metamaterial absorbers are an alternative that are used in fifth-generation fighter plane stealth to improve the isolation of MIMO antenna, imaging, sensing, Radar cross-section and electromagnetic interference reduction [8-9]. Moreover, multi-band absorbers that are polarization as well as incident wave insensitive have been designed and validated with a thickness lower than the conventional absorbers having absorptivity above 90% for S, and X-band applications [10-11].

In this paper, an ultra-thin microwave absorber is designed that works at three frequency bands i.e. band-1(5.42-5.74 THz), band-2 (7.98-8.84THz), and band-3(11.36-11.54THz) with high absorptivity in X-band. The absorption for the three intended bands is above 90% for TE & TM polarization. The three-layered structure is simple to design and fabricate using commercially available copper-clad laminate. Section-II shows the design of the proposed MMA structure. Section-III depicts the simulated results of the proposed MMA in terms of return loss, absorbance, reflectivity, current density, RCS, and electric field distribution. Conclusion of the proposed designing work is summarized in section-IV.

II. DESIGN OF MMA STRUCTURE

The proposed metamaterial absorber comprises two oval geometries organized as a flower-shaped pattern having three ring-shaped slots cut at the center and a metallic ground plane separated by a dielectric layer, as shown in Figure 1. The MMA is designed on an FR-4 substrate with 19.6μm height (h) and copper plated (σ = 4.2×10^{7}S/m) of 3μm thickness (t). FR4 is commercially available and an alternative to the expensive dielectrics with satisfactory response for the THz range of interest [37-38]. The ground made up of a copper conductor acts as conducting mirror so that there is no transmission of the incident wave and thus converting the incident wave into heat that minimizes the reflected wave which produces high absorbance. The radius and width of all the rings are optimized and the geometrical details of the structure are mentioned in Table-1 and shown in Figure 1.



(a) (b)

FIGURE 1 Triple band meta-material absorber structure (a) Front view (b) Enlarged view

TABLE-I Dimensions of Proposed MMA shown in Figure 2

R1 (μm)	R2 (μm)	R3 (μm)	R4 (μm)
8.5	10	11.3	26
S (μm)	**h (μm)**	**t (μm)**	
36	19.6	3	

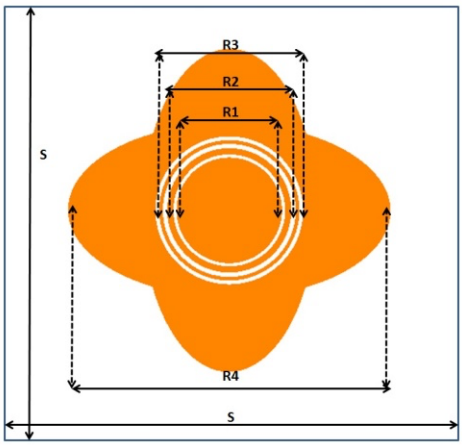


FIGURE 2 Dimensions of a triple-band absorber

Absorption is measured by reflection coefficient (S_{11}) and transmission coefficient (S_{21}) using the simple relation given in (1) [21]:

$$A(\omega) = 1 - |S_{11}|^2 - |S_{21}|^2 \quad (1)$$

The ground plane is a metallic sheet therefore no transmission takes place, so the transmission coefficient (S_{21}) becomes zero. So, (1) is reduced to (2) [21]:

$$A(\omega) = 1 - |S_{11}|^2 \quad (2)$$

$$|S_{11}|^2 = |S_{11co}|^2 - |S_{11cross}|^2 - |S_{21co}|^2 - |S_{21cross}|^2 \quad (3)$$

Equation (4) comes from (2) and (3) [21],

$$A(\omega) = 1 - |S_{11co}|^2 - |S_{11cross}|^2 - |S_{21co}|^2 - |S_{21cross}|^2 \quad (4)$$

Where, S_{11co} and $S_{11cross}$ are the reflection coefficient of co-polarized and cross-polarized EM waves as shown in Figure 5(c).

The S-parameters are calculated using the simulations done on CST-Microwave Studio for a frequency sweep from 5THz to 12THz. The absorber impedance should be comparable to the free space impedance to achieve perfect absorption. The normalized impedance can be calculated using (5):

$$\mathbf{Z} = \sqrt{\frac{(1+S_{11})^2 - S_{21}^2}{(1-S_{11})^2 - S_{21}^2}} = \sqrt{\frac{(1+S_{11})^2}{(1-S_{11})^2}} \quad \mathbf{as\ S_{21} = 0} \quad (5)$$

Equivalent Circuit Analysis-Transmission line theory is used to explain electromagnetic wave absorbers such as Salisbury, Jaumann, and Dallenbach absorbers, etc. This theory is applicable to THz metamaterial absorbers because of scalability. Metamaterials using transmission line theory can be modeled as metasurface impedance connected parallel to the dielectric impedance. Z_{in} & Z_m are the input impedance and impedance of the top layer of the designed metasurface absorber comprising series R, L & C components. Also, X_L & X_C are inductive reactance and capacitive reactance of the equivalent circuit respectively. Z_t is the dielectric impedance of the designed absorber by (6) & (7) [21].

$$\mathbf{Z_m = R + j(X_L - X_c) = R + j\left(\omega L - \frac{1}{2\pi fC}\right)} \quad (6)$$

$$\mathbf{Z_t = j\frac{Z_0}{\sqrt{\varepsilon_r}} \tan\left(\frac{2\pi f\sqrt{\varepsilon_r}h}{c}\right)} \quad (7)$$

The total impedance of the metasurface absorber Z_{in} and reflection coefficient is given as:

$$\mathbf{\frac{1}{Z_{in}} = \frac{1}{Z_m} + \frac{1}{Z_t}} \quad (8)$$

$$\mathbf{\Gamma = \frac{Z_{in} - Z_0}{Z_{in} + Z_0}} \quad (9)$$

Equations (10) and (11) indicate the constraints upon relating s (6)-(9) in agreement to match with the free space impedance, i.e. $Z_{in} = Z_0$.

$$\frac{1}{Z_0} = \frac{R}{R^2 + \left(2\pi fL - {}^1/_{2\pi fC}\right)^2} \quad (10)$$

$$\frac{\sqrt{\varepsilon_r}}{Z_0} \cot\left(\frac{2\pi f\sqrt{\varepsilon_r}d}{c}\right) = \frac{\left(2\pi fL - {}^1/_{2\pi fC}\right)}{R^2 + \left(2\pi fL - {}^1/_{2\pi fC}\right)^2} \quad (11)$$

Put the value of $\frac{1}{Z_0}$ and from (10) in (11) and change "cot" form to "tan" form gives (12) that can be further solved by (13):

$$\frac{R\sqrt{\varepsilon_r}}{R^2 + \left(2\pi fL - {}^1/_{2\pi fC}\right)^2} cot\left(\frac{2\pi f\sqrt{\varepsilon_r}d}{c}\right) = \frac{\left(2\pi fL - {}^1/_{2\pi fC}\right)}{R^2 + \left(2\pi fL - {}^1/_{2\pi fC}\right)^2} \quad (12)$$

$$\frac{R\sqrt{\varepsilon_r}}{tan\left(\frac{2\pi f\sqrt{\varepsilon_r}d}{c}\right)} = \left(2\pi fL - {}^1/_{2\pi fC}\right) \quad (13)$$

Putting value of $\left(2\pi fL - {}^1/_{2\pi fC}\right)$ from (11) to (13) gives (14):

$$\frac{1}{Z_0 tan\left(\frac{2\pi f\sqrt{\varepsilon_r}d}{c}\right)} = \frac{\frac{1}{tan\left(\frac{2\pi f\sqrt{\varepsilon_r}d}{c}\right)}}{R\left(1 + \frac{\varepsilon_r}{tan^2\left(\frac{2\pi f\sqrt{\varepsilon_r}d}{c}\right)}\right)} \quad (14)$$

The value of R, L, and C can be estimated from (15) and (16) as:

$$\mathbf{R = \frac{Z_0 \tan^2\left(\frac{2\pi f\sqrt{\varepsilon_r}d}{c}\right)}{\varepsilon_r + \tan^2\left(\frac{2\pi f\sqrt{\varepsilon_r}d}{c}\right)}} \quad (15)$$

$$\mathbf{\frac{\sqrt{\varepsilon_r}Z_0 \tan\left(\frac{2\pi f\sqrt{\varepsilon_r}d}{c}\right)}{\varepsilon_r + \tan^2\left(\frac{2\pi f\sqrt{\varepsilon_r}d}{c}\right)} = \left(2\pi fL - {}^1/_{2\pi fC}\right)} \quad (16)$$

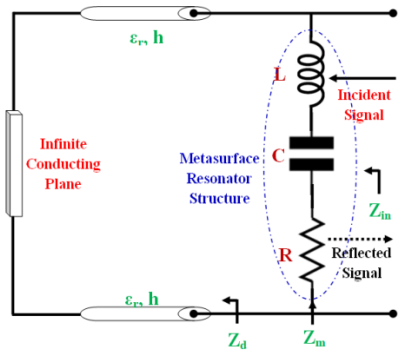


FIGURE 3 Shows the equivalent impedance circuit model of the absorber circuit in terms of series Resistance (R), Inductance (L) & Capacitance (C)

For maximum absorption condition, the impedance of the absorber should be equal to the characteristic impedance of free space as per maximum power transfer theorem i.e. $Z_{in} = Z_0 = 377\Omega$ in [5-9].

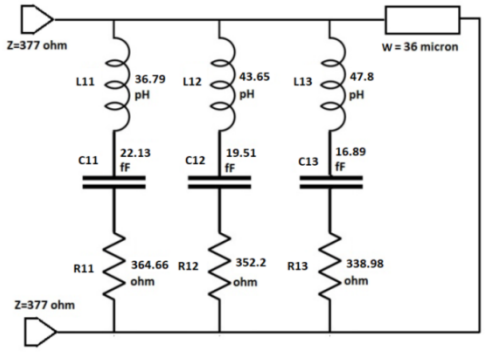
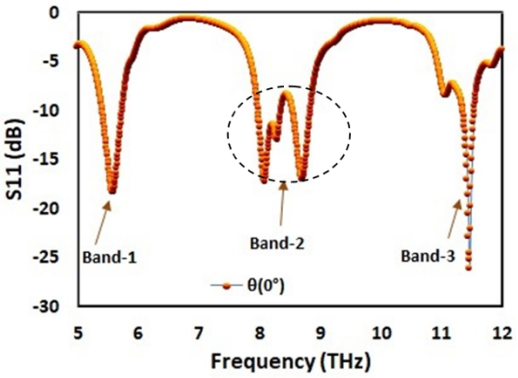


FIGURE 4 Equivalent circuit model of proposed metasurface absorber

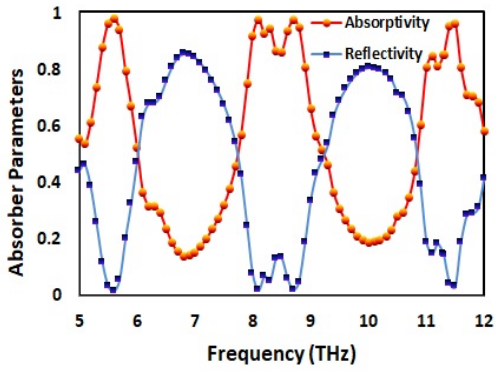
Figure4 shows the equivalent circuit diagram of proposed MMA with three distinct resonance frequencies 5.58, 7.98-8.84, 11.45THz having absorbance level more than 90%. For 5.58, 7.98-8.84, 11.45THz resonance frequency with components value are R_{11} = 364.66Ω, R_{12} = 352.2Ω, R_{13} = 338.98Ω, L_{11} = 36.79pH, L_{12} = 43.65pH, L_{13} = 47.8pH, C_{11} = 22.13fF, C_{12} = 19.51fF, C_{13} = 16.89fF respectively.

III. RESULTS & DISCUSSION

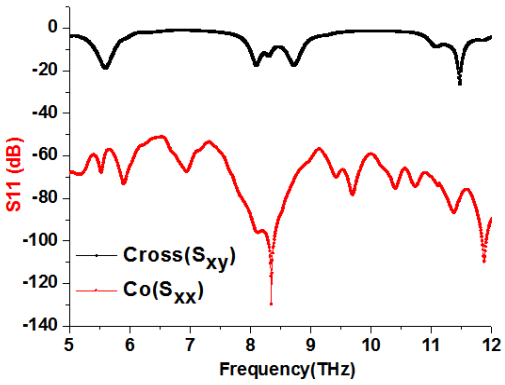
The proposed triple-band MMA is simulated with unit cell boundary condition on the x-y axis and incident wave on the z-axis to generate a return loss curve as shown in Figure 5 (a). Absorptivity at the intended band 5.58THz, 8.71THz, and 11.46THz is above 90% and for band-1(5.42-5.74 THz), band-2 (7.98-8.84THz), and band-3(11.36-11.54THz) having above 90% as shown in Figure 5 (b). The absorption of the proposed MMA is calculated by (2) [9]. The absorber structure is horizontally and vertically polarized waves to obtain co and cross-reflection coefficients by assigning the Floquet ports on the top of the unit cell as indicated in Fig 5(c). Figure 5(d) shows that the phase change due to resonance, indicating the phase variation of ±180° at 5.58THz, 8.71THz, and 11.46THz. Fig 5(e) indicates the resonance due to the individual ring in the absorber structure at 7.54THz, 9.41 THz and 10.418 THz respectively. However, the resonating frequencies of the final structure are difference due to the inter-capacitance between the rings. Fig 5(f) indicates the circuit simulation in comparison to the CST simulation. It shows three resonance frequencies with a small shift.



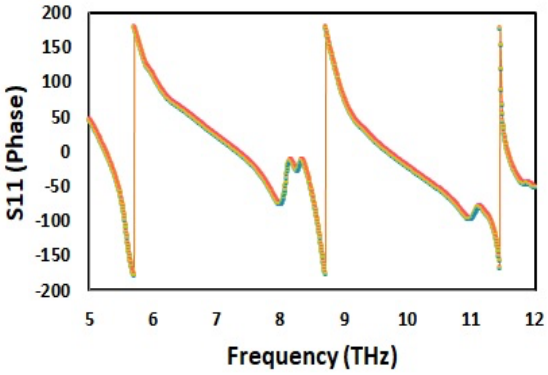
(a)



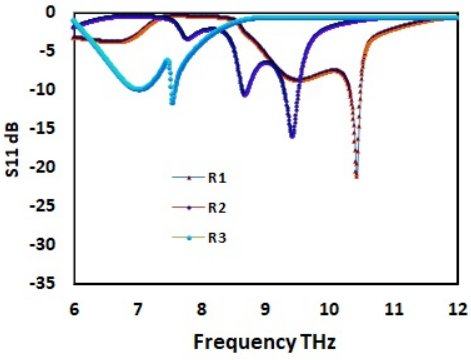
(b)



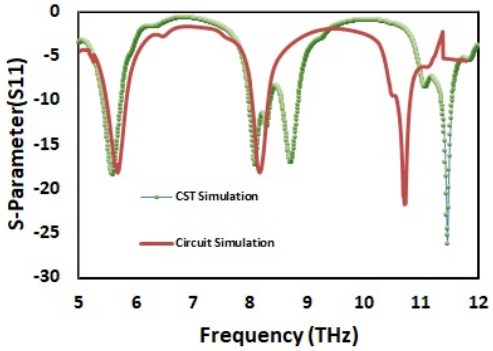
(c)



(d)



(e)

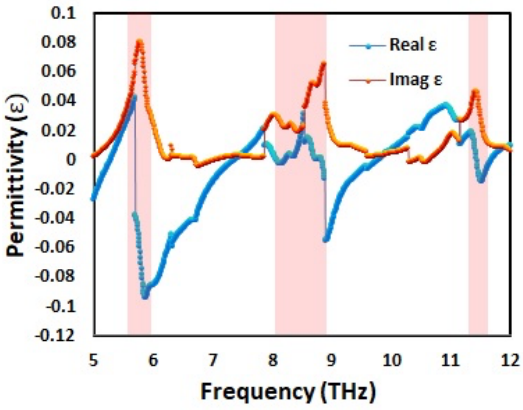


(f)

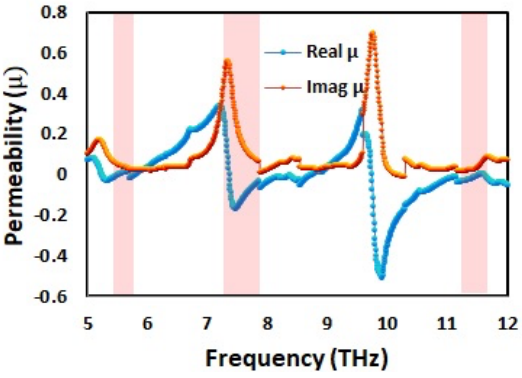
FIGURE 5 Simulated Results of proposed triple-band absorber (a) Return loss (b) Absorptivity and Reflectivity (c) Co & Cross S_{11} (d) Phase shift of proposed Multiband absorber at different frequencies, (e) S_{11} for individual ring structure of Radius R1, R2 & R3 indicating resonance frequency as 7.54THz, 9.41 THz and 10.418 THz respectively, (f) comparison of circuit simulation and CST simulation.

Figure 6 indicates the results of absorbers that are required to recognize the exotic behavior of the material in terms of effective relative permeability/permittivity and refractive index, due to geometrical alterations, not reflected in natural materials. Relative permeability is having a negative peak value, at the resonant frequencies. Figure 6 (a), (b), and (c) show the negative permittivity, permeability, and refractive index at three resonance frequencies respectively. Also, Figure 6 (d) indicates the material (Copper & FR4) behaviour in the desired frequency band. The single layer absorber can be modelled using Drude's formulation for the metal conductivity as given in (17)[39]:

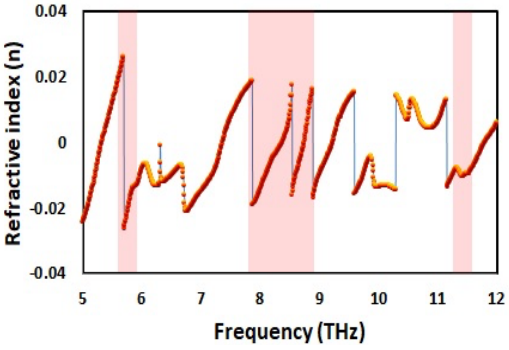
$$\sigma(\omega,\mu_c,\tau,t) = -j\frac{e^2 k_B t}{\pi h^2(\omega - j2\tau)}\left(\frac{\mu_c}{k_B t} + 2ln\left(e^{-\frac{\mu_c}{k_B t}} + 1\right)\right) \quad (17)$$



(a)



(b)



(c)

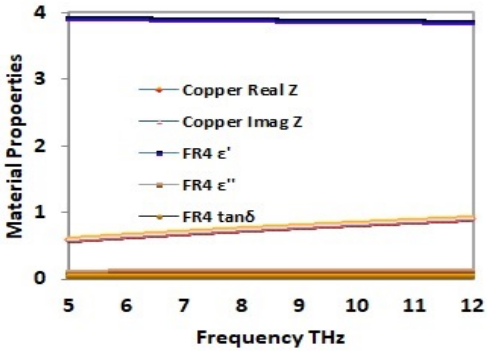
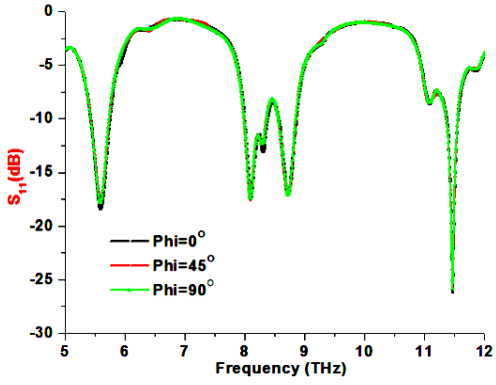


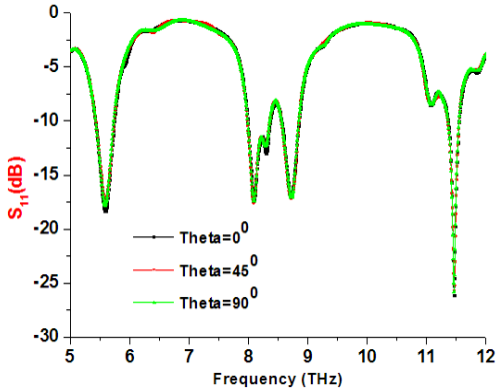
FIGURE 6 Simulated Results concerning the frequency sweep from 5 THz to 13 THz (a) Permittivity (b) Permeability(c) refractive index of the triple-band absorber (d) Material properties viz. Complex permittivity and Loss Tangent for FR4 and Impedance of copper conductor at the desired frequency band.

where, the parameters ω, μ_c, t, τ, h, e and k_B are radian frequency, chemical potential, temperature, scattering rate, Plank's constant, electronic charge and Boltzmann's constant.

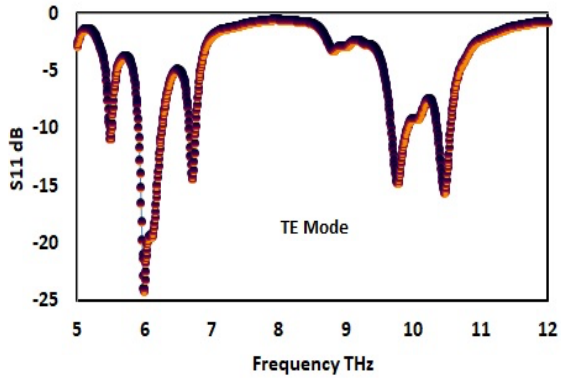
Figure 7 depicts the performance of the absorber under extreme conditions of orientation and wave incidence. In Figure 7 (a) polarization angle "ϕ" is varied from 0° to 90° with an incident wave at normal i.e. "$\theta = 0°$". It is observed that the absorber designed is insensitive to the polarization angle. The incidence angle is varied from 0° to 90° with "$\phi = 0°$", it is found that the absorber is also insensitive to incidence angle, thus reducing RCS at three frequency bands as shown in Figure 7(b). RCS reduction is used in stealth for fifth-generation aircraft to minimize the reflected power for detecting the stealth aircraft.



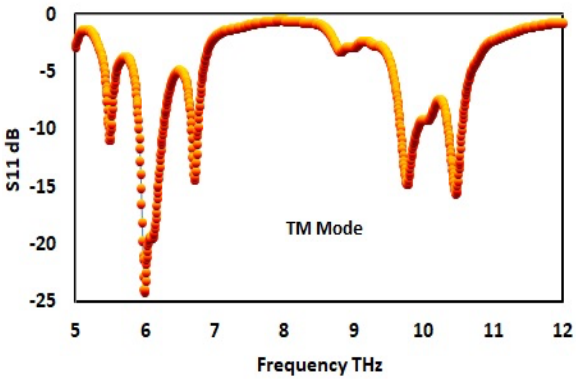
(a)



(b)



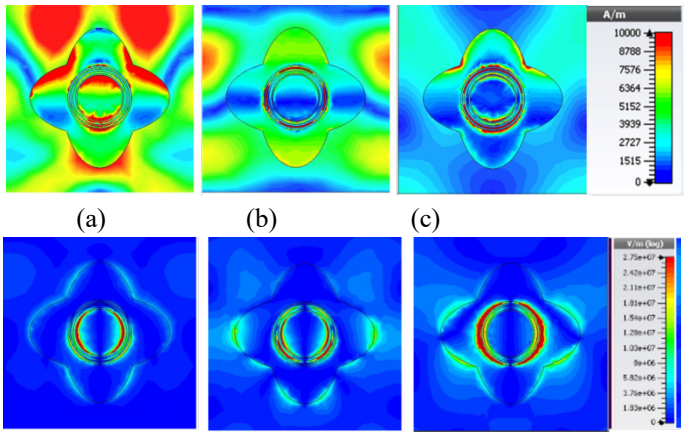
(c)



(d)

FIGURE 7 Simulated reflection coefficient "S_{11}" of the absorber for various (a) polarization angles "ϕ" at normal incidence (b) incidence angles "θ" at "ϕ = 0. "(c)TE polarization, (d) TM polarization

The polarization angle ϕ, is varied from 0° to 90°keeping incident angle θ = 0°, it can be observed from Figure 7(a)there is no variation in absorptivity, which indicates that the absorber is polarization insensitive. Also, it is clear from Figure 7(b) that if the incident angle θ is varied from 0° to 90°, the structure holds good reliability. Therefore, the absorber is insensitive to the angle of incidence of the approaching electromagnetic wave. Fig 7 (c) and (d) indicate the TE and TM polarization respectively.



(a) (b) (c)

(d) (e) (f)

FIGURE 8 Simulated magnetic field H_z at (a) 5.58THz, (b) 8.71THz and (c) 11.46THz and Electric field E_z at (d) 5.58THz, (e) 8.71THz and (f) 11.46THz of proposed absorber frequency respectively

It can be stated from the results that the designed absorber is insensitive to incident angle variations as well as polarization angle variations. The absorber structure comprising the ring resonators has peak absorptivity at three bands i.e.5.58THz, 8.71THz, and 11.46THz. Figure8 shows the structure behavior with the approaching electromagnetic wave indicating the electric and magnetic fields. The simulations are performed using the field monitors in the CST-microwave studio with appropriate periodic boundary conditions (PBC). The top and bottom walls are set with perfect electric conductor (PEC) boundary conditions. Also, perfect magnetic conductor (PMC) boundary conditions are applied to the side walls. The remaining sides are assigned as ports.

Figure8 (a)-(f) shows the magnetic and electric field variations with the operating frequency. Figure9 (a), (b), and (c) show that for lower frequencies the outer structure is responsible for absorption, and at higher frequencies, the electric field is concentrated more on the inner rings. The proposed MMA provides good absorbance and low profile compared with the referred articles as shown in Table-II.

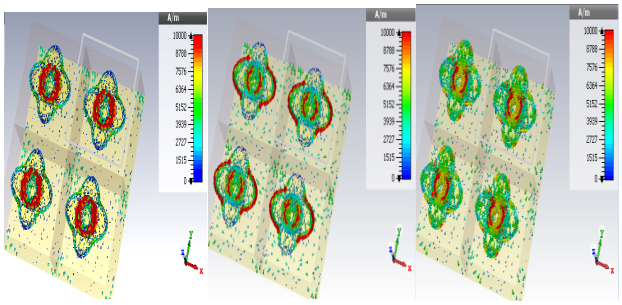


FIGURE 9 Surface current density of proposed absorber at (a) 5.58THz (b) 8.71THz and (c) 11.46THz respectively.

TABLE-II Comparison table of proposed MMA with existing reference

Frequency	Absorbance Band	Absorbance (%)	Thickness (↖)	Polarization insensitivity/ Reference
0.2–2.4 THz	2	>75	0.13↖	Yes[6]
0.1–1 THz	5	>90	0.036↖	No[13]
0–2 THz	5	>94	0.027↖	No[14]
0–4.5 THz	5	>98	1.515↖	Yes[15]

0.5–4 THz	4	>75	0.13λ	Yes[16]
1.9 & 5 THz	2	>95	0.68λ	No[23]
0.1-2THz	6	>95	0.13λ	Yes[25]
1-8THz	4	>90	-	Yes[26]
8-14GHz	3	>98	0.010λ	Yes[27]
2-5THz	4	>99	0.134λ	Yes[28]
4-12THz	7	>90	0.274λ	Yes[29]
6-18GHz	5	>90	-	Yes[35]
85-345THz	4	>99	-	Partial[36]
5.58-11.45THz	**3**	**>91.2**	0.013λ	**Yes [Proposed]**

IV. CONCLUSION

A single layer ultra-thin multi-band THz MMA with polarization-insensitive has been presented with high absorbance at three distinct frequencies 5.58, 8.71, and 11.45THz. The proposed structure has high absorbance with ultra-thin 0.013λ thickness with existing literature used in Table-II. So this MMA is used in many defense applications like stealth Fifth-generation stealth aircraft for RCS reduction, MIMO antenna isolation improvement, EMI shielding, energy harvesting, EMC reduction, THz imaging [18], optoelectronic devices, and sensing in the intended band.

REFERENCES

[1] Hui Feng Ma, Sanming Hu, Yun Gui Ma, Yun Lai, and Karu Esselle, "The Applications of Metamaterials", International Journal of Antennas and Propagation, vol. 20, pp.1-2, (2015).
[2] Mohammed R Hashemi, Semih Cakmakyapan, and Mona Jarrahi, "Reconfigurable metamaterials for terahertz wave manipulation", Reports on Progress in Physics vol. 80, pp. 1-42 (2017).
[3] Y. Khanna and Y. K. Awasthi, "Wideband Ultra-thin Metamaterial Absorber for Ku & K- Band Applications," 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2020.
[4] Saxena, Gaurav, Y. K. Awasthi, and Priyanka Jain. "Design of Metasurface Absorber for Low RCS and High Isolation MIMO Antenna for Radio Location & Navigation." AEU-International Journal of Electronics and Communications, vol. 20, pp. 31171-2(2021).
[5] R. Xing, C. Dong, B. Zhu, Y. Gao, and S. Jian, "Numerical Analysis on Ribbon-ArraySheet Coupled Graphene Terahertz Absorber," in IEEE Photonics Technology Letters, vol. 28, pp. 2207-2210, 2016.
[6] Y. Wang et al., "Properties and Sensing Performance of All-Dielectric Metasurface THz Absorbers," in IEEE Transactions on Terahertz Science and Technology, vol. 10, pp. 599-605, 2020.
[7] R. Mishra, R. Panwar and D. Singh, "Equivalent Circuit Model for the Design of Frequency-Selective, Terahertz-Band, Graphene-Based Metamaterial Absorbers," in IEEE Magnetics Letters, vol. 9, pp. 1-5, 2018.
[8] Y. Wang et al., "Properties and Sensing Performance of All-Dielectric Metasurface THz Absorbers," in IEEE Transactions on Terahertz Science and Technology, vol. 10, pp. 599-605, 2020.
[9] M. Qu, T. Chang, G. Guo, and S. Li, "Design of Graphene-Based Dual-Polarized Switchable Rasorber/Absorber at Terahertz," IEEE Access, vol. 8, pp. 127220- 127225, 2020.
[10] M. Biabanifard, A. Arsanjani, M. S. Abrishamian, and D. Abbott, "Tunable Terahertz Graphene-Based Absorber Design Method Based on a Circuit Model Approach," IEEE Access, vol. 8, pp. 70343-70354, 2020.
[11] Jin Huang, Jining Li, Yue Yang, Jie Li, Jiahui Li, Yating Zhang, and Jianquan Yao, "Broadband terahertz absorber with a flexible, reconfigurable performance based on hybrid-patterned vanadium dioxide metasurfaces," Optics Express vol. 28, pp. 17832-17840, 2020.
[12] A. Chen and Z. Song, "Tunable Isotropic Absorber With Phase Change Material VO2," in IEEE Transactions on Nanotechnology, vol. 19, pp. 197-200, 2020.
[13] He, Y., Wu, Q. & Yan, S. "Multi-Band Terahertz Absorber at 0.1–1 THz Frequency Based on Ultra-Thin Metamaterial", Plasmonics, vol. 14, pp. 1303–1310, 2019.
[14] Sabah, C., Mulla, B., Altan, H. et al. "Cross-like terahertz metamaterial absorber for sensing applications", Pramana - J Phys vol. 91, pp.17-23, 2018.
[15] Tianhua Meng, Dan Hu, Qiaofen Zhu, "Design of a five-band terahertz perfect metamaterial absorber using two resonators", Optics Communications, vol. 415, pp. 151-155, 2018.
[16] Yang Liu, Sheng-Qiu Xu, Ming Liu, Xin-Guang Hu, Yong-Fa Duan, Lin Yi, "Tunable multiband terahertz absorber based on a one-dimensional heterostructure containing semiconductor", Optik, vol. 170, pp. 203-209, 2018.
[17] Saxena, Gaurav, Y. K. Awasthi, and Priyanka Jain. "High Isolation and High Gain Super-Wideband (0.33-10 THz) MIMO Antenna for THz Applications." Optik, vol. 223, pp. 165335, 2020.
[18] Khanna, Y., Awasthi, Y.K. Ultra-Thin Wideband Polarization-Insensitive Metasurface Absorber for Aviation Technology. Journal of Electronic Materials, vol. 49, pp. 6410–6416, 2020.
[19] Y. Dong, P. Liu, D. Yu, G. Li, and L. Yang, "A Tunable Ultra-broadband Ultrathin Terahertz Absorber Using Graphene Stacks," in IEEE Antennas and Wireless Propagation Letters, vol. 16, pp. 1115-1118, 2017.
[20] W. Pan, X. Yu, J. Zhang, and W. Zeng, "A Novel Design of Broadband Terahertz Metamaterial Absorber Based on Nested Circle Rings," in IEEE Photonics Technology Letters, vol. 28, pp. 2335-2338, 2016.
[21] Saxena, Gaurav, R. L. Yadava, Priyanka Jain, and Y. K. Awasthi. "Triple Band Polarization Insensitive Ultra-Thin Metamaterial Absorber for EMC and RCS Reduction in X-Band Applications." In IEEE, 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 772-775. 2020.
[22] Khan, B., Kamal, B., Ullah, S. et al. Design and experimental analysis of dual-band polarization converting metasurface for microwave applications. Scientific Reports vol. 10, pp.15393, 2020.
[23] Barzegar-Parizi, Saeedeh, Amir Ebrahimi, and Kamran Ghorbani. "High-Q dual-band graphene absorbers by selective excitation of graphene plasmon polaritons: Circuit model analysis." Optics & Laser Technology 132 (2020): 106483.
[24] Yan, Fei, Li Li, Ruoxing Wang, Hao Tian, Jianlong Liu, Jianqiang Liu, Fengjun Tian, and Jianzhong Zhang. "Ultrasensitive tunable terahertz sensor with graphene plasmonic grating." Journal of Lightwave Technology vol. 37, pp. 1103-1112, 2019.
[25] Biabanifard, Mohammad, and Mohammad Sadegh Abrishamian. "Multi-band circuit model of tunable THz absorber based on graphene sheet and ribbons." AEU-International Journal of Electronics and Communications vol. 95 pp. 256-263, 2018.
[26] Barzegar-Parizi, Saeedeh, and Amir Ebrahimi. "Ultrathin, polarization-insensitive multi-band absorbers based on graphene metasurface with THz sensing application." JOSA B vol. 37, pp. 2372-2381,2020.
[27] Wu, Jun. "Tunable multi-band terahertz absorber based on graphene nano-ribbon metamaterial." *Physics Letters A* vol. 383, pp. 2589-2593, 2019.
[28] Zamzam, Pouria, Pejman Rezaei, and Seyed Amin Khatami. "Quad-band polarization-insensitive metamaterial perfect absorber based on bilayer graphene metasurface." Physica E: Low-dimensional Systems and Nanostructures vol. 128, pp. 114621, 2021.
[29] Cheng, Y., Zou, Y., Luo, H. et al. Compact Ultra-Thin Seven-Band Microwave Metamaterial Absorber Based on a Single Resonator Structure. Journal of Elec Materi vol. 48, pp. s3939–3946, 2019.
[30] Wang, Yingying, Zeqiang Chen, Danyang Xu, Zao Yi, Xifang Chen, Jian Chen, Yongjian Tang, Pinghui Wu, Gongfa Li, and Yougen Yi. "Triple-band perfect metamaterial absorber with good operating angle polarization tolerance based on split ring arrays." Results in Physics vol. 16, pp. 102951, 2020.
[31] Amiri, Majid, Farzad Tofigh, Negin Shariati, Justin Lipman, and Mehran Abolhasan. "Wide-angle metamaterial absorber with highly insensitive absorption for TE and TM modes." Scientific reports vol. 10, pp. 1-13, 2020.
[32] Ghods, Mohammad Mahdi, and Pejman Rezaei. "Ultra-wideband microwave absorber based on uncharged graphene layers." Journal of Electromagnetic Waves and Applications vol. 32, pp. 1950-1960, 2018.
[33] Yi, Zao, Hang Lin, Gao Niu, Xifang Chen, Zigang Zhou, Xin Ye, Tao Duan, Yong Yi, Yongjian Tang, and Yougen Yi. "Graphene-based tunable triple-band plasmonic perfect metamaterial absorber with good angle-polarization-tolerance." Results in Physics vol. 13, pp. 102149, 2019.

[34] Yongzhi Chenga, Hui Luo, and Fu Chen, “Broadband metamaterial microwave absorber based on asymmetric sectional resonator structures”, Journal of Applied Physics vol. 127, pp. 214902, 2020.
[35] Fen Zhang, Chao Jiang, Qiang Wang, Zixiang Zhao, Yan Wang, Zuojuan Du, Chunqi Wang and Xiaozhong Huang, “A multi-band closed-cell metamaterial absorber based on a low-permittivity all-dielectric structure”, Applied Physics Express, vol. 13, pp. 084001, 2020.
[36] Hongsen Zhang, Yongzhi Cheng, Fu Chen, “Quad-band plasmonic perfect absorber using all-metal nanostructure metasurface for refractive index sensing”, Optik, vol. 229 pp. 166300, 2021.
[37] Limie Qi and Chang Liu, “Terahertz wide angle metamaterial absorber fabricated by printed circuit bosrd technique”, Journal of Applied Physics, vol. 124, pp. 153101, 2018.
[38] Ben-Xin Wang, Yuanhao He, Pengcheng Lou and Wenhui Xing, “Design of a dual-band terahertz metamaterial absorber using two identical square patches for sensing application” Nanoscale Advances, vol. 2, pp 763-769, 2020.
[39] Gaurav Varshney, “Reconfigurable Graphene Antenna for THz Applications: A Mode Conversion Approach” Nanotechnology, vol. 31, pp. 135208, 2020.

International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934

ORIGINAL ARTICLE

Multi-purpose hybrid optimization (DA-ALO) algorithm for efficient job shop scheduling issues

Vineet Kumar[1], Radhey Shyam Mishra[2]

[1]*Department of Mechanical Engineering, IKG Punjab Technical University, Jalandhar, India*
[2]*Department of Mechanical Engineering, Delhi Technological University Delhi, India*

Article Information

Received: 08 December 2020
Revised: 11 January 2021
Accepted: 18 January 2021
Available online: 23 January 2021

Keywords:

Job Shop Scheduling Problem;
Make Span Time;
Dragon fly Algorithm;
Ant Lion Optimization
Algorithm, Hybrid Algorithm

Abstract

The Job-Shop Scheduling (JSSP) is the most vital modern exercises, particularly in manufacturing industry. In this paper the Hybrid (DA-ALO) is presented in the fasteners industry, combination type plant layout is used. There are numerous types of machines, numerous operations to be performed on the numerous jobs. This is the reason in these days for controlling the make span time in the fasteners industry. For solving the scheduling problem in the fasteners industry, it is necessary to use hybrid algorithms i.e. Dragon fly Algorithm (DA) and Ant Lion Optimization(ALO) algorithm that forms Hybrid (DA-ALO). It is concluded that there is 2% gain in the make span time by using the hybrid model. The above said gain is very significant and recommended in the fasteners industry. ©2021 ijrei.com. All rights reserved

1. Introduction

Each job consists of a group of operation that required to be controlled in a fixed pre-specified order and every operation is to be operated on a special resource for a predetermined duration of time. In this context, emphasis has been done to minimizing the completion time of the end task of the job. In the last two decades' metaheuristics approaches such as PSO algorithm, ABC algorithm, GA etc. have been applied to resolve the JSSP. A couple of frameworks have been planned in the writing for JSSP that is essentially in view of the optimization algorithm, for example, Dragonfly Optimization (DO) algorithm and Antlion Optimization (ALO) is assessed for beating the issues. This segment depicts to limit the makes span time and tardiness for different jobs and the ideal esteem is acquired by the mixture (DA-ALO) optimization algorithm by conquering the benchmark issues and getting the minimize the make span time. In this paper 17 benchmark problems have been taken of the fasteners industry and each problem consists of numerous types of products and corresponding required machines. One problem i.e. LA21 whose size is 15×20. It means that there are 15 kinds of Jobs/Products having different lengths, diameters and designs & each product is required 20 machines irrespective of their size and shape. It is one of the complex and complicated problem among the all 17 benchmark problems. The developed algorithm (Hybrid DA & ALO) will be executed on the working platform

Corresponding author: R.S. Mishra
Email Address: hod.mechanical.rsm@dtu.ac.in
https://doi.org/10.36037/IJREI.2021.5107

of MATLAB 2015a & the outcome/result will be studied and compared with the traditional strategies of DA and ALO algorithm. Sharma Nirmala et al 2018 [1] had analyzed JSSP is an imperative combinatorial optimization issue in the field of machine scheduling. This article exhibits an adjusted ABC algorithm to explain JSSP. The (Beer froth artificial bee colony algorithm) Be F ABC has been surveyed on 25 benchmark test issues and contrasted and other condition of-craftsmanship calculations. Further, it is connected to tackle 62 surely understood occasions of discrete JSSP. The acquired numerical outcomes and factual examination delineate that the proposed algorithm is equipped in managing the discrete JSSP. Huang Rong-Hwa et al 2017 [2]. had researched a few novel hybrid ant colony optimizations (ACO)- based algorithms to determine multi-target job-shop scheduling issue with a square with estimate parcel part. The fundamental issue examined in this strategy is part of jobs and tradeoff between parcel part expenses and make the traverse. The destinations that are utilized to quantify the nature of the created plans are weighted-total of make traverse, the lateness of jobs and parcel part cost. An ACO based calculation is utilized to produce calendars and improve the outcome and minimize the make span time tardiness. Li Jun-qing et al 2013 [3] had proposed a hybrid algorithm joining particle swarm optimization (PSO) and Tabu Search (TS) was proposed to take care of the job shop scheduling issue with fuzzy handling time. The question is to limit the greatest fuzzy fulfillment time, i.e., the fuzzy make span. In the proposed algorithm, PSO plays out the worldwide hunt, i.e., the investigation stage, while TS leads the neighborhood to seek, i.e., the abuse procedure. The proposed algorithm is tried on sets of the notable benchmark occurrences. Through the examination of trial comes about, the exceedingly compelling execution of the proposed algorithm appears against the best performing outcome. KS Sree Ranjini et al 2017 [4]. had proposed optimization algorithm in view of the static and dynamic swarming conduct of dragonflies. Because of its effortlessness and effectiveness, DA has gotten enthusiasm of specialists from various fields. The p best and g best idea of Particle Swarm Optimization (PSO) are added to regular DA to manage the look procedure for potential candidate arrangements and PSO is then introduced with p best of DA to additionally misuse the hunting space. The outcomes demonstrate that MHDA gives preferred execution over customary DA and PSO. Also, it gives focused outcomes as far as meeting, precision, and pursuit capacity when contrasted with state-of-the-art algorithms. Zhang Rui et al 2014 [5]. had proposed to propose a hybrid Differential Evolution (DE) algorithm for the job shop scheduling issue with random handling times under the target of limiting the normal aggregate lateness (a measure of benefit quality). To begin with, they propose an execution evaluate for generally looking at the nature of candidate arrangements. At that point, a parameter perturbation algorithm was connected as a neighborhood look module for quickening the joining of DE. At long last, the K-outfitted bandit display is used for decreasing the computational weight in the correct assessment of arrangements in view of recreation. The computational outcomes on various scale test issues approve the viability and proficiency of the proposed approach.

2. Methodology

Job Shop Scheduling Problem (JSSP) has been to a great degree of the most important industrial model and it has ascended insignificance because of the demands of industry. The methodology gives a JSSP issue set of jobs on an arrangement of machines and each of having particular and conceivably extraordinary times and conveyance. The object is to discover a preparing time which limits a given make time to lessen the cost. This examination of JSSP gives almost seventeen benchmark issues are considered to lessen the make span time used to empower distinctive optimization techniques. The make span time and tardiness esteem limits and getting the target work as best by utilizing the hybrid (DA-ALO) optimization algorithms. This methodology gives the optimal yield esteem with limit the time when contrasted with actual BKS.

2.1 Assumption

Consider the different 17 benchmarks issue and this strategy is identified with the job and the procedure utilized different methodologies. To assess the execution of the framework, most often considered execution measure, for example, minimization of make span for add up to finishing to the time of the procedures are considered. The procedure was legitimate with any sort of job and framework expects the adaptable job shop conditions where period might be skipped. This job the resulting hypothesis is being done and they are as clarified underneath.

- Let J_i = {1,2, ...n} be set n jobs are scheduled.
- M_j = {1,2, ...m} be a set m machines are schedule.
- At a time, it is possible to process only one operation on one machine.
- Execution of each operation M j requires a resource from a set of alternative machines
- Processing of operations on the machines should not be interrupted.
- Release times and due dates are not specified.

The make span value to be definite as below: each job has a number of operations and each function has an explicit make value at this place

2.2 Objective Function

This job examination considers multi-target initial one is slightest make span time and of different jobs performed in machine.

$$M_j = \min (M_t) \quad (1)$$

2.3 An Optimization Algorithm for JSSP

Consider different jobs with the relating machine and every one of them achieve an alternate time and distinctive conveyance period. We will probably limit the make span time and tardiness by utilizing the optimization algorithm, for example, Dragonfly (DA) and Ant Lion (AL) Optimization Several benchmark issues are settled by the JSSP issue with hybrid (DA-AL) and getting the optimal incentive with BKS.

.

2.4 Hybrid Model for (DA-ALO)

In JSSP, the two swarms-based hybrid approach model is utilized to get the ideal value. This hybrid approach is acquired to resolve optimization scheduling issues, for example, seventeen benchmark issues and it needs to limit the make span time and tardiness for each job and machines with ideal to esteem

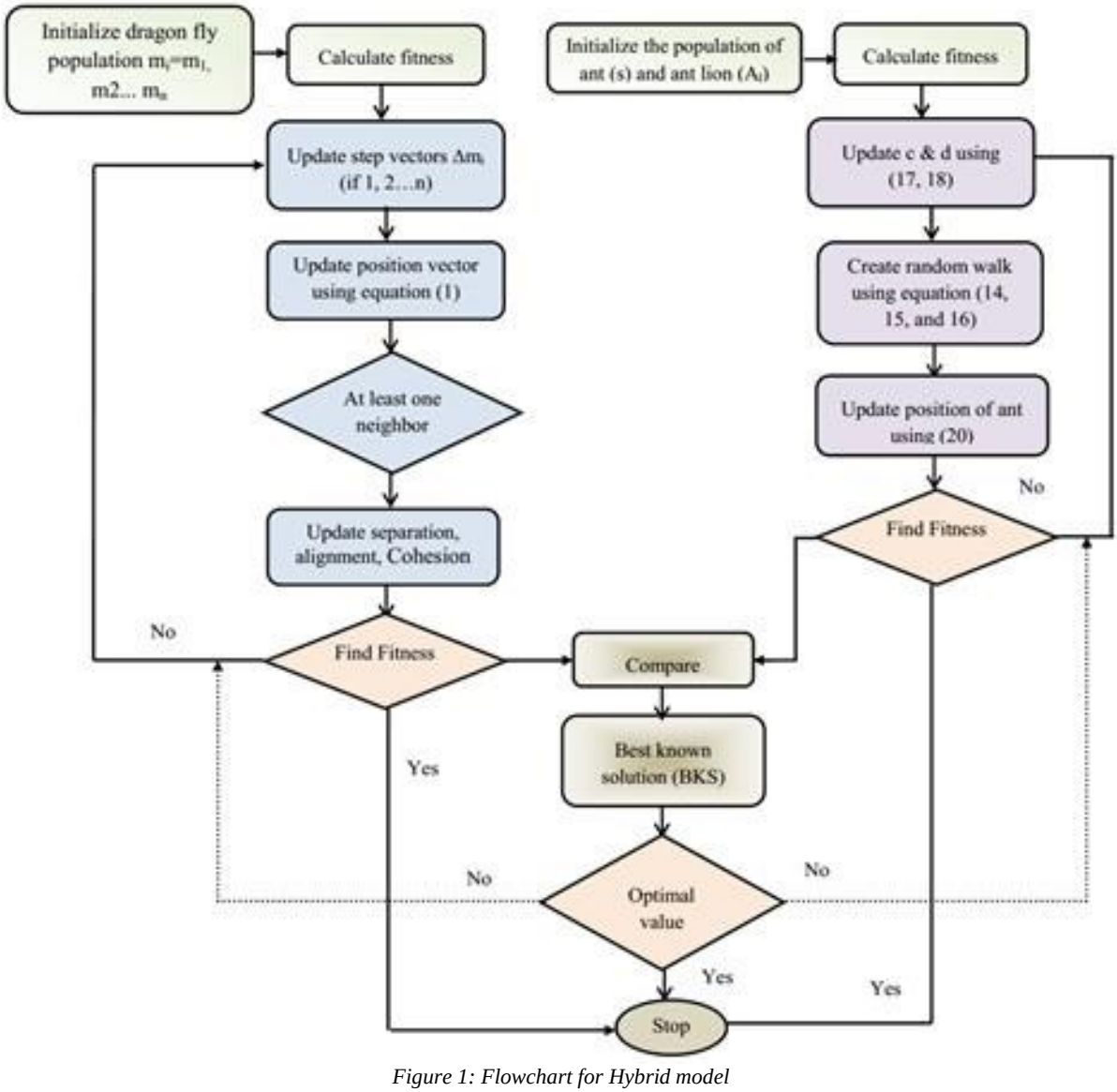


Figure 1: Flowchart for Hybrid model

The hybrid model algorithm is explained in the above figure. The detailed explanation for DA and ALO algorithm with the behavioral process has been shown. The left portion depicts the DA algorithm and right portion depicts the ALO algorithm and compare the output with both the algorithms. Whichever the best-known solution, then check it for optimal value. If yes, then stop and if no repeat the whole process for DA as well as ALO algorithm.

3. Results and Discussion

Fig-1 demonstrates the comparable analysis of make span time for the benchmark problems ABZ5, ABZ6, FT10, LA05, LA10, LA11, LA12, LA14. The performance of proposed algorithms is analyzed and compared among one another. For the problem ABZ5, DA takes 1250 minutes, ALO takes 1260 minutes and hybrid DA-ALO takes 1215 minutes. From the analysis, the least make span time is gained in the proposed DA-ALO mathematical model. Similarly, make span time comparative analysis of other problems like ABZ6, FT10, LA05, LA10, LA11, LA12, LA14. is depicted in that figure-2.

The comparative analysis of make span time for the benchmark issues LA21, LA30, LA31, ORB01, ORB02, ORB03, ORB04, ORB05, ORB06 have been carried out. For the benchmark problem LA21, DA takes 1010 minutes, ALO takes 1015 minutes and DA-ALO takes 990 minutes. Similarly, for LA30 benchmark problem, DA takes 1380 minutes, ALO takes 1405 minutes and DA-ALO takes 1360 minutes. In the same manner, make span time a comparative analysis of other benchmark problems like LA31, ORB01, ORB02, ORB03, ORB04, ORB05, ORB06 are represented in chart. On comparing those benchmark issues, the least make span time is accomplished by the developed hybrid DA-ALO algorithm

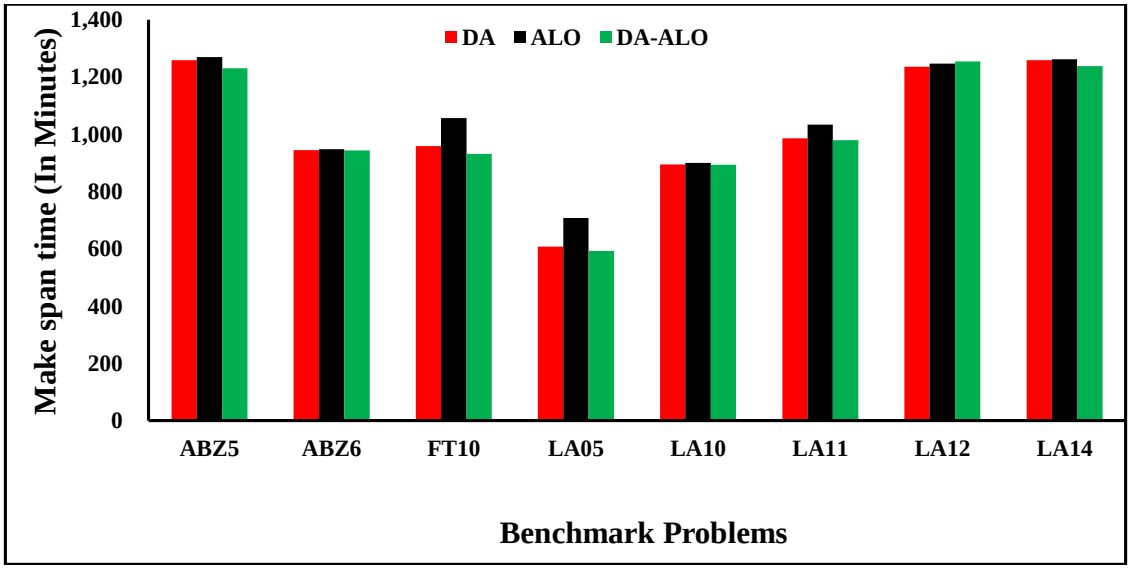


Figure 2: Comparable analysis of make span time for the benchmark problems

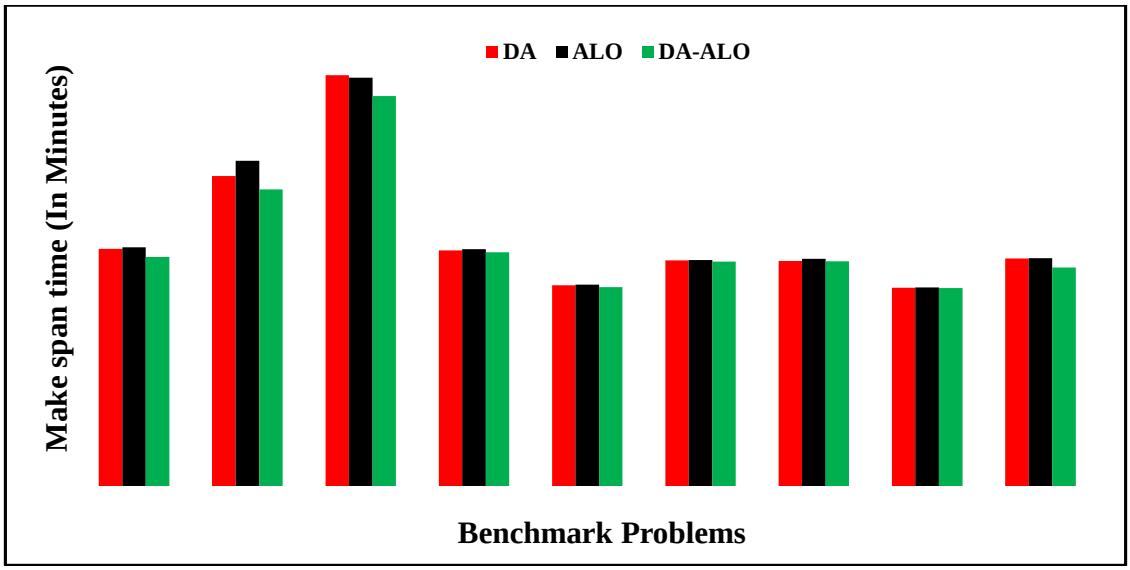


Figure 3: Comparable analysis of make span time for the benchmark problems

Fig-3 shows the bar chart which illustrates the comparative analysis of make span time for the benchmark issues LA21, LA30, LA31, ORB01, ORB02, ORB03, ORB04, ORB05, ORB06. For the benchmark problem LA21, DA takes 1010 minutes, ALO takes 1015 minutes and DA-ALO takes 990 minutes. Similarly, for LA30 benchmark problem, DA takes 1380 minutes, ALO takes 1405 minutes and DA-ALO takes 1360 minutes. In the same manner, make span time a comparative analysis of other benchmark problems like LA31, ORB01, ORB02, ORB03, ORB04, ORB05, ORB06 are represented in chart. On comparing those benchmark issues, the least make span time is accomplished by the developed hybrid DA-ALO algorithm.

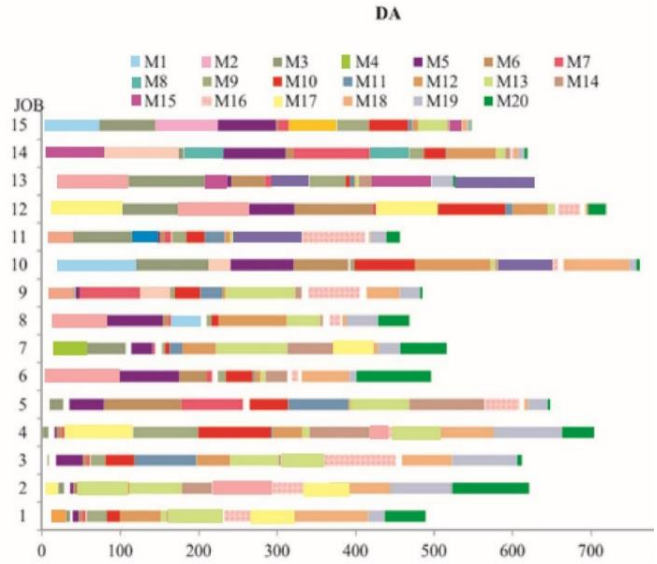


Figure 4: Gantt-Chart found by the DA based method for the LA21 (15X20) problem

The Fig-4 presents the Gantt-Chart found by the DA based method for the LA21 (15X20) problem. In the Fig-4, different combination of operations was done for LA21 (15X20) problem and found the make span time. While running the DA based algorithm, comparison has been done for make span time of all the products on all the machines.

Fig-5 presents the Gantt-Chart found by the algorithm ALO based method for the LA21 (15X20) problem. In the Figure, it is noted that a different combination of operations can be done for LA21 (15X20) problem and While running the ALO based algorithm, comparison has been done for make span time of all the products on all the machines. Fig-6 presents the Gantt-Chart found by the developed hybrid algorithm DA-ALO based method for the LA21 (15X20) problem. In the Figure, it is noted that a different combination of operations can be done for LA21 (15X20) problem and found the make span as a minimum. When comparing all algorithms, the developed hybrid DA-ALO achieves minimum make span time for all benchmark problems.

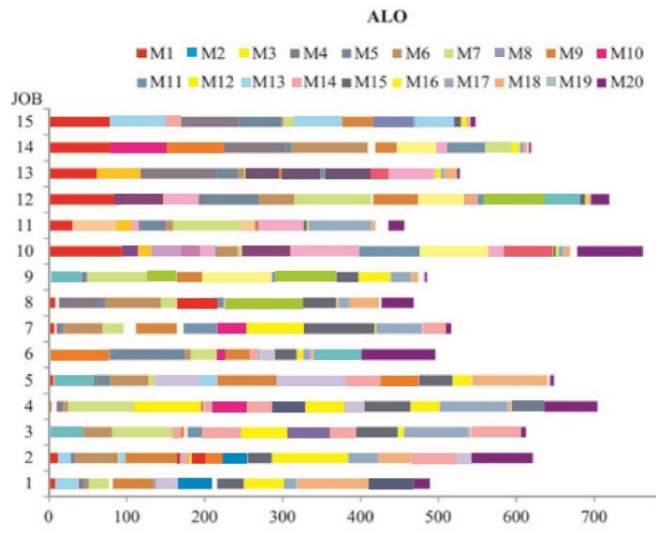


Figure 5: Gantt-Chart by the algorithm ALO based method for the LA2

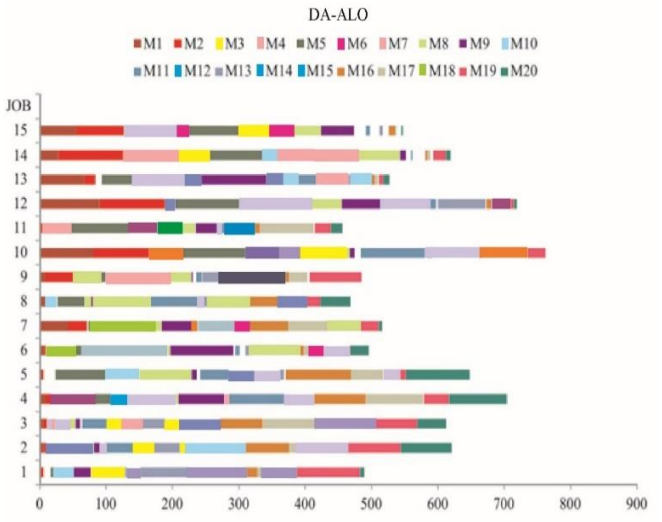


Figure 6: Gantt-Chart hybrid algorithm DA-ALO

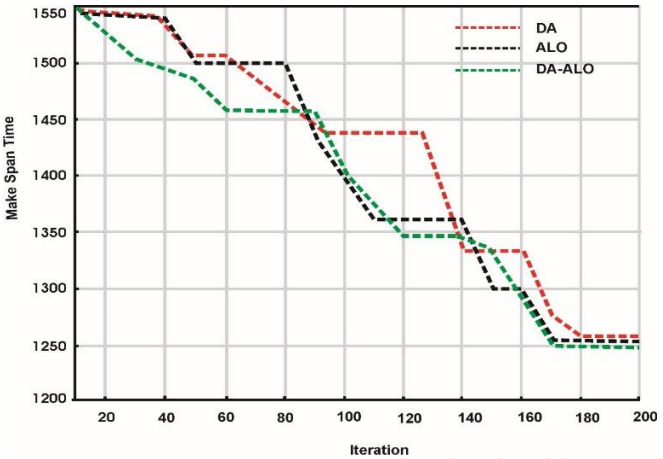


Figure 7: Convergence in ORB01 benchmark problem

The convergence graph analyzing the benchmark problems of *ORB01*. The performance of different algorithms such as DA, ALO and DA-ALO algorithm for 20th, 40th, 60th, 80th, 100th, 120th, 140th, 160th, 180th& 200thiteration are shown in fig-7 respectively for the benchmark problem ORB01 by DA algorithm, in Fig-7, it was found that the fitness value starts from 1550 and then decrease to 1250 at 175 iterations and remains constant up to 200th iteration. For ALO algorithm the fitness value starts from 1550 and then decrease to 1255 at 170 iterations and remains constant up to 200th iteration. For the hybrid DA-ALO algorithm the fitness value starts from 1550 and then decrease to 1245 at 170 iterations and remains constant up to 200th iteration. Fig-08: shows the convergence in LA11. It was found in the DA algorithm, the fitness value starts from 1550 and then decrease to 1340 at 150 iterations and remains constant up to 200th iteration. For ALO algorithm the fitness value starts from 1550 and then decrease to 1245 at 190 iterations and remains constant up to 200th iteration. For the hybrid DA-ALO algorithm the fitness value starts from 1550 and then decrease to 1230 at 110 iterations and remains constant up to 200th iteration.

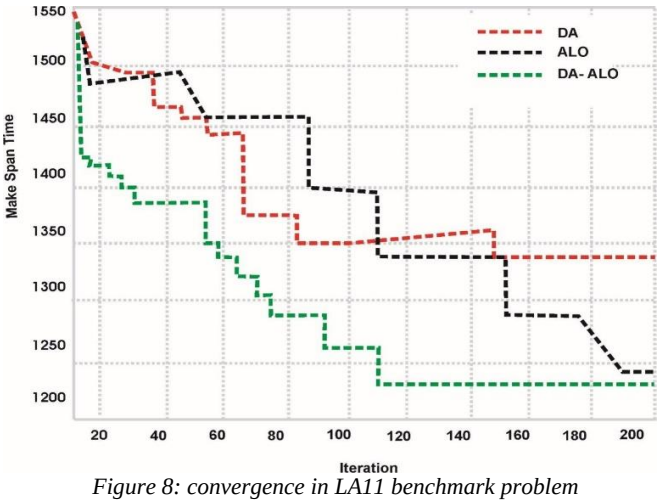


Figure 8: convergence in LA11 benchmark problem

4. Conclusions

From the validation perspective, I have compared all 17 benchmark issues by the optimization algorithm such as DA, ALO and hybrid (DA-ALO) is utilized to obtain optimal values with best fitness function of the algorithm and minimize the given objectives. The DA model has been used and same model has been implemented in the industry. Results are obtained. Another model using Ant Lion (ALO) has also used. The model has been implemented on the same set of data as for DA algorithm and deduce the result. The main model by hybrid of above two algorithms has been developed. The same set of data collected from the industry has been implemented on the developed model i.e. (Hybrid of DA and ALO). It is observed that the result of the hybrid model developed in present work are better than the two independent models i.e. DA and ALO. On observing the result, it is concluded that there is 2% gain in the make span time by using the hybrid model. The above said gain is very significant and recommended in the industry under observation.

References

[1] Sharma, N., Sharma, H., & Sharma, A. Beer froth artificial bee colony algorithm for job-shop scheduling problem. Applied Soft Computing, 68, 2018, 507-524.

[2] Huang, R. H., & Yu, T. H. An effective ant colony optimization algorithm for multi-objective job-shop scheduling with equal-size lot-splitting. Applied Soft Computing, 57, 2017, 642-656.

[3] Li, J. Q., & Pan, Y. X. A hybrid discrete particle swarm optimization algorithm for solving Fuzzy job shop scheduling problem. The International Journal of Advanced Manufacturing Technology, 66, 2013, 583-596.

[4] Manukumar N.M and Manjunatha Hiremath, A Hybrid Algorithm for Face Recognition using PCA, LDA and ANN, International Journal of Mechanical Engineering and Technology 9, 2018, 85–93.

[5] Nikhil Gajra, Shamsuddin S. Khan and Pradnya Rane, Private Cloud Data Security: Secured User Authentication by using Enhanced Hybrid Algorithms, International Journal of Computer Engineering and Technology (IJCET), 5, 2014, 80-87.

Cite this article as: Vineet Kumar, R.S. Mishra, Multi-purpose hybrid optimization (DA-ALO) algorithm for efficient job shop scheduling issues, International journal of research in engineering and innovation (IJREI), vol 5, issue 1 (2020), 62-67. https://doi.org/10.36037/IJREI.2021.5107.

Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Multiview Running and Walking Gait Analysis using the First and Third Person Data

To cite this article: Nikita Malik and Sudipta Majumdar 2021 *J. Phys.: Conf. Ser.* **2070** 012138

View the article online for updates and enhancements.

You may also like

- Validity, reliability, and feasibility of the uSense activity monitor to register physical activity and gait performance in habitual settings of geriatric patients
Martin Bongartz, Rainer Kiss, André Lacroix et al.
- The Research of Gait Recognition Based on Deep Learning: A Case Study of the Missing Elderly
Mingyang Liu
- Human gait recognition based on feature extraction of support vector machine and pattern network algorithm
Shahla A. AbdAlKader



This content was downloaded from IP address 14.139.251.100 on 01/12/2021 at 05:29

Multiview Running and Walking Gait Analysis using the First and Third Person Data

Nikita Malik[1]*, Sudipta Majumdar[1]

[1]Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, 110042, India

E-mail: *nikitamalik70@gmail.com

Abstract. Gait recognition, which allows to recognise someone from a distance, has a lot of applications. The gait of a person is commonly used as a biometric approach to identify or categorise them by gender and age. Biometric systems are a fast-growing subject that necessitates the development of creative solutions to problems that have plagued previous attempts. By initially looking at the design of a gait detection system, two different types of gait datasets have been reported and presented in this paper. While running at the same time, the first person (FP) data containing the camera motion gathered from the movement of the volunteer's body and the third person (TP) data captured from a distant view were recorded. The dataset contains a total of 23 participants (14 males and 9 females). The discussion is expanded to include a comparison of the results obtained using TP and FP data, as well as an examination of the physiological motions recorded while running and walking.

1. Introduction

Biometric systems have been around for a long time, but they are continually evolving as new and effective human identification methods are added to the list. Human gait recognition is one of those newer approaches that has seen a lot of research and yet requires more. The study of human motion using brain and eye observations, as well as sensors that detect bodily movements, is known as human gait. The technology can be used to do remote biometric identification without the need for an observer. It excels in the field of biometrics due to its ability to recognize objects from a distance. It is a technology that extracts features from a person's body motions in order to identify them. It has a wide range of applications and is commonly utilized in sports, health monitoring, surveillance, criminal investigation, and gender and age classification. Many gait acquisition technologies exist, including wearable sensors data (accelerometers, gyroscopes, pressure and force sensors), high-definition cameras, and non-wearable sensors data. Model-based and appearance-based strategies are the two types of gait identification approaches. Some earlier state-of-the-art surveys are available for gait recognition which provides a detailed overview regarding the progress made so far in the area of human gait recognition systems. The evaluation part has been a history and the area of interest for many researchers is the application part in which not much evaluation has been done till now. The work is divided into six sections. In section 2, the framework of the gait recognition system is briefly outlined, followed by a full description of the proposed work in section 3. In addition, in section 4, the recorded dataset including the running patterns of 23 volunteers is described and investigated, and in section 5, the results acquired from the databases are compared and discussed, followed by a conclusion and a scope for future work.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

2. Framework of the Gait Recognition System

Gait data acquisition, pre-processing of the acquired data, feature extraction, data reduction or feature selection, and classification and recognition are the five stages of the gait identification system, as shown in fig.1.

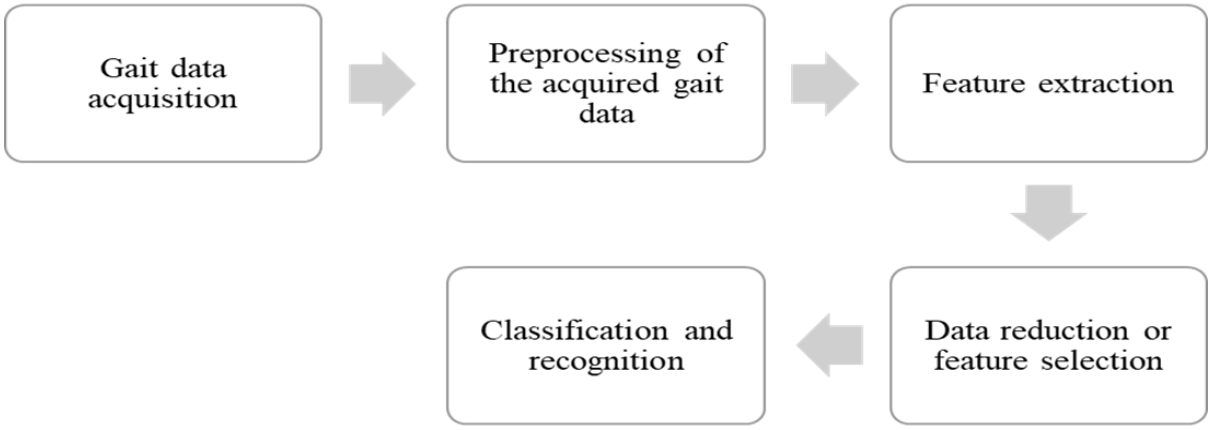


Figure 1. Framework of Gait Recognition System.

2.1. Data Recording and Preprocessing Techniques

Sensor-based, video-based, and radar-based data collecting techniques are all viable options. [1] The data recording and processing techniques are accelerometers, floor-sensors, camera, and radar. Depending on the nature of the application, the cell phones are increasingly incorporating accelerometers. The device's reduced size allows it to be easily carried in a pocket. The device makes use of Bluetooth with a distant station to carry out the connection. The calculations are done at a remote location. The subject of the investigation must be required to carry the gadget with him at all times.

RGB-D camera and Microsoft Kinect cameras are some examples of the data acquisition camera-based devices in which the RGB-D camera is used to input information about a person's depth and color. Kinect is a Microsoft-designed device that uses an infrared projector and sensors to acquire depth data. The infrared laser fired by Kinect creates an erected design of the bodily surface. Based on the principle of structured light, the infrared camera detects the element of depth in this network. Kinect comprises of a color video camera with a unique microphone configuration.

Different types of floor-sensors like capacitive, resistant and coax cable (used just as a delay line) are used for acquiring different types of gait data. Following data gathering, noise removal and background modelling are performed using filters (Gaussian filters, mean filters, and so on) and background/foreground subtraction methods (frame difference, GMM, and so on).

2.2. Feature Extraction

Following the pre-processing of obtained data, the next step is to extract some characteristics from the acquired gait data, which is noise-free and segregated using background subtraction algorithms. A variety of methods are available for extracting features categorized into model-based and model-free techniques. The goal of model-based feature representation is to model the human body and extract features from it. Model-based feature representation is size and view independent. However, it is dependent on the video quality. On the other hand, the goal of model-free feature representation is to process a human silhouette's whole motion or shape. There are two major advantages to feature representation without using a model. For starters, it is unaffected by video quality. As a result, it can be used far away from the human object. Second, it is less expensive to compute than model-based feature representation. As a result, model-free feature representation is becoming increasingly common.

2.3. Feature Selection/ Reduction

Feature reduction is a type of dimension reduction strategy that can help improve the gait biometric system's effectiveness and efficiency by removing some unimportant features. As stated below in table 1, there are a variety of approaches for reducing features.

Table 1. Survey of some state-of-the-art feature selection algorithm.

Feature sel.	Algorithm
CSA+DATER [2]	The pre-processing algorithm CSA, which is matrix-based, reduces noise and keeps the best information, while DATER is utilized to improve classification abilities.
RF [3]	A backward feature elimination search technique looks across the subspaces while the Random Forest algorithm scores the features.
PCA [4]	On wavelet features, PCA performs feature separation. Dimensionality features can be efficiently separated for classifiers to use as features up to the 6th principal component.
WPCA [5]	In WPCA, eigenvectors were separated using the square root of the eigenvalues. The positive impact of eigenvectors that convey specific knowledge increases, whereas the negative impact of eigenvectors with higher eigenvalues declines.
BD [6]	Using the principles of binomial distribution, BD applied to random characteristics retrieves some successful features (mean and variance). A simple algorithm is then constructed to select/extract the best features.
FEcS [7]	Using the FEcS technique, the fused feature vector (FV) produces entropy and skewness vectors and selects the optimal subsets of features.
Firefly algorithm and Skewness based approach [8]	On fused vector, the Firefly algorithm (a meta-heuristic technique) was implemented. Using fused vectors, a skewness-based technique finds the best features.
LDA [9]	LDA reduces information loss ($m < n$ where n = dimension of feature data) by using an orthogonal transformation technique to provide m-dimensional feature data that are unrelated and utilized for classification. The algorithm chooses the direction in which substantial data variation occurs. As the orthogonal transformations' axis.

2.4. Classification

It's the final step in the gait recognition process. This stage's accuracy is mainly reliant on the preceding levels. Machine learning algorithms, which can be supervised, unsupervised, or reinforcement learning algorithms, are used to classify data. For human gait classification, classifiers such as Nearest-Neighbour (NN), KNN (k-nearest neighbour), Multilayer Perceptron, Decision Tree, SVM, Nave Bayes, Random Forest, DCNN, Regression Techniques, Random Tree, BayesNet, QDA (quadratic discriminant analysis), Neuro-Fuzzy classifier, CART (classification and regression tree), PNN (probabilistic neural network), FIS (fuzzy interference system) were used in the recent research work.

3. Proposed Work

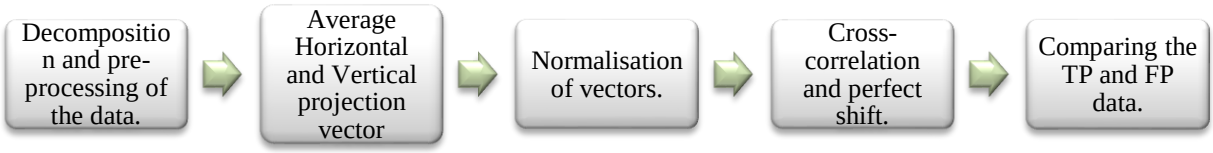


Figure 2. Framework of the Proposed Approach.

- As shown in Fig.2, the first-person (FP) video was split into frames, and certain pre-processing operations were carried out to remove the video's empty and turning regions.
- Each image frame is divided into its R, G, and B colour components, with horizontal and vertical projection vectors calculated for each component. After calculating three separate horizontal and vertical projection vectors for each of the three components (Red, Green, and Blue), the average projection vectors are calculated using the sample mean.
- Normalizing both the horizontal and vertical projection vectors after calculating the average horizontal and vertical projection vectors yields normalised horizontal and normalised vertical projection vectors.
- The peak value of cross-correlation is calculated for each consecutive normalised horizontal projection vector value, and the position or perfect shift value at which the peak value of cross-correlation occurs is investigated further. Similarly, for each subsequent normalised vertical projection vector value, a cross-correlation value is calculated, and the peak value of cross-correlation is acquired to study the position or perfect shift value at which the peak value of cross-correlation occurs.
- The number of local maxima peaks in an oscillation is compared to the number of steps (or gait cycles) a person took while running using all of the horizontal perfect shift values acquired. Similarly, the number of local maxima peaks in an oscillation is compared to the number of steps count (or gait cycles) a person has by plotting all of the vertical perfect shift values acquired against time.

4. Dataset and Description

The implementation of a novel model on a database is critical to its development and implementation. Testing on a database of appropriate size and variant factors is required to obtain successful results for the model. In the last several years, a number of datasets in the field of human gait identification have been made publicly available for research purposes. Every other database has been shown to be effective in terms of a variety of parameters, including baggage carrying conditions, garment variations, covariates, changes in location or view in footwear, indoor or outdoor circumstances, and so on. Table 2 shows a descriptive survey-based study of certain publicly available human gait datasets.

Table 2. Available Public Datasets used for Human Gait Analysis.

Data	sub./seq.	cam/view/style
UMD	55 sub.	2 cam/ 4 view
CASIA A [10]	20 sub/ 240 seq.	3 views
CASIA B [11]	124 sub.	11 views/ 4 style
CASIA C [12]	153 sub./ 1530 seq.	1 cam
floor-sensor-based [13]	15 sub.	Different footwears

Georgia-Tech [14]	18 sub./ 20 seq.; 15 sub./ 268 seq.	multiple view
TUM-IITKGP [15]	35 sub./ 840 seq.	1 cam/ 6 style
HUMANID [16]	74 sub./ 452 seq.	2 cam/ 12 style
Motion-Recording-Sensor-Based Dataset [17]	50 sub./ almost 100 seq.	2 styles
MIT [18]	24 sub./ 194 seq.	1 cam/ 1 view
Human Activities and Postural Transitions Dataset	30 sub.	style: 3 static postures/ 3 dynamic motions
HID UMD 1 and 2	25 sub./ 100 seq.; 55 sub./ 222 seq.	2 views
Ground Reaction forces	62 sub.	Data taken as body mass
TUM-GAID	305 sub./ 3370 seq.	Multimodal gait recognition, RGB-D sensor with a 4-channel microphone array
OU-ISIR A, B and D	34 sub./ 408 seq.; 68 sub./ 1350 seq.; 185 sub./ 370 seq.	25 views/ 9 speed variations/ 32 style
UCSD	6 sub./ 42 seq.	1 cam (stationary camera mounted at an elevation)
UWB Impulse Radar Prototype	-	The individual marched/ walked with either one-arm or two-arm swings.
SOTON large dataset	115 sub./ 2128 seq.	multiple view
First wave-radar-based dataset	49 sub.	Signal reflected from the person’s torso, legs, and arms.
SOTON small dataset	12 sub.	3-4 view
SOTON temporal dataset	25 sub/ 2280 seq.	multi-view

4.1 Dataset Generation

The purpose of recording this database, which contains two categories of data, is to serve both research and practical needs.

Two sets of data are collected using two separate phone cameras on a database of 23 participants (14 males and 9 females). One camera is mounted on a tripod stand for collecting third-person running data, and the tripod stand is positioned at a perpendicular bisector drawn at line segment AB. The line segment AB connects the points A and B on the running track. The side perspective of a person's running manner is captured by the TP data. The other camera, which is used to collect first-person (FP) data, is attached to the volunteer's body with the use of a waist belt. The FP data, as well as the timestamp information needed to sync it with the TP data, is recorded with the use of a phone camera. The FP data is then utilised to analyse the camera's motion in order to estimate body movements.

Only one volunteer's running data is collected at a time, and two cameras record two sorts of data. A person was instructed to simply run with his or her natural running style while wearing the FP camera on his or her belt, while the TP camera on the tripod stand was turned on at around the same time. Data was collected from participants of both genders in the following age groups: below 5 years, 5-14, 15-24, 25-34, 35-49 and 50-65 years of age.

4.2 Precautions Taken

- The number of steps in both the TP and FP camera recorded data should be the same to infer the two forms of data from one another.

- While recording the data, neither of the cameras should have direct sunlight in their view. In the frame of view of the cameras, there should be no sharp shadows thrown by the subject or volunteer. Data should be recorded after sunset and before nightfall, or on cloudy days should be recorded.
- A significant number of samples from various age groups, with enough examples in each category, should be available to raise doubts about the statistical significance. Any category may only be thoroughly examined if a vast number of instances are included.

5. Discussion

The running database simulations are compared to the database simulations that include the walking patterns of the same 23 participants.

Table 3. The steps count and positive peaks count from the obtained horizontal perfect shift plot for the volunteers walking on straight line AB: classified as underweight, normal, overweight, and obese are used to compare the TP and FP video results.

AGE GROUP (relation B/W TP and FP data)	GENDER - AGE	STEPS COUNT (TP)	PEAK COUNT (FP-HORIZO-NTAL)	BMI (Body Mass Index)	Health
Below 6 (exactly double)	M – 5 years	38	76	16.5	Normal
6-14 (almost triple)	F – 7 years	31	70	13.2	Underweight
	F – 10 years	25	73	16.8	Normal
	M – 12 years	25	74	15.3	Normal
	M – 12 years	23	73	16.8	Normal
	M – 13 years	24	70	16.4	Normal
	M – 13 years	24	60	15.1	Underweight
	M – 14 years	22	54	15.4	Underweight
15-30 (almost double)	M – 15 years	22	45	19.2	Normal
	F – 15 years	27	56	21.3	Normal
	F – 16 years	27	55	20.6	Normal
	M – 21 years	26	53	27.1	Overweight
	M – 22 years	23	59	30.8	Obese
	F – 25 years	24	50	23.4	Normal
	F – 25 years	25	51	22.4	Normal
31-49 (almost same)	F – 32 years	24	25	23.9	Normal
	M – 34 years	24	27	27.5	Overweight

	M – 42 years	21	28	30	Obese
	F – 44 years	26	29	31.7	Obese
	M – 47 years	24	27	24.9	Normal
	M – 48 years	23	30	30.9	Obese
50-65 (almost double)	F – 54 years	29	58	27.6	Overweight
	M – 55 years	24	42	24	Normal

Table 4. The steps count and positive peaks count from the obtained horizontal perfect shift plot for the volunteers running on straight line AB: classified as underweight, normal, overweight, and obese are used to compare the TP and FP video results.

AGE GROUP (relation B/W TP and FP data)	GENDER - AGE	STEPS COUNT (TP)	PEAK COUNT (FP-HORIZO-NTAL)	BMI (Body Mass Index)	Health
Below 6 (almost double)	M – 5 years	23	40	16.5	Normal
6-14 (almost double)	F – 7 years	15	26	13.2	Underweight
	F – 10 years	18	34	16.8	Normal
	M – 12 years	18	35	15.3	Normal
	M – 12 years	15	32	16.8	Normal
	M – 13 years	22	44	16.4	Normal
	M – 13 years	22	44	15.1	Underweight
	M – 14 years	20	38	15.4	Underweight
15-30 (almost double)	M – 15 years	14	34	19.2	Normal
	F – 15 years	18	26	21.3	Normal
	F – 16 years	19	28	20.6	Normal
	M – 21 years	16	24	27.1	Overweight
	M – 22 years	15	30	30.8	Obese
	F – 25 years	16	25	23.4	Normal
	F – 25 years	16	25	22.4	Normal
31-49 (almost double)	F – 32 years	21	39	23.9	Normal

	M – 34 years	17	32	27.5	Overweight
	M – 42 years	20	39	30	Obese
	F – 44 years	25	48	31.7	Obese
	M – 47 years	18	34	24.9	Normal
	M – 48 years	19	38	30.9	Obese
50-65 (almost double)	F – 54 years	26	54	27.6	Overweight
	M – 55 years	20	39	24	Normal

Some discussions can be conducted using the tabular data provided in table 3 and table 4.

The volunteers ranged from overweight to obese, and when their walking data was assessed, the camera motion was higher for obese people than for underweight people. In table 3, the number of positive peaks in an oscillation obtained from the FP data plot was exactly double that of the number of steps obtained from the TP video data in the case of a 5-year-old healthy male, and the data obtained from the FP videos is almost three times that obtained from the TP videos in the next age group, which is 6-14 years old. In terms of steps count and peak, there is a difference for non-healthy volunteers, and the FP data for underweight individuals is less than three times that of the TP data, indicating less camera movement. The data acquired from FP videos for the 15-30 range is twice that obtained from FP videos, indicating that camera movement was higher in fat and overweight volunteers than in healthy subjects. Peak counts from FP videos are substantially equal to steps counts from TP videos exhibiting less camera movement while walking for volunteers aged 31 to 49.

Whereas in table 4, the number of positive peaks in an oscillation obtained from the FP data plot was almost double that of the number of steps obtained from the TP video data in the case of a 5-year-old healthy male, also in the case of rest of the age groups: a uniformity can be seen in terms of the camera motion as for every age group the data obtained from the FP videos is almost twice that the data obtained from the TP videos. While running, there is a uniform shakiness or movement of the volunteers' bodies, which causes the camera to move. This is due to the fact that when compared to walking, running involves more vigorous body motions.

6. Conclusion and Future Direction

A small-scale database is proposed, which contains data from 23 volunteers and records their running motion (14 males and 9 females). For study, two types of data are collected: FP (first person) and TP (third person). The FP data is used to perform proper camera motion analysis, whereas the TP data is simply utilised to track gate cycles or the number of steps count for various subjects. According to the simulated experiments the running motion of the volunteers leads to more camera movements caused by vigorous bodily movements, the camera movements for every volunteer belonging to different age groups are uniformly double than the number of step counts estimated from the TP videos data. In the case of running, the camera motion is irrespective of the individual's BMI and health, and there is more shakiness or movement of the camera in running than in walking. When walking, an underweight volunteer of any age group has fewer body movements and thus produces fewer camera movements than a healthy volunteer, whereas an obese and overweight volunteer of any age group has more body movements and thus produces more camera motion than a healthy participant, according to the findings. Depending on the age group and gender, the camera moves in different ways. Future analysis could infer FP and TP data from one another, such as predicting joint angles and determining the cross-correlation between the two databases for both running and walking motions. Individual recognition can be accomplished by analysing their stride, which entails deducing the FP and TP data from one another.

References

[1] Kumar M, Singh N, Kumar R, Goel S, and Kumar K 2021 Gait recognition based on vision systems: systematic survey *J. of Vis. Comm. and Imag. Repre.* **vol 75** pp. 103052-64.

[2] Dong X, Shuicheng Y, Dacheng T, Lei Z, Xuelong L and Hong-Jiang 2006 Z Human Gait Recognition with Matrix Representation *IEEE Trans. on Circ. and Sys. for Video Tech.* **vol 16** pp. 896-03.

[3] Dupuis Y, Savatier X and Vasseur P 2013 Feature subset selection applied to model-free gait recognition *Imag. and Vis. Computing* **vol 31** pp. 580-91.

[4] Yuwono M, Su SW, Guo Y, Moulton BD and Nguyen HT 2014 Unsupervised nonparametric method for gait analysis using a waist-worn inertial sensor *Appl. Soft Computing* **vol 14** pp. 72-80.

[5] Uzun-Per M and Gökmen M 2018 Face recognition with Patch-based Local Walsh Transform *Sig. Proc.: Image Comm.* **vol 61** pp. 85-96.

[6] Arshad H, Khan MA, Sharif M, Yasmin M, and Javed MY 2019 Multi-level features fusion and selection for human gait recognition: an optimized framework of Bayesian model and binomial distribution *Int. J. Mach. Learn. & Cyber*. **vol 10** pp. 3601–18.

[7] Arshad H, Khan MA, Sharif M, Yasmin M, Tavares JM, Zhang YD and Satapathy SC 2020 A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition *Expert Systems* pp. 12541-56.

[8] Mehmood A, Khan MA, Sharif M, Muhammad, Khan S, Shaheen M, Saba T, Riaz, Riaz N and Ashraf I 2020 Prosperous Human Gait Recognition: An End-to-End System based on Pre-trained CNN Features Selection *Multi. Tools and App.*

[9] Guo H, Li B, Zhang Y, Zhang Y, Li W, Qiao F, Rong X and Zhou S 2020 Gait Recognition Based on the Feature Extraction of Gabor Filter and Linear Discriminant Analysis and Improved Local Coupled Extreme Learning Machine *Mathe. Prob. in Eng.* **vol 2020** pp. 245-54.

[10] Yu S, Tan D and Tan T A 2006 Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition *18th Int. Conf. on Pattern Recog. (ICPR'06)* Hong Kong China pp. 441-44.

[11] Wang L, Tan T, Ning H and Hu W 2003 Silhouette analysis-based gait recognition for human identification *IEEE Trans. on Pattern Analy. and Mach. Intell.* **vol 25** pp. 1505-18.

[12] Tan D, Huang K, Yu S and Tan T 2006 Efficient Night Gait Recognition Based on Template Matching *18th Int. Conf on Pattern Recog. (ICPR'06)* Hong Kong China pp. 1000-03.

[13] Orr RJ and Abowd GD 2000 The smart floor: a mechanism for natural user identification and tracking *CHI EA '00: CHI '00 Extended Abstracts on Human Factors in Computing Systems*, NY USA pp. 275–76.

[14] Johnson AY and Bobick AF 2001 A Multi-view Method for Gait Recognition Using Static Body Parameters *AVBPA Audio- and Video-Based Biometric Person Authentication* pp. 301-1.

[15] Hofmann M, Sural S and Rigoll G 2011 Gait Recognition in the Presence of Occlusion: A New Dataset and Baseline Algorithms *Proc. of the 19th Int. Conf. in Central Europe on Comp. Graph. Vis. and Comp. Vis.* pp. 99–04.

[16] Phillips PJ, Grother P, Sarkar S, Vega IS and Bowyer K 2002 Baseline Results for the Challenge Problem of Human ID Using Gait Analysis *5th IEEE Int. Conf. on Automatic Face and Gesture Recog. (FGR)* with CD-ROM Washington D.C. USA pp. 137-142.

[17] Gafurov D, Snekkenes E and Bours P 2007 Gait Authentication and Identification Using Wearable Accelerometer Sensor *IEEE Wor. on Aut. Identification Advan. Techn.* pp. 220 – 25.

[18] Collins RT, Gross R and Shi J 2002 Silhouette-based human identification from body shape and gait *Proc. of 5th IEEE Int. Conf. on Automatic Face Gesture Recog*. pp. 366-71.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/355927843

NATURAL LANGUAGE PROCESSING AND SENTIMENT ANALYSIS

Article *in* International Research Journal of Computer Science · November 2021

DOI: 10.26562/irjcs.2021.v0810.001

CITATIONS
0

READS
28

1 author:



Aryaman Sharma
Delhi Technological University
2 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

NATURAL LANGUAGE PROCESSING AND SENTIMENT ANALYSIS View project

All content following this page was uploaded by Aryaman Sharma on 05 November 2021.

The user has requested enhancement of the downloaded file.

International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842
Issue 10, Volume 08 (October 2021) https://www.irjcs.com/archives

NATURAL LANGUAGE PROCESSING AND SENTIMENT ANALYSIS

Aryaman Sharma
Undergraduate student, Computer Engineering
Delhi Technological University
aryamansharma0408@gmail.com

Publication History
Manuscript Reference No: **IRJCS/RS/Vol.08/Issue10/OCCS10080**
Received: 03, October 2021
Accepted: 21, October 2021
Published: 01, November 2021
Citation: Aryaman (2021). NATURAL LANGUAGE PROCESSING AND SENTIMENT ANALYSIS. International Research Journal of Computer Science, X- October Issue, 2021, Pgs 237-242
doi: https://doi.org/10.26562/irjcs.2021.v0810.001
Editor: Dr.A.Arul Lawrence Selvakumar, Chief Editor, IRJCS, AM Publications, India
Copyright: ©2021 This is an open access article distributed under the terms of the Creative Commons Attribution License; which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: Natural Language Processing is one of the branches of Artificial Intelligence that has only recently entered the spotlight. Apple Siri, Amazon Alexa, and, more recently, Google Duplex are just a few of the most well-known instances of NLP in work, with the technology delivering outstanding human-machine interactions. By 2023, there are estimated to be eight billion digital voice assistants in use, owing to their popularity. With such a large-scale use, the data generated from these interactions would also be immense. This untapped goldmine of data can be used to further research and development of Natural Language Processing and can be used in many industries such as healthcare, technology, business etc. Sentiment analysis, with the help of Natural Language Processing can help industries process these huge datasets faster and efficiently and can be used, for example, in the healthcare industry to diagnose patients and develop diagnostic models for detecting chronic disease in its early stages. Web 2.0 has allowed massive user data to be generated which can be tapped to extract valuable information for various reasons that an individual, policy maker(s), organizations and Governments might need.

Keywords: Natural Language Processing, Healthcare, Technology, Artificial Intelligence, Data.

INTRODUCTION

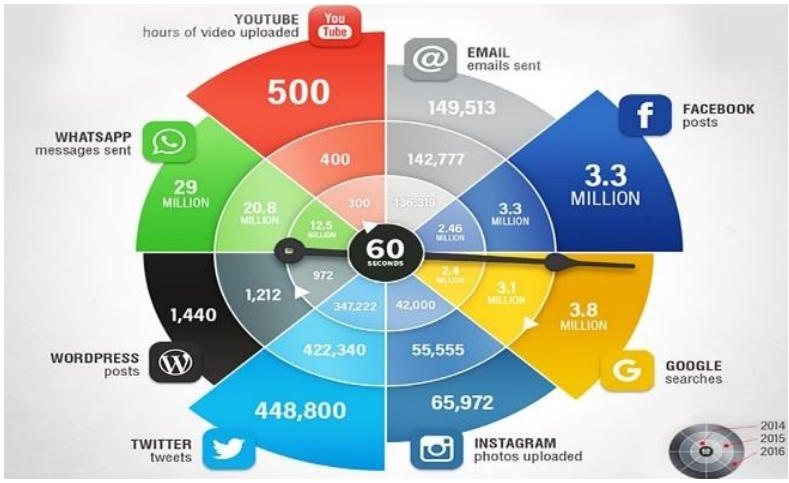
Natural language processing refers to a system's capacity to process sentences written in natural languages such as English rather than specialized artificial computer languages such as C, C++, JAVA, PROLOG, and so on. It is a branch of artificial intelligence that examines data in order to better comprehend natural human language. Natural language processing, or NLP, is a branch of machine learning that focuses on analyzing the nuances of how people communicate with one another so that it may replicate how we speak to and with one another without the constraints of the English language. It is impossible to overestimate the relevance of computer communication in ordinary language. Natural language processing (NLP) is undoubtedly the best example, with intelligent chatbots poised to revolutionize customer service and beyond. While computers have always been extremely useful for abstract tasks such as quantification, real humans have always presented a problematic interface. Despite the fact that computing technologies enable extremely rapid and precise communication channels, machines have never been satisfactory at understanding how and why we interact. NLP is focused to deciphering the cooperation between computers and machines through the use of language.

Today's consumers generate a vast amount of data through digital means. Because the data is created across multiple platforms, it is tough for businesses to keep track of it and respond effectively. Natural language processing (NLP) solutions give businesses a reliable way to keep track of a large amount of user-generated material across several digital channels. Companies can actively engage customers by replying to their queries or complaints using NLP techniques. NLP tools also assist businesses in installing various content filters, such as filtering undesired information or implementing spam filters on websites or social media pages. Businesses utilize NLP models to keep the quality of information on forums up to par. There will be no harmful backlinks or unsolicited advertising as a result of this. It can also auto-filter user-generated information that could be considered as hate speech by detecting “banned” words or vulgarism.NLP has become popular as a result of big data technologies, as this approach can now be used to handle enormous volumes of text data at unprecedented speeds using cloud/distributed computing. Machines can now interpret more language-based data than humans can, without becoming fatigued and in a consistent and unbiased manner. [1].

© 2014-21, IRJCS-All Rights Reserved

Recent strides and research interest in the field of Natural Language Processing have made this area one of the most interesting, contemporary and important topics for study. The onset of big data and its collection, processing etc. have made it impossible for man to do all those by themselves. To put things in perspective, 1 Exabyte (10^18) of data is created on the internet every day, which is roughly the same as 250 million DVDs worth of data. Humankind generates the same amount of data in two days as it did from the origin of civilization until 2003. All these factors and points prove that NLP is going to guide and help us shape the future. The ongoing covid-19 pandemic has brought many challenges in various industries such as Healthcare, Technology, Retail etc. With these challenges comes opportunities to implement newer technologies to combat it. We shall explore these cases with special emphasis in the healthcare industry as a case study in this review research paper.

The objective of this study will be to explore how Natural Language Processing started being used in Sentiment Analysis. From simple chatbots to clinical diagnosis of patients suffering from psychological disorders, we shall see the flexibility of Natural Language Processing used in gauging emotions and sentiments. We shall also try to gauge the future and relevance of this topic while reviewing the work done towards it. We shall also see how mountains of user generated data can be sifted through easier via NLP to understand user review about a product or service by analyzing the sentiments in order to improve the same.



Reference: "What happens online in 60 seconds? | Smart Insights," *Smart Insights*, Mar. 30, 2017. https://www.smartinsights.com/internet-marketing-statistics/happens-online-60-seconds/.

INCEPTION

Syntactic Structures, written by Noam Chomsky, was published in 1957. He revolutionized existing linguistic principles in it, stating that the sentence structure must be changed for a computer to grasp a language. Chomsky developed a type of grammar known as Phase-Structure Grammar to achieve this objective, which painstakingly transformed natural English phrases into a structure that computers could understand. The overarching goal was to construct an artificial intelligence (AI) machine capable of thinking and communicating like a human brain.[2]

Natural Language Processing was given a fresh lease on life when the idea and need for Artificial Intelligence surfaced in the 1960s. LUNAR was founded in 1978 by W.A. Woods with the objective of studying, comparing, and interpreting chemical data on lunar rock and soil composition that had accumulated as a result of the Apollo moon landings, as well as answering the associated question. In the 1980s, computational grammar became a hot topic of study, and it was linked to the science of reasoning for meaning and taking into consideration the user's preferences and intents. In the 1990s, the rate of expansion of NLP quickened with the availability of parsers, grammars, tools, and practical materials related to NLP.[3].

In the 1980s, computational grammar became a hot topic of study, and it was linked to the science of reasoning for meaning and taking into consideration the user's preferences and intents. In the 1990s, the rate of expansion of NLP/MT quickened. Parsers, grammars, tools, and practical resources related to NLP/MT become available as a result of the parsers.

The history of natural language processing would be incomplete without including ELIZA, a chatbot programbuilt at MIT's Artificial Intelligence Laboratory between 1964 and1966.Joseph Weizenbaum was the one who came up with the idea. The analysis of key words is the first step in Eliza's input processing. Key words are significant, and scripts provide rankings to them. If key words are present, they are sorted by rank, and the input sentence is processed according to the rule associated with the key word with the highest rank.

© 2014-21, IRJCS-All Rights Reserved

Eliza uses a transformation rule that consists of two parts: breakdown and reassembly, after recognizing important words. The decomposition rule divides the input into chunks based on key words and applies the appropriate transformations. Then, when reassembling the sentence, the decomposition rule creates reassembly rules to follow. This rule rearranges the pieces of the input sentence formed by the decomposition rule in a suitable format. Overall, Eliza employs a set of rules to generate a response from user input.[4]. Because sentiment analysis is a relatively young topic of study, various terms have been invented to characterize it. In sentiment analysis, the words opinion mining, subjectivity analysis, review mining, and assessment extraction are all interchangeable. The purpose of subjectivity analysis, also known as subjectivity categorization, is to figure out if a statement or document expresses the author's thoughts or sentiments or only delivers facts. The majority of articles that use the phrase "sentiment analysis" focus on how to classify evaluations favorably or negatively.

MORDERN DAY USE AND APPLICATIONS

Nowadays, everyone wants the machine to communicate, and Natural Language Processing is the only means for a computer to do so (NLP). Take, for example, Amazon's Alexa, a talking assistant. It receives an inquiry through the medium of voice and can respond through the same media, i.e., voice. It can be used to ask questions, search for information, play music, and even book a cab. The technique for Natural Language Processing (NLP) in Alexa is demonstrated in this basic figure. Alexa is only one example; these talking computers, also known as Chatbots, can handle even the most complex interactions and operations connected to streamlined business using simply NLP. Previously, chatbots were only used for customer interaction and had limited conversational capabilities due to their rule-based nature. However, with the advent of Natural Language Processing and its integration with Machine Learning and Deep Learning, chatbots can now handle a wide range of tasks, including Human Resources and Health. [5][6]

Understanding customers has long been a priority for companies. However, all too frequently, these businesses fall into the trap of making product selections based on archaic marketing strategies like the consumer mood index. The consumer sentiment index provides useful information about macroeconomic conditions, but it does not assist in addressing consumer demand. Furthermore, conventional wisdom holds that offering a diverse range of products both online and in-store is the best strategy for improving the chances that customers will find ones they enjoy. Unfortunately, this strategy does not always work because many customers are perplexed by all of the options. To obtain information about customer demand, sentiment analysis uses natural language processing (NLP), text mining, and data mining capabilities. NLP is a branch of machine learning and artificial intelligence that enables machines to read and understand human language. It develops a means to evaluate millions of postings that are uploaded every day to social media sites like Facebook and Twitter because it is done by machines. Sentiment analysis examines the text for positive, negative, or neutral sentiment. It can also consider the meaning of the words as well as the context in which they were spoken.[7]

Customers are increasingly expressing their needs, ideas, preferences, and complaints online, which is why sentiment analysis is becoming popular among retailers and technology companies. Businesses can use sentiment analysis and natural language processing to learn what their consumers are saying and how they feel about the products they're selling. Retailers and IT businesses may now use social media to understand their consumers' mood in real time, learning how they feel about things on store shelves, store layouts, and ads. Furthermore, sentiment analysis may inform organizations about how strongly customers feel about a product and which aspects are responsible for those feelings. Technology companies may be able to turn their fortunes around in the future by incorporating sentiment analysis into their product releases. With real-time data, the merchandising team could make adjustments to rekindle public interest in the company's products.

FUTURE AND APPLICATIONS IN WEB 2.0

Future opinion-mining systems will necessitate a larger and more diverse knowledge base of commonplace and commonsense data. More comprehensive knowledge must be linked to reasoning techniques impacted more deeply by human psychology and thought. Researchers will be able to better understand natural language opinions as a result of this, and the gap between (unstructured) multimodal data and (structured) machine-processable data will be bridged. Combining scientific theories of emotion with the practical engineering goals of analyzing sentiments in natural language text will result in more bio inspired approaches to the design of intelligent opinion-mining systems capable of dealing with semantic knowledge, making analogies, learning new affective knowledge, and detecting, perceiving, and "feeling" emotions. The capacity to tap into public emotion on social media is becoming more widely recognized as a valuable tool for market research, consumer segmentation, and stock price forecasting in strategic marketing planning and maneuvering. The healthy growth of the big data framework has fueled this evolution of technology adoption, causing applications based on Sentiment Analysis (SA) in big data to become commonplace for enterprises.

© 2014-21, IRJCS-All Rights Reserved

The term "Web 2.0" is commonly used to denote a new way of thinking about data and data interactions on the internet. This evolution affects both technologies and practices. This progress may be seen in the way users engage with one another by acting directly on the website's content (for example, by adding tags to ease future search/retrieval/classification). As a result, Web 2.0 exemplifies community, dynamicity, and web interactions. As a result, websites are more reliant on tools such as wikis and blogs. [8][9]

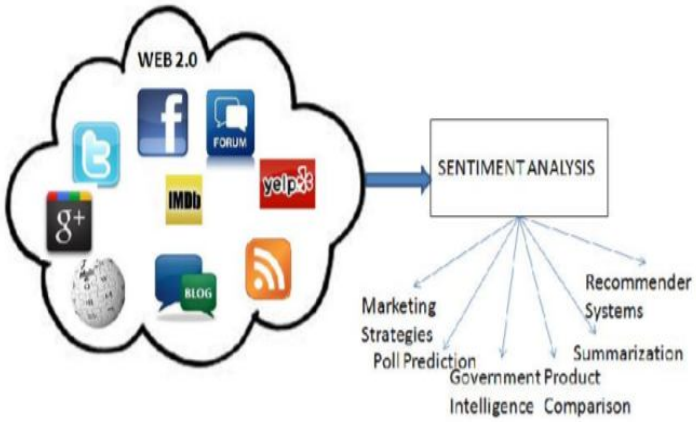
The area of sentiment analysis study has been rapidly expanding as a result of the huge and diverse data given by Web 2.0 applications. The data used in sentiment analysis comes from blogs, review sites, forums, microblogging sites, wikis, and social media sites. A review site is a website where users can post unfavorable comments about people, businesses, products, and services. The majority of sentiment analysis is done on movie and product review sites. Because the purpose of a review is to evaluate a single object, it is a single domain problem. Sentiment research on review sites can help manufacturers and future purchasers of a product. Manufacturers can use reviews to determine how well a product is received. They can deduce which characteristics the reviewers liked and disliked.[6]. The process of extracting meaningful conclusions is a tricky and long process. Data collection is the first step. One of the most crucial elements in the sentiment analysis process is this. Everything will depend on the quality of the data gathered and how it has been annotated or labeled from here on out. Live APIs for social media can be used to upload data. A news API can help you gather data from a variety of news sources, whereas a Facebook API can provide you access to all of the platform's publicly available data. Open-source repositories such as Kaggle, Amazon reviews, and Yelp can also be used. If you have data that you already have from a CRM application, you can manually upload that onto the sentiment analysis API as a .csv file. Customer relationship management (CRM) is a software program that helps you to keep track of all of your company's customer and prospect contacts and interactions. The second step is to analyze the data. The type of information in the data will determine how it is processed: text, image, video, or audio.

Audio transcription - To ensure that any video or audio file (e.g., podcast) in the data is not disregarded, the audio from the video data is transcribed using speech to text software.
Caption overlay - Any subtitles that appear in the video are evaluated for any appearing entities, aspects, or subjects that you have flagged as important.
Image overlay - Using OCR, the software can recognize and collect any images in a video or text data (optical character recognition)

Text extraction- During the sentiment analysis process, all text is recognized and extracted in the same way. Emojis and hashtags are also included, as they are an important aspect of social media sentiment research. Emojis must never be excluded from data processing because this could result in false positives or negatives. The third step would be to analyze the dataThere are several subtasks that need to be done for this stage of the sentiment analysis process. The dataset that will be used to train the model must be pre-processed and manually labeled before it can be used. By comparing the correctly categorized data with the wrongly classified data, this labelled data will be utilized to train the model. This will aid in the development of a brand's bespoke model. Custom Tags - During this stage of the process, custom tags for aspects and themes such as brand mentions, product names, and so on will be created for the data. Once the model has been trained, it will automatically partition text depending on these specially constructed tags. Topic Classification - A text's theme is assigned by the topic classifier. This sentence, for example, will be labeled with the topic "clothing" since "the dresses were wonderful, and I discovered some pretty excellent scarves as well." Sentiment Analysis - At this point, the platform isolates each aspect and theme, which is subsequently analyzed for sentiment. The range of sentiment scores is from -1 to +1. Zero can be used to describe a neutral statement.

This polarity assignment is significant because, while the platform awards individual scores to different characteristics such as convenience, speed, cleanliness, functionality, drinks, ambience, and so on, it is the aggregate score that is used to determine the audience's sentiment toward the brand. So, if three of the seven elements are rated poorly (-.65) and four are rated positively (+.5), the sentiment analyzer will assign an average score to the brand's overall sentiment. The fourth and final step would be to visualize the data. The insights are immediately translated into actionable reports in the form of graphs and charts when all of the steps in the sentiment analysis process have been completed. These reports can then be distributed among teams. These visual reports are extremely significant since they allow the organization to see granular, aspect-based outcomes. When the organization get an average score for their brand or service, they can use the sentiment analysis dashboard to filter the data to see which features received high scores and which received low scores. This will help them determine which areas require more attention than others. As a result, one gains actionable information at this point of the sentiment analysis steps, which they can utilize to determine the best course of action for a companies or service's growth ambitions.[10][11][12]

© 2014-21, IRJCS-All Rights Reserved

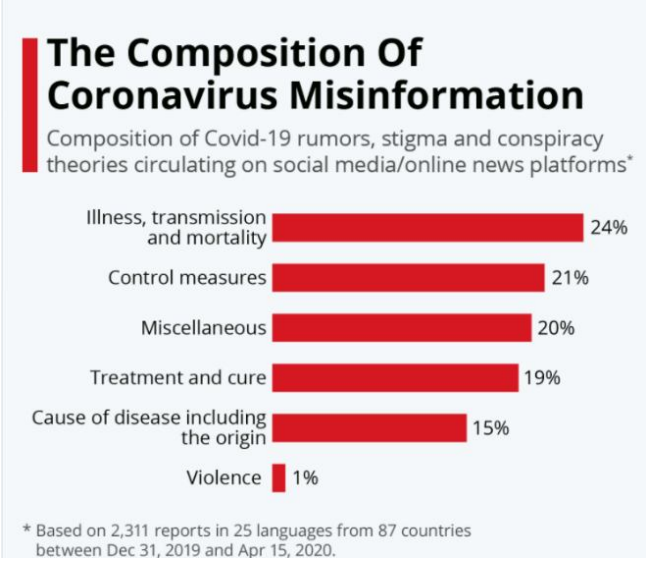


Reference: A. Kumar and T. M. Sebastian, "Sentiment analysis: A perspective on its past, present and future," *Int. j. intell. syst. appl.*, vol. 4, no. 10, pp. 1–14, 2012.

CASE STUDY: COVID -19 AND NATURAL LANGUAGE PROCESSING

The COVID-19 pandemic has had a tremendous impact on society, both as a result of COVID-19's severe health impacts and as a result of public health measures taken to slow its spread. Many of these issues stem from information needs, and attempts to fulfill these needs have resulted in an information overload for both scholars and the general public. Many of the information demands made urgent by the COVID-19 epidemic can be addressed using natural language processing (NLP), a branch of artificial intelligence that translates human language.

The elderly, those with impairments, and those with numerous comorbidities are among the susceptible categories who have been identified as being at a greater risk of getting more severe and life-threatening COVID-19 infections. The COVID-19 pandemic led healthcare providers and systems around the world to curtail or limit less-urgent treatments, such as rehabilitation for persons with acute and chronic diseases and disorders. For some patients, the treatment delay is annoying but not life-threatening. A delay or interruption in treatment can have a substantial impact on recovery and efficacy for other patients. Deferring rehabilitation therapy is undesirable since it leads to poorer physical and psychological consequences for patients, as well as increasing the strain on the healthcare system in the future to deal with the mounting backlog. Rehabilitation has become increasingly important during the COVID-19 pandemic. Rehabilitation is needed to meet the needs of people with acute and chronic illnesses, as well as to help people recover from severe COVID-19 infections that require long-term intensive care and respiratory support. Rehabilitation for post-COVID patients has been demonstrated to be costly to healthcare systems, with the average cost of post-COVID patients' rehabilitation services being nearly twice that of non-COVID patients' rehabilitation services.



Reference: "The Composition of Coronavirus Misinformation," *Statista Infographics*. https://www.statista.com/chart/22527/composition-of-covid-19-misinformation/.

© 2014-21, IRJCS-All Rights Reserved

Artificial intelligence (AI) and machine learning (ML) approaches can be used to support healthcare professionals and healthcare delivery at this moment, when healthcare resources are being stressed because to the epidemic. In-person meetings with doctors, occupational therapists, and physiotherapists can be supplemented with these AI/ML technologies. A case-study summary of our work on an AI/ML and Natural Language Processing (NLP) system for a telephone-based Rehabilitation Advice Line will also be presented as an example of such a system. These technologies can also provide continuity of care when in-person appointments are too risky due to subsequent waves of the COVID-19 epidemic. Furthermore, beyond the pandemic's urgent needs, the implementation of these devices will continue to be beneficial for providing care to isolated and rural populations.[13] A misinformation epidemic accompanied the COVID-19 health disaster. Due to the lack of knowledge about the virus's nature and origins, conspiracy theories arose, which were widely disseminated on YouTube and other online social media platforms. Efforts to detect and filter relevant falsehoods were expedited by YouTube, but the results were mixed. [14]. People have expressed a variety of emotions on social media sites such as Twitter throughout the surge of COVID-19 instances and tougher lockdowns. During COVID-19, social media played a crucial role in motivating academics to use NLP and machine learning technologies to analyze data. Deep learning sentiment analysis revealed that World Health Organization Tweets failed to provide public guidance. Sentiment analysis was used to investigate the impact of a nationwide lockdown in India due to the COVID-19 outbreak, and it was discovered that people in India viewed the fight against COVID19 positively, with the majority agreeing with the government's decision to announce the lockdown in order to flatten the curve. The sheer number of tweets enable researchers and policy makers to observe the impact and result of steps taken to curb the virus. [15]

REFERENCES

1. "The basics of NLP and real time sentiment analysis with open-source tools," *www.linkedin.com*. https://www.linkedin.com/pulse/basics-nlp-real-time-sentiment-analysis-open-source-tools-%C3%B6zg%C3%BCr-gen%C3%A7 (accessed Sep. 23, 2021).
2. "NLP - overview," *cs.stanford.edu*. https://cs.stanford.edu/people/eroberts/courses/soco/projects/2004-05/nlp/overview_history.html (accessed Sep. 23, 2021).
3. Researchgate.net. [Online]. Available: https://www.researchgate.net/publication/247926251_The_Lunar_Science_Natural_Language_Information_System_Final_Report. [Accessed: 23-Sep-2021].
4. Researchgate.net. [Online]. Available: https://www.researchgate.net/publication/221152425_A_Multimodal_Eliza_Using_Natural_Language_Processing_and_Emotion_Recognition. [Accessed: 25-Sep-2021].
5. B. Lutkevich and E. Burns, "What is natural language processing? An introduction to NLP," Techtarget.com, 02-Mar-2021. [Online]. Available: https://searchenterpriseai.techtarget.com/definition/natural-language-processing-NLP. [Accessed: 27-Sep-2021].
6. X. Fang and J. Zhan, "Sentiment analysis using product review data," *Journal of Big Data*, vol. 2, no. 1, Jun. 2015, Doi: 10.1186/s40537-015-0015-2.
7. "(PDF) Sentiment analysis using product review data," ResearchGate. https://www.researchgate.net/publication/316666447_Sentiment_analysis_using_product_review_data.
8. P. Durville and F. Gandon, "Filling the Gap between Web 2.0 Technologies and Natural Language Processing Pipelines with Semantic Web," IEEE Xplore, Oct. 01, 2009. https://ieeexplore.ieee.org/document/5291527 (accessed Oct. 13, 2021).
9. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?" Proceedings of the ACL-02 conference on Empirical methods in natural language processing - EMNLP '02, 2002, Doi: 10.3115/1118693.1118704.
10. Á. García-Crespo, R. Colomo-Palacios, M. Mencke, and J. M. Gómez-Berbís, "CUSENT: Social Sentiment Analysis Using Semantics for Customer Feedback," Social Web Evolution: Integrating Semantic Applications and Web 2.0 Technologies, 2009. https://www.igi-global.com/chapter/social-web-evolution/29290 (accessed Oct. 14, 2021).
11. J. A. Morente-Molinera, G. Kou, K. Samuylov, R. Ureña, and E. Herrera-Viedma, "Carrying out consensual Group Decision Making processes under social networks using sentiment analysis over comparative expressions," Knowledge-Based Systems, vol. 165, pp. 335–345, Feb. 2019, Doi: 10.1016/j.knosys.2018.12.006.
12. K. Unnamalai, "Sentiment Analysis of Products Using Web," Procedia Engineering, vol. 38, pp. 2257–2262, 2012, Doi: 10.1016/j.proeng.2012.06.272.
13. J. Carriere *et al.*, "Case Report: Utilizing AI and NLP to Assist with Healthcare and Rehabilitation During the COVID-19 Pandemic," *Frontiers in Artificial Intelligence*, vol. 4, Feb. 2021, Doi: 10.3389/frai.2021.613637.
14. H. O.-Y. Li, A. Bailey, D. Huynh, and J. Chan, "YouTube as a source of information on COVID-19: a pandemic of misinformation?" *BMJ Global Health*, vol. 5, no. 5, p. e002604, May 2020, Doi: 10.1136/bmjgh-2020-002604.
15. A.H. Alamoodi et al., "Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review," Expert Systems with Applications, Oct. 2020, Doi: 10.1016/j.eswa.2020.114155.

© 2014-21, IRJCS-All Rights Reserved

View publication stats

Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Neural network approach for faster optical properties predictions for different PCF designs

To cite this article: Hardik Kumar *et al* 2021 *J. Phys.: Conf. Ser.* **2070** 012001

View the article online for updates and enhancements.

You may also like

- Collective Study of Polar Crown Filaments in the Past Four Solar Cycles
Yan Xu, Werner Pötzi, Hewei Zhang et al.
- Dispersion tailoring of photonic crystal fibers for flat-top, coherent, and broadband supercontinuum generation
Seyed Ali Seyed Hashemi and Mina Noori
- Migration of Solar Polar Crown Filaments in the Past 100 Years
Yan Xu, Dipankar Banerjee, Subhamoy Chatterjee et al.



This content was downloaded from IP address 14.139.251.100 on 01/12/2021 at 05:28

Neural network approach for faster optical properties predictions for different PCF designs

Hardik Kumar[1], **Tanya Jain**[1], **Mritunjay Sharma**[1], **Kamal Kishor**[1]

[1]Department of Applied Physics, Delhi Technological University, Delhi, India
E-mail: *hardikkumar_2k17ep31@dtu.ac.in*

Abstract. Photonic Crystal Fibres (PCFs) are emerging as an alternative to standard fibres for applications in many disciplines like fibre lasers & amplifiers, imaging, spectroscopy and telecommunications. They have superior light guiding properties compared to ordinary Optical Fibres (OFs). This paper illustrates the potential of neural networks to efficiently and accurately compute the optical properties of PCFs including solid-core, hollow-core and multi-core designs. The proposed method takes a range of design parameters and wavelengths as input to predict PCF optical properties like effective index, effective mode area, confinement loss and dispersion desired for optimal specifications. The neural network approach is significantly better in terms of the low computational runtimes (~5 milli-sec) required for predicting the properties against the longer runtimes (~18 sec) required for similar calculations by traditional numerical methods.

Keywords: Photonic Crystal Fibres, Optical Properties, Artificial Neural Network, Numerical Simulation, Computation Time

1. Introduction

Photonic Crystal Fibres (PCFs) are a special class of optical fibres that are characterized by a periodic arrangement of microcapillaries that form the fibre's cladding around a solid or hollow defect core [1]. They have the advantage of being low loss and endlessly single mode along with a high birefringence and low confinement loss [2]. This paper extends the past approach of using Machine Learning (ML) technique to predict the optical properties of a solid core photonic crystal fibre design based on variation in input design parameters like no. of rings, pitch length, diameter of holes and wavelength of light [3]. A simple yet effective neural network model is trained for three different design types of PCFs namely, solid core, hollow core and a six core multi-core PCF to predict their important optical properties. This is a robust model which has the capacity to generalise the results while maintaining high accuracy and negligible compute times.

The designing of photonic crystal fibres can be an exacting task because of the many time and resource draining computations that are required to be done for arriving at the optimum combination of PCF design attributes. The effective mode index, effective area, confinement loss and dispersion values at different combinations of these input parameters like hole diameter and pitch length vary greatly in magnitude with slight changes in dimensions. The combinations with high loss and high dispersion values are sub-optimal and need to be improved. The best possible optimal combination of PCF parameters where light propagates in the fundamental mode with the least loss is obtained by iterative analysis. The numerical methods generally used for this process are Finite Element Method, Finite Difference Time Domain method, Coupled Mode Model (CMM), and Plane-Wave Admittance Transfer (PWAT) [4]. All of these methods have high compute times in the range of tens of hours and also use heavy computer resources [5].

Recently, Machine Learning and Deep Learning techniques are finding applications in solving complex mathematical problems in the field of optics and photonics including plasmonics [6], nano-photonic structures [7], neuromorphic photonics [8], bio-printing [9], metamaterials [10] and optical communications and networks

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

[11]. The deep learning neural network approach is based on identifying complex patterns in the training dataset which are then used to make connections between the given and desired attributes. The bigger and wider the provided training dataset will be, the more efficient and reliable will be the learning pattern and predictions of the model. This way a neural network can make predictions about the new data without being explicitly programmed. This saves precious time that would be otherwise wasted if using traditional algorithms. This paper serves the purpose to highlight the usefulness of deep learning predictions in the designing of photonic crystal fibres.

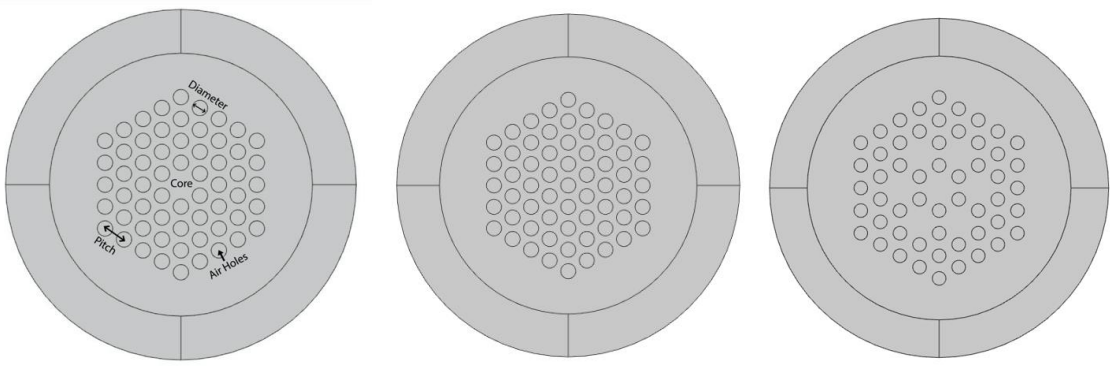
A brief about different PCF designs and considered optical properties are explained in Section 2. The detailed proposed approach and neural network architecture used for modelling is explained in section 3. Prediction results are explained for different types of PCF Design in Section 4. The conclusion of this paper is discussed in the last section, i.e., Section 5.

2. Designs Considered and Optical Properties

For our study we have considered simple Solid Core, Hollow Core and Multi Core PCF designs. The main variation amongst the designs is in the arrangements of air holes around the silica core which results in the wave-guiding structure. A solid core PCF has a solid silica core in the centre of fibre with air holes surrounding it. A hollow core PCF differs from Solid core PCF by having a hole filled with air or a different material in the centre instead of solid silica. A multicore PCF has multiple cores of solid silica. The PCFs have a core of pure silica as in case of solid and multicore PCFs or air as in the case of hollow core with a cladding made up of air-holes separated by pitch length Λ and of diameter d arranged in a hexagonal lattice. We have taken four and five concentric rings designs into consideration. The thickness of fibre is 12 μm and thickness of PML is taken as 4 μm.

2.1 Solid Core PCF:

A Solid Core PCF is a type of photonic crystal fibre which gets its wave guiding characteristics from an order of very small and closely spaced air holes that run through whole length of fibre as shown in Figure 1(i) [12]. The light propagation mechanism in a solid core PCF can be explained via Modified Total Internal Reflection theory [13]. A solid core PCF with a large enough core has a capability of being single mode for a large range of wavelength[2] [14]. It finds its application in the field of telecommunications. If the solid core is big enough then light can be guided through the fibre with high power without the power density causing non-linear effects [15]. This allows distance between repeaters to be greatly increased.



(i) (ii) (iii)

Figure 1. Concentric hexagonal ring designs with 4 rings: (i) Solid core, (ii) Hollow Core and (iii) Multi-core designs taken into consideration.

2.2 Hollow Core PCF:

A hollow-core fibre is an optical fibre in which the light propagates inside a hollow zone of the PCF, such that just a small fraction of the light is guided inside solid fibre(generally made of silica) [16]. Here hollow core PCF design is achieved by having a hollow hole in the centre of the fibre instead of a solid core as seen in Figure 1 (ii). The hollow hole may contain air or some other gaseous material [17]. We have considered a hollow core fibre which have air holes to train our model. Light propagation in hollow core PCFs can be explained only by

using photonic band gap effect [18]. The wavelength range for confinement of these fibres is very small, of the order of few nanometres in the visible or infrared spectrum.

2.3 Multi Core PCF:
Most optical fibres have a single fibre core, which is usually located on the fibre axis. However, there are also specialty fibres containing multiple cores, which may be arranged on a ring around the fibre axis or on some 2D arrangement [19]. Multi-core PCFs guidance principle is that of modified total internal reflection just like solid core PCFs and are designed for high power supercontinuum generation due to their high beam quality output good thermal dissipation properties and large effective mode area [20]. We have taken a design with 6 cores for light guiding into consideration as shown in Figure 1 (iii).

2.4 Desired Optical Properties
For our study we have considered the following optical properties of PCF to compare their values with variation in design parameters.

2.4.1 Effective Refractive Index. It is a figure calculated using the band structure of the surrounding arrangement of holes. It refers to an "effective index" associated with the largest value possible for the propagation constant β for the frequency in the microstructure. Which means at a particular frequency, light with a component β of the wave vector along the axis of the holes greater than a specific β_{MAX} value cannot propagate in the micro-structured part of the fibre [15].

2.4.2 Effective Mode Area. It a measure of the area effectively swept by a waveguide in the transverse dimensions [21]. The effective area is a very crucial quantity. It was put in place to evaluate non-linearities: lesser effective area leads to a higher density of power required for non-linear influences to be noteworthy [22], [23]. The spot-size radius w for a gaussian beam is related to the effective area through the following equation [24]:

$$A_{eff} = \pi w^2 \tag{1}$$

Thus, it is crucial in the context of micro-bending loss , confinement loss [25], splicing loss [26], macro-bending loss [23] and numerical aperture [24]. We used the below formula to calculate the effective mode area:

$$A_{eff} = \frac{\left(\iint_{\Omega}(|Ht|^2\, dxdy\,)\right)^2}{\iint_{\Omega}|Ht|^4\, dxdy} \tag{2}$$

Where **Ω** denotes area enclosed within the computational domain and H_t denotes transverse magnetic field vector.

2.4.3 Dispersion. Dispersion represents all occurrences causing light pulses to broaden while propagating. Dispersion is essentially caused by 4 basic causes [27], [28]: Material dispersion, Inter-modal dispersion, Polarization mode dispersion, and Waveguide dispersion. In optical telecommunications, to keep the bandwidth as large as possible, the optic pulses keep their starting widths. That is because if optic pulses spread, they will overlap further down the path and cannot be differentiated by the receiver. Chromatic dispersion (D) is the dispersion which is the consequence of combined effects of waveguide dispersion and material dispersion. When chromatic dispersion has value greater than zero (the dispersion regime is said to be anomalous), shorter wavelengths are propagated at higher pace than longer wavelengths. In the opposite case, when D is less than zero, the dispersion regime is said to be normal [29]. The following equation has been used to calculate dispersion [30]:

$$D = -\frac{\lambda}{c}\frac{d^2 Re(neff)}{d\,\lambda^2}\ \text{(ps/km.nm)} \tag{3}$$

Where Re stands for the real part of the effective refractive index and c denotes the free-space speed of light.

2.4.4 Confinement loss. Light propagation because of modified total internal reflection between core and micro structured cladding which consist inclusions in a matrix. Irrespective of the approach used to substantiate propagation in PCFs, if guidance is due to a fixed number of holes layer, leakage from the core to the outer matrix of material is unavoidable. These are losses due to the finite extent of the cladding as Confinement losses [31]. The following equation has been used to calculate confinement loss [32]:

$$\alpha_c = 8686 \left(\frac{2\pi}{\lambda}\right) \mathrm{Im}(n_{eff}) \text{ dB/m} \quad (4)$$

Where Im stands for the imaginary part of effective index n_{eff}.

3. Proposed Approach

This section discusses our proposed approach which includes major modules as: Dataset Generation, Pre-processing, Modelling Neural Network. An overview of the approach is presented in the Figure 2.

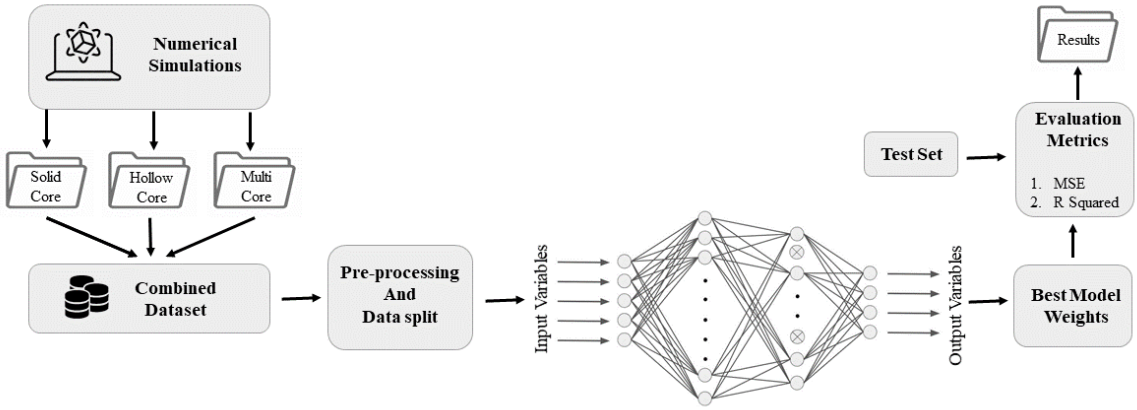


Figure 2. Proposed approach for our task.

3.1 Dataset Generation

The dataset is prepared by running simulations in COMSOL Multiphysics 5.4. For computation of Effective mode index (n_{eff}), Dispersion, Effective mode area (A_{eff}) and Confinement Loss, the input design parameters are varied as follow:

- Pitch [um]: Range (0.8-2.0) with step of 0.2
- Diameter of Holes [um]: Range (0.6-0.9) with step of 0.1
- Wavelength [um]: Range (0.5-1.8) with values divided into 20 steps
- Number of Rings: 4 and 5
- Types of PCF: Solid Core, Multi Core and Hollow Core
- Mesh Size: Finer

The total computed values i.e. around 3359 data instances are combined together for further processing.

3.2 Pre-processing

The various feature values in the dataset have non-uniform ranges of values which is not an ideal condition for neural network modelling. We have used MinMax normalization which rescales the data into [0,1] range values. It improves the robustness of the model accuracy and also reduces the model training time [33]. Categorical features like type of PCF which contain string values are mapped to integers by label encoding.

The dataset is then shuffled to eliminate the bias towards any particular set of values and keep similar distribution in all sets. The data was split into three sets namely, training, validation and test sets in the aspect ratios of 80%, 10% and 10% respectively.

3.3 Modelling Neural Network

In supervised machine learning, Artificial Neural Network (ANN) is considered one of the best techniques because of its ability to learn highly non-linear complicated patterns and features [34]. Neural network architecture consists of a number of layers and each layer can have multiple nodes (or Neurons). Here we have considered a regression problem to predict PCF optical properties as output.

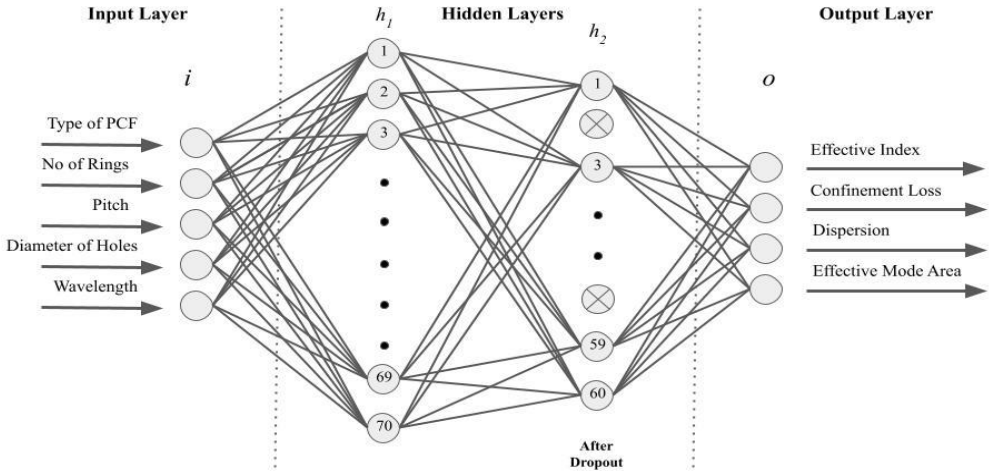


Figure 3. Proposed Neural Network design

The architecture as shown in Figure 3. contains one input layer, two hidden layers and an output layer all connected to each other via nodes. The nodes in input and output layers contain the vectored data. The nodes in hidden layers are basically responsible to convert the input data into corresponding output data by learning the underlying complex rules or patterns.

For training of our neural network, we used Rectified Linear Unit (ReLU) as activation function which is responsible for approximating the non-linear learning function [35]. To optimise the weights of nodes during the training process Adam optimizer was selected as it performs well on big datasets as compared to other optimizers [36]. We set the learning rate parameter to 0.001. Our second and third hidden layers contain 70 and 60 nodes, respectively. Generally, it's a good idea to keep relatively more nodes in early layers as they learn the more complex patterns. In vanilla feed forward neural networks, every node of each layer is interconnected to one another [37]. However, it has been observed often that when training with such architecture, the model suffers overfitting i.e., it predicts well on the training data but suffers relatively more loss on test data. To overcome this a dropout technique is used which basically cuts-off a few connections between nodes randomly [38]. We used dropout in our third layer with value set to 10%. For back-propagation phenomenon [39], mean squared error is calculated between the predicted and actual values and then used iteratively to update the weights of hidden layer nodes [40].

For our experiment we have chosen 2000 epochs because that gives an acceptable stable MSE on both training and validation data. Once these hyperparameters were finalised and the model was trained, we generated the predicted output values for the unseen data of the test set.

4. Results and Discussion

This section presents the performance of our neural network model on validation set and on the unseen test set. The difference between the true and predicted values of target variables are evaluated on the basis of some error metrics used in regression tasks.

Mean squared error (MSE) is one of the metrics to gauge model performance in regression task [40]. For N observation it is defined as the averaged squared difference between the predicted and true values (5), where yi and y are the predicted and true value of data points, respectively. Smaller the value of MSE, smaller the difference between predicted values and actual values and hence model can be considered well trained. Another

metric used here is R Squared (R2), it is a statistical measure of fit which indicates how much variation of a dependent (or target) variable is explained by independent variable(s) in a regression problem [41]. Its value ranges between [0,1], where 1 indicates that the predicted values perfectly fit the data (6).

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \tag{5}$$

$$R^2 = 1 - \frac{Unexplained\ Variation}{Total\ Variation} \tag{6}$$

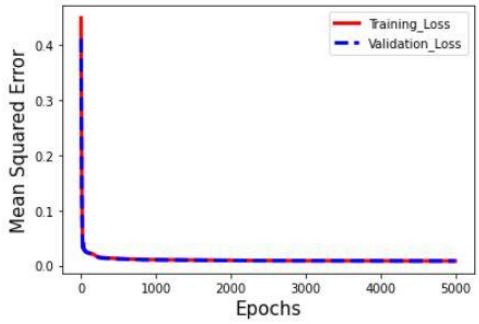


Figure 4. Mean Squared Error Vs No of Epochs

The validation set is available to the model in training process to evaluate the loss on every epoch and update it accordingly. The Figure 4 shows the plot of MSE for both training and validation set for different number of epochs. The negligible difference can be observed between the two and how quickly they converge to a stable low MSE loss.

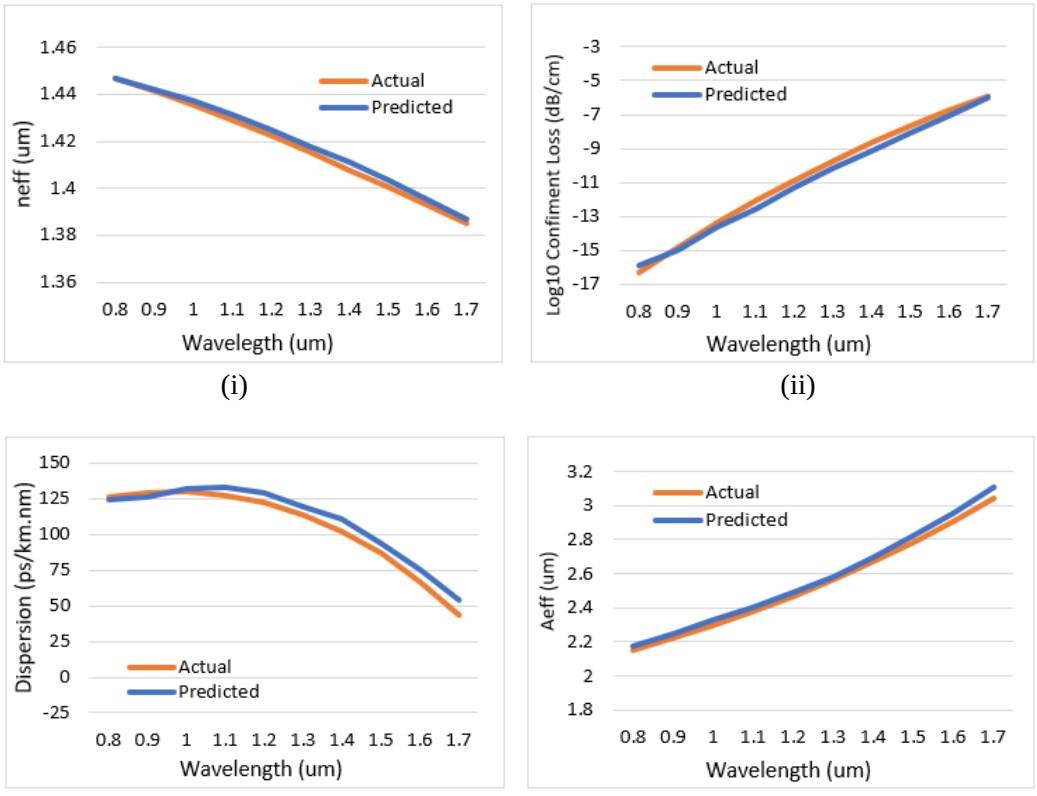


Figure 5. Plot of Actual and Predicted Optical Properties values of Solid Core PCF against different wavelengths

The variation in predicted and actual values of properties of a solid core PCF design is plotted in Figure 5. Generally, with increase in wavelength the effective index value of fundamental mode decreases, the same can be observed in Fig 5(i). The predicted n_{eff} values deviates a bit in the range of 1.3-1.5 um wavelength. Confinement loss values are plotted in Fig 5(ii), Logarithmic conversion is used to obtain a linear relation against wavelength as real values were mostly confined to a particular range value. Dispersion predicted values suffer from a relatively significant error as compared to real values which can be observed in Fig 5(iii). Effective mode area values also follow the same pattern as other predictions, the difference between actual and predicted values are almost negligible.

Following tables show the performance of the Neural Network in prediction across the three chosen PCF designs:

Table 1. Performance indices for output optical properties of Solid Core PCF

	Validation Set		Test Set	
	MSE	*R2*	*MSE*	*R2*
Effective Index (neff)	0.001	0.99	0.001	0.99
Confinement Loss	0.001	0.98	0.002	0.99
Dispersion	0.009	0.81	0.016	0.83
Effective Mode Area (Aeff)	0.001	0.99	0.002	0.98
Combined Data	**0.003**	-	**0.004**	-

Table 2. Performance indices for output optical properties of Multi Core PCF

	Validation Set		Test Set	
	MSE	*R2*	*MSE*	*R2*
Effective Index (neff)	0.001	0.99	0.001	0.99
Confinement Loss	0.003	0.98	0.004	0.97
Dispersion	0.019	0.90	0.036	0.86
Effective Mode Area (Aeff)	0.001	0.99	0.002	0.93
Combined Data	**0.005**	-	**0.009**	-

Table 3. Performance indices for output optical properties of Hollow Core PCF

	Validation Set		Test Set	
	MSE	*R2*	*MSE*	*R2*
Effective Index (neff)	0.001	0.99	0.001	0.99
Confinement Loss	0.009	0.82	0.013	0.89
Dispersion	0.001	0.99	0.001	0.99
Effective Mode Area (Aeff)	0.001	0.98	0.002	0.98
Combined Data	**0.003**	-	**0.005**	-

Table 1 show the prediction errors of various optical properties of Solid Core Design. It can be observed that the MSE of n_{eff}, confinement loss and A_{eff} are negligibly small and R2 value is close to 1 in both validation and test set. There is relatively more error in dispersion as compared to other values but overall, the MSE on combined data is consistent and very small. The same trend continues on the Hollow Core design as well. From Table 3, a consistent error of both validation set and test can be observed on combined data. However, in Table 3 of Multi-Core PCF design, the error in Dispersion value is relatively more. The reason for this could be the uneven distribution of values in dataset. However, the significance of 0.036 MSE is still very small. The high R2 value across all the predictions showed that model performs consistently well with negligible loss. R2 is generally not used to evaluate multi-output predictions as every variable can have different data distribution and a combined good fit may not be achieved.

Computation Time

The main motivation of this work was to leverage the advantage of machine learning techniques in terms of faster and accurate prediction of values given the training parameters. For our experiments we used a workstation with Intel i5 dual core processor, 8GB RAM running window 10 operating system.

Table 4. Simulation Time with COMSOL Multiphysics and Neural Network

	Size of Mesh Element					Neural Network Model	
	Normal	*Fine*	*Finer*	*Extra Fine*	*Extremely Fine*	*Training (Whole Data)*	*Testing*
Time Taken	7 sec	12 sec	**18 sec**	26 sec	38 sec	**19 sec**	**5 milli sec**

The size of mesh was fixed as Finer for all our simulations. It can be observed from Table 4 that our trained neural network with 2 layers having 70 nodes in 1st layer and 60 nodes in 2nd layer took 19 seconds for training on whole dataset of 3359 simulation values to learn the complex pattern while just 5 milliseconds to predict new set of optical properties using the trained model. However, while using COMSOL Multiphysics given input design parameters, it took 18 seconds just to compute one set of optical properties which is far more as compared to 5 milliseconds of our neural network model. Also, it can be observed that with increase in mesh size, simulation time also increases drastically.

5. Conclusions

This paper presents that a simple neural network can be successfully used as a generalised model to predict the optical properties of the various diverse PCFs efficiently. The compute times and computer resources required by the model are significantly lower than used by tradition numerical method techniques. The trained neural network requires only 5 milli seconds to predict optical properties for a new set of PCF design parameters. The same single calculation would take 18 seconds using COMSOL Multiphysics. The overall MSE error of predicted optical properties observed in different PCF design are in range of 0.004-0.009 which is highly insignificant. Such a neural network approach carries the potential to be exploited more in fields where numerical methods are heavily used for computation and simulation.

Disclosures

The authors declare no conflicts of interest.

Acknowledgment

The authors gratefully acknowledge the support provided by “TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communications at Delhi Technological University (Formerly Delhi College of Engineering), Delhi” through “Mission Reach” program of Technology Vision-2020, Government of India

References

[1] Photonic Crystal Fibers R. Buczynski, Information Optics Group, Faculty of Physics, Warsaw University Pasteura 7, 02-093 Warsaw, Poland.
[2] T. A. Birks, J. C. Knight, and P. S. J. Russell, “Endlessly single-mode photonic crystal fiber,” *Opt. Lett.*, vol. 22, no. 13, p. 961, Jul. 1997, doi: 10.1364/ol.22.000961.
[3] S. Chugh, A. Gulistan, S. Ghosh, and B. M. A. Rahman, “Machine learning approach for computing optical properties of a photonic crystal fiber,” *Opt. Express*, vol. 27, no. 25, p. 36414, Dec. 2019, doi: 10.1364/oe.27.036414.
[4] R. Kotyński, M. Dems, and K. Panajotov, “Waveguiding losses of micro-structured fibres-plane wave method revisited,” *Opt. Quantum Electron.*, vol. 39, no. 4–6, pp. 469–479, Mar. 2007, doi: 10.1007/s11082-007-9092-6.
[5] P. A. Agbemabiese and E. K. Akowuah, “Numerical analysis of photonic crystal fiber of ultra-high birefringence and high nonlinearity,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Dec. 2020, doi: 10.1038/s41598-020-77114-x.
[6] J. Baxter, A. Calà Lesina, J. M. Guay, A. Weck, P. Berini, and L. Ramunno, “Plasmonic colours predicted by deep learning,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Dec. 2019, doi: 10.1038/s41598-019-44522-7.
[7] I. Malkiel, A. Nagler, U. Arieli, M. Mrejen, L. Wolf, and H. Suchowski, “Deep Learning for Design and Retrieval of Nano-photonic Structures.”
[8] A. N. Tait *et al.*, “Machine Learning With Neuromorphic Photonics,” *J. Light. Technol. Vol. 37, Issue 5, pp. 1515-1534*, vol. 37, no. 5, pp. 1515–1534, Mar. 2019, Accessed: Feb. 26, 2021. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=jlt-37-5-1515.
[9] C. Yu and J. Jiang, “A perspective on using machine learning in 3D bioprinting,” *Int. J. Bioprinting*, vol.

6, no. 1, pp. 4–11, 2020, doi: 10.18063/ijb.v6i1.253.
[10] G. Moon, J. ryul Choi, C. Lee, Y. Oh, K. H. Kim, and D. Kim, "Machine learning-based design of meta-plasmonic biosensors with negative index metamaterials," *Biosens. Bioelectron.*, vol. 164, p. 112335, Sep. 2020, doi: 10.1016/j.bios.2020.112335.
[11] F. N. Khan, Q. Fan, J. Lu, G. Zhou, C. Lu, and A. P. T. Lau, "Applications of machine learning in optical communications and networks," in *Optics InfoBase Conference Papers*, Mar. 2020, vol. Part F174-OFC 2020, p. M1G.5, doi: 10.1364/ofc.2020.m1g.5.
[12] M. P. Fok and P. R. Prucnal, "Switching based on optical nonlinear effects," in *Optical Switches: Materials and Design*, Elsevier Ltd, 2010, pp. 181–205.
[13] C. Markos, J. C. Travers, A. Abdolvand, B. J. Eggleton, and O. Bang, "Hybrid photonic-crystal fiber," *Rev. Mod. Phys.*, vol. 89, no. 4, p. 045003, Nov. 2017, doi: 10.1103/RevModPhys.89.045003.
[14] J. C. Knight, T. A. Birks, P. S. J. Russell, and D. M. Atkin, "All-silica single-mode optical fiber with photonic crystal cladding," *Opt. Lett.*, vol. 21, no. 19, p. 1547, Oct. 1996, doi: 10.1364/ol.21.001547.
[15] F. Zolla *et al.*, *Foundations of photonic crystal fibres, 2nd edition*. Imperial College Press, 2012.
[16] "RP Photonics Encyclopedia - hollow-core fibers, photonic bandgap fibers, air-guiding fibers." https://www.rp-photonics.com/hollow_core_fibers.html (accessed Feb. 27, 2021).
[17] B. Ung and M. Skorobogatiy, "Transmission and propagation of terahertz waves in plastic waveguides," in *Handbook of Terahertz Technology for Imaging, Sensing and Communications*, Elsevier Inc., 2013, pp. 28–61.
[18] F. Benabid, "Hollow-core photonic bandgap fibre: New light guidance for new science and technology," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 364, no. 1849, pp. 3439–3462, Dec. 2006, doi: 10.1098/rsta.2006.1908.
[19] L. Michaille, C. R. Bennett, D. M. Taylor, and T. J. Shepherd, "Multicore photonic crystal fiber lasers for high power/energy applications," *IEEE J. Sel. Top. Quantum Electron.*, vol. 15, no. 2, pp. 328–336, Mar. 2009, doi: 10.1109/JSTQE.2008.2012134.
[20] H. Xiang and Y. Jiang, "Fiber Bragg grating inscription in multi-core photonic crystal fiber by femtosecond laser," *Optik (Stuttg).*, vol. 171, pp. 9–14, Oct. 2018, doi: 10.1016/j.ijleo.2018.06.020.
[21] "RP Photonics Encyclopedia - effective mode area, fiber, waveguide." https://www.rp-photonics.com/effective_mode_area.html (accessed Feb. 27, 2021).
[22] G. Bin Ren, Z. Wang, S. Q. Lou, and S. S. Jian, "Effective area of photonic crystal fiber," *Tien Tzu Hsueh Pao/Acta Electron. Sin.*, vol. 32, no. 5, pp. 723–726, May 2004, doi: 10.1364/oe.10.000341.
[23] N. T. Sørensen *et al.*, "Spectral macro-bending loss considerations on photonic crystal fibres," *IEE Proc. - Optoelectron.*, vol. 149, no. 5–6, pp. 206–210, 2002, Accessed: Feb. 27, 2021. [Online]. Available: https://orbit.dtu.dk/en/publications/spectral-macro-bending-loss-considerations-on-photonic-crystal-fi.
[24] N. A. Mortensen, J. R. Folkenberg, P. M. W. Skovgaard, and J. Broeng, "Numerical Aperture of Single-Mode Photonic Crystal Fibers," 2002. Accessed: Feb. 27, 2021. [Online]. Available: http://www.crystal-fibre.com.
[25] T. P. White, R. C. Mcphedran, C. M. De Sterke, L. C. Botten, and M. J. Steel, "Confinement losses in microstructured optical fibers," 2001.
[26] D. Marcuse, "Loss Analysis of Single-Mode Fiber Splices," *Bell Syst. Tech. J.*, vol. 56, no. 5, pp. 703–718, 1977, doi: 10.1002/j.1538-7305.1977.tb00534.x.
[27] R. Vallée, "The Optical Fibre," Springer, Berlin, Heidelberg, 2001, pp. 21–70.
[28] A. W. Snyder and J. D. Love, *Optical Waveguide Theory*. Springer US, 1984.
[29] M. Lucki, "Photonic Crystal Fibers with Optimized Dispersion for Telecommunication Systems," in *Recent Progress in Optical Fiber Research*, InTech, 2012.
[30] R. Ahmad, M. Komanec, D. Suslov, and S. Zvanovec, "Modified Octagonal Photonic Crystal Fiber for Residual Dispersion Compensation over Telecommunication Bands," doi: 10.13164/re.2018.0010.
[31] K. Ahmed, M. S. Islam, and B. K. Paul, "Design and numerical analysis: Effect of core and cladding area on hybrid hexagonal microstructure optical fiber in environment pollution sensing applications," *Karbala Int. J. Mod. Sci.*, vol. 3, no. 1, pp. 29–38, Mar. 2017, doi: 10.1016/j.kijoms.2017.02.001.
[32] A. Kajla and S. Gupta, "Evaluation of Confinement Loss of Different Photonic Crystal Fibers on the Basis of Varying the Size and Shape of Holes." Accessed: Feb. 27, 2021. [Online]. Available: www.ijert.org.
[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning, ICML 2015*, Feb. 2015, vol. 1, pp. 448–456, Accessed: Feb. 27, 2021. [Online]. Available: https://arxiv.org/abs/1502.03167v3.

[34] S. J. Russell, P. Norvig, J. F. Canny, J. M. Malik, and D. D. Edwards, "Artificial Intelligence A Modern Approach," 1995.
[35] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines."
[36] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," Dec. 2015, Accessed: Feb. 27, 2021. [Online]. Available: https://arxiv.org/abs/1412.6980v9.
[37] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015, doi: 10.1016/j.neunet.2014.09.003.
[38] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," 2014.
[39] Y. Lecun, "A theoretical framework for back-propagation." Morgan Kaufmann, pp. 21–28, 1988, Accessed: Feb. 27, 2021. [Online]. Available: https://nyuscholars.nyu.edu/en/publications/a-theoretical-framework-for-back-propagation-2.
[40] "Mean Squared Error (MSE)." https://www.probabilitycourse.com/chapter9/915meansquarederror MSE (accessed Feb. 27, 2021).
[41] Devore, Jay L. (2011). Probability and Statistics for Engineering and the Sciences (8th ed.) pp. 508–510.

Electrical Engineering
https://doi.org/10.1007/s00202-021-01424-0

ORIGINAL PAPER

Nonlinear adaptive normalized Huber control algorithm for 5-level distribution static compensator

Praveen Bansal[1] · **Alka Singh**[1]

Received: 11 March 2021 / Accepted: 9 October 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract
This paper proposes a new normalized Huber adaptive control algorithm for shunt compensation. A 5-level cascaded multilevel inverter (CHB-MLI) is controlled as a shunt active power filter (SAPF) unit for a medium voltage distribution system. The proposed algorithm is designed, modelled, and tested to control the SAPF unit under normal and distorted grid conditions. It is employed to extract the fundamental active power component of load current. This extracted effective component is used to generate source reference currents by multiplying with the unit vector template. Furthermore, the switching of CHB-MLI is performed using phase-shifted PWM scheme. The algorithm helps to mitigate the reactive power and unwanted harmonics generated by the nonlinear load. A single-phase scale-down prototype model is developed for experimental investigation and testing purpose in the laboratory. The associated voltage and current sensors and control circuits are controlled by the dSpace 1104 digital microprocessor. Simulation and experimental results show good performance of the proposed algorithm in steady-state and dynamic load conditions. A fair performance comparison is also done with other conventional adaptive algorithms.

Keywords Shunt active power filter · Multilevel inverter · Power quality · Nonlinear load · Adaptive control algorithms

1 Introduction

It is well known that good dynamic performance and cost-effective inverter control is a prerequisite for effective shunt filtering. As in the case of a grid-tied system, a two-level inverter usually faces problems such as malfunctioning of switches, high switching losses, high output harmonics, and the requirement of large-sized filter circuits. Therefore, a two-level inverter is not a preferred choice in medium- and high-voltage applications. Multilevel inverter (MLI) is emerging as a popular area of extensive investigations [1]. MLI aggregates several small voltages to achieve high voltage. Besides, MLI possesses low switching losses, low electromagnetic interference (EMI), and the possibility of redundant states [2].

✉ Praveen Bansal
pbansal444@gmail.com

Alka Singh
alkasingh.eed@gmail.com

[1] Department of Electrical Engineering, Delhi Technological University, New Delhi, Delhi 110042, India

The literature survey presents three benchmark MLI topologies [1–6]. The first one is introduced by Nabae et al. [3] and is named 3-level Diode Clamped Multilevel Inverter (DCMLI). This introduced the concept of midpoint Neutral Point Clamped (NPC) inverter. The second well-known configuration is Flying Capacitor MLI (FCMLI) introduced by Meynard and Foch [4] and Lavieville et al. [5] in the 1990s. It involves the basic idea to use capacitors instead of clamping diodes and is also known as capacitor clamped inverter. In the mid-1970s, Baker and Bannister [6] proposed a series connection of H-bridges popularly known as cascaded H-bridge MLI (CHB-MLI). It is a packed module structure, easy to use, and requires simple switching. In this paper, CHB-MLI is controlled as a SAPF unit triggered by phase-shifted (PS) modulation scheme [7].

Recent years have seen an increasing dependency on power electronics devices in domestic, industrial, and commercial applications. This has made our life easier, but various power quality (PQ) problems have also cropped up. Injection of harmonics in the distribution system is one common outcome of the proliferation of power electronics-based devices. The converter used in the system offers highly nonlinear characteristics and results in pollution of the electrical

Published online: 05 November 2021

grid. At the point of common coupling (PCC), various power quality (PQ) issues such as mal-operation of relays, overheating of conductors, interference with nearby communication lines, neutral burning, and voltage swell/sag are commonly observed and these issues directly affect the life span of various industrial loads connected nearby [8]. Therefore, a solution to such problems is pertinent, and SAPF is one of the best solutions available to mitigate PQ issues.

The effectiveness of SAPF depends upon how accurately the control algorithm extracts the reference current and compensates for the harmonics generated by the load. Various time and frequency domain algorithms have been presented in the literature [9–18]. Some of the conventional algorithms are very popular and effective such as synchronous reference frame theory (SRFT) [9]; instantaneous reactive power theory [10]; i_d–i_q algorithm which transforms the current signals to rotating reference frame [11, 12]; algorithms based on the cancellation of double frequency component [13]; several new algorithms have been proposed and include composite and model observer-based algorithm [14, 15]; kernel incremental learning algorithm [16], etc. These algorithms require internal tuning of parameters, substantially high arithmetical operations, and additional filters. Besides, various frequency-domain algorithms such as Kalman filter [17] and Fourier transform [18]-based algorithms, have also been implemented for the control of SAPF and distribution static compensator (DSTATCOM).

In recent years, adaptive control algorithms have been the subject of extensive investigation as a means for extracting the fundamental component. The reason for the increased use of adaptive techniques is that automatic adjustment of parameters, simplified approach, and good performance under transient conditions are achieved. Various adaptive algorithms [19–29] have been reported in the literature. Paper [19] discusses the multiple complex coefficient filter MMCF-SOGI-based control algorithm. Manoj et al. [20] have presented an adaptive zero attraction least mean square (ZA-LMS) algorithm; Naidu et al. [21] have implemented various adaptive control algorithms such as amplitude adaptive notch filter (AANF), the quasi-Newton (QN) and the sliding mode observer (SMO); Alam [22] have presented steady-state linear Kalman filter for compensating PQ problems; optimal step LMS algorithm has been discussed in [23]; SRLMMN control technique has been presented in [24]; Euclidean direction search-based adaptive technique in [25]. Shyu [26] has implemented model reference adaptive control; Widrow–Hoff adaptive algorithm has been presented by Modi [27] for DSTATCOM, backstepping adaptive controller for single-phase SAPF is presented by Jayashankar [28], Trilochan [29] has implemented the sparse LMS algorithm; and the Z^2-proportionate diffusion algorithm [30] has also been reported in the recent years. Different algorithms show varied response in terms of convergence speed, dynamic response, complexity and reduction in harmonic content. Moreover, all the algorithms have taken a conventional two-level inverter for verification of results, and most of the papers have not discussed the system performance under distorted grid conditions.

This paper is based on a new adaptive algorithm, namely normalized Huber [31] proposed for harmonic elimination and shunt compensation. The control of CHB-MLI is developed using the NHuber algorithm. The performance of NHuber is compared to conventional LMS, RLS algorithms and synchronous reference frame theory (SRFT). The algorithm is developed to work satisfactorily in mitigating harmonics of load current at a faster convergence rate with low steady-state error and high transient stability.

The main contributions in the presented work are:

1. To realize the model and development of 1-phase shunt active power filter (SAPF) employed for 5-level cascaded H-Bridge inverter (CHB-MLI) to showcase the advantage of MLI over the conventional 2-level inverter
2. A new NHuber adaptive control algorithm is implemented for power quality improvement and system performance in comparison with well-known conventional control algorithms such as LMS, RLS and SRFT
3. Proposed model is developed using MATLAB/Simulink platform, and extensive investigations have been done
4. Scaled-down prototype model is developed in the laboratory and tested under dynamic load conditions
5. Self-balancing dc link voltage of 100 V each is maintained without using any additional complex voltage balancing circuits.
6. Simulation and experimental results under load variations as well as sag/swell in supply voltage are investigated.

The proposed NHuber algorithm is tested to control the 5-level shunt active power filter to mitigate the harmonics generated by the load current. The proposed system is tested under various dynamics such as load variations, voltage sag and swell and even in distorted grid conditions. Further, different adaptive algorithms show varied performance under different operating conditions. After analysing the adaptive control algorithms [19–29], almost all the authors have focussed the work on implementing conventional 2-level inverters used as DSTATCOM or SAPF. It was further observed from literature survey that these adaptive algorithms have been tested at higher loads to achieve compliance with IEEE-519 level. However, the proposed N-Huber algorithm system is tested at lower load current, and it shows exemplary performance in terms of %THD content. Both simulation and experimental results are in complete agreement.

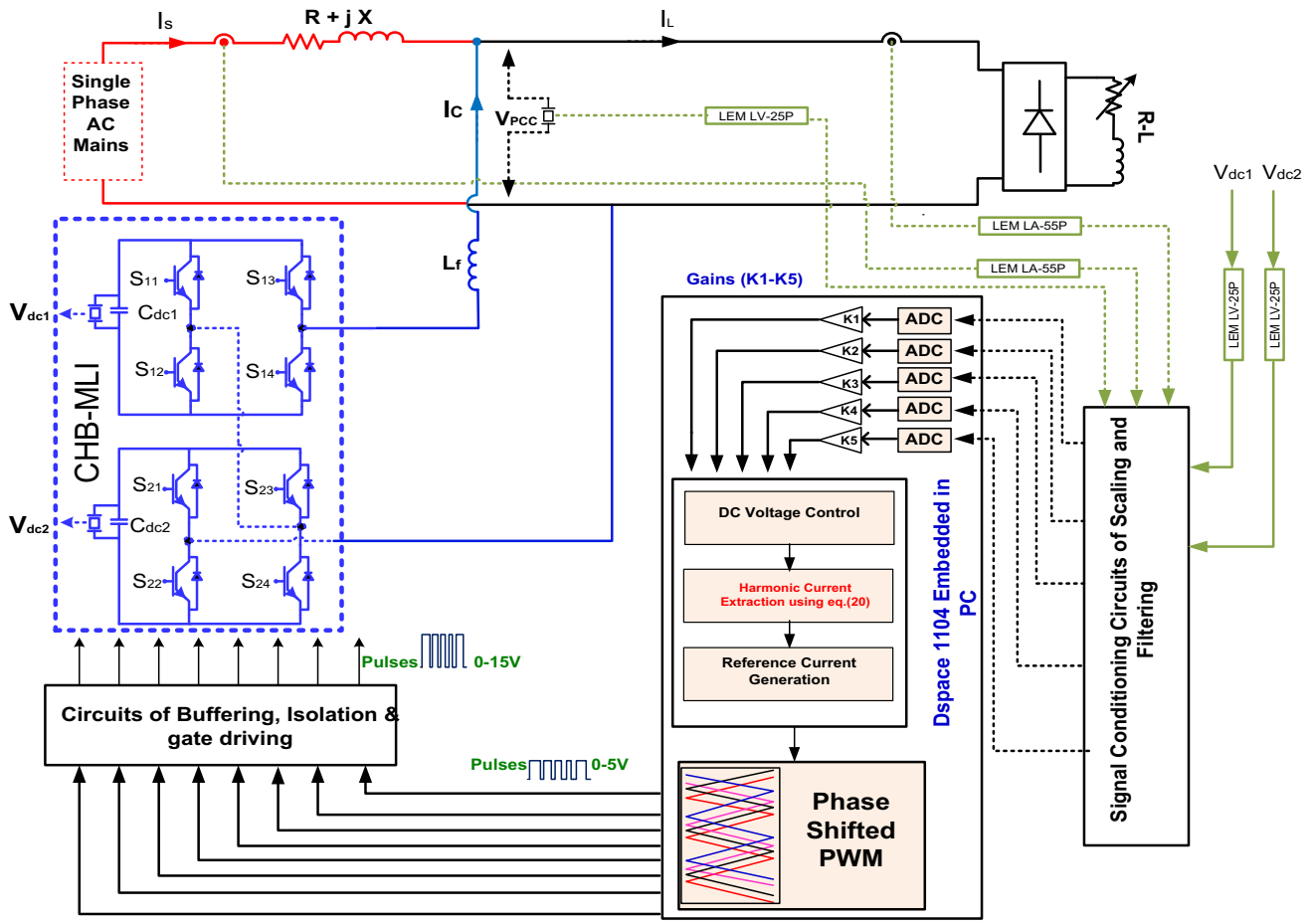


Fig. 1 Schematic diagram of CHB-MLI-based SAPF

The orientation of the paper is as follows. Section 3 describes the detailed modelling of CHB-MLI to design values of interfacing inductor (L_f), dc-link capacitors (V_{dc1} and V_{dc2}), and to select dc reference values ($V_{dc\text{-}ref}$). The detailed mathematical modelling of NHuber algorithm is presented in Sect. 4, and its subsection estimates the unit vector template and active loss component. Furthermore, the reference current generation and firing pulses for CHB-MLI are discussed. Sections 5 and 6 present the simulation and experimental results. Section 7 presents comparison results with other conventional adaptive algorithms. The final section presents the concluding remarks of the proposed algorithm for power quality applications.

2 System description

Figure 1 shows the schematic diagram of single-phase 5-level CHB-MLI SAPF connected to single-phase grid AC mains feeding the nonlinear load. To control the SAPF, the sensed input variables are voltage at the point of common coupling (PCC), load current (i_L), source current (i_s), and dc bus voltages V_{dc1}, V_{dc2}. An interfacing inductor (L_f) is connected to reduce the AC output ripples. The SAPF unit is current-controlled to inject suitable compensating current in phase opposition to eliminate the harmonics generated by the load current. For effective operation of the CHB-MLI, both the dc-link voltages must be maintained at a constant level and this function is realized by the conventional PI controller. The system is simulated using MATLAB/SIMULINK.

3 5-Level CHB-MLI configuration

The generalized system configuration is shown in Fig. 2. Two H-bridge inverters are connected in cascade to generate 5-level AC output voltage. Here, the generalized function 'S_z' defines a switching function per unit of output voltage, which is mathematically given in (1). It is switched in such a way that two switches in each bridge do not remain switched on simultaneously.

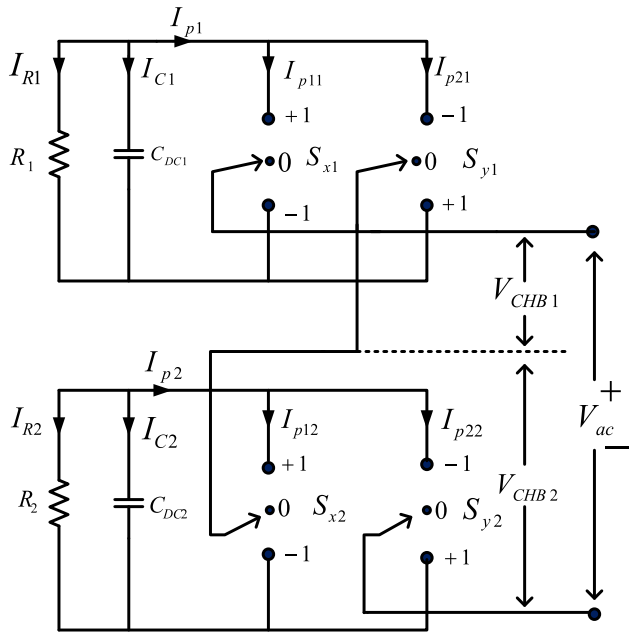


Fig. 2 Generalized schematic diagram of 5-level CHB-MLI

Table 1 Switching states and corresponding voltage levels

Voltage levels	V_{CHB1}	V_{CHB2}	S_{x1}	S_{y1}	S_{x2}	S_{y2}
$2V_{dc}$	V_{dc}	V_{dc}	+ 1	+ 1	+ 1	+ 1
V_{dc}	V_{dc}	0	+ 1	+ 1	− 1	+ 1
	V_{dc}	0	+ 1	+ 1	+ 1	− 1
	0	V_{dc}	− 1	+ 1	+ 1	+ 1
	0	V_{dc}	+ 1	− 1	+ 1	+ 1
0	$-V_{dc}$	V_{dc}	− 1	− 1	+ 1	+ 1
	V_{dc}	$-V_{dc}$	+ 1	+ 1	− 1	− 1
	0	0	+ 1	− 1	− 1	+ 1
	0	0	+ 1	− 1	+ 1	− 1
	0	0	− 1	+ 1	− 1	+ 1
	0	0	− 1	+ 1	+ 1	− 1
$-V_{dc}$	$-V_{dc}$	0	− 1	− 1	− 1	+ 1
	$-V_{dc}$	0	− 1	− 1	+ 1	− 1
	0	$-V_{dc}$	+ 1	− 1	− 1	− 1
	0	$-V_{dc}$	− 1	+ 1	− 1	− 1
$-2V_{dc}$	$-V_{dc}$	$-V_{dc}$	− 1	− 1	− 1	− 1

$$S_z = \begin{cases} 1 \quad \text{if} \quad S_{x1}, S_{y1}, S_{x2}, \quad \text{and} \quad S_{y2} = 1 \\ \frac{1}{2} \quad \text{if} \quad \begin{cases} \text{case} - 1\ S_{x1}, S_{y1} = 1 \quad \text{and} \quad (S_{x2} \& S_{y2}) \text{ are of opposite sign} \\ \text{case} - 2\ S_{x2}, S_{y2} = 1 \text{ and} \quad (S_{x1} \& S_{y1}) \text{ are of opposite sign} \end{cases} \\ 0 \quad \text{if} \quad \begin{cases} \text{case} - 1\ S_{x2,} S_{y2} = 1 \quad \text{and} \quad (S_{x1} \& S_{y1}) \text{ are of opposite sign} \\ \text{case} - 2\ S_{x1}, S_{y1} = 1 \quad \text{and} \left(S_{x2} \,\&\, S_{y2}\right) \text{are of opposite sign} \\ \text{case} - 3\ (S_{x1} \& S_{y1}) \quad \text{and} \quad (S_{x2} \,\&\, S_{y2}) \text{ are of opposite sign} \end{cases} \\ -\frac{1}{2} \quad \text{if} \quad \begin{cases} \text{case} - 1\ S_{x1}, S_{y1} = -1 \quad \text{and} (S_{x2} \,\&\, S_{y2}) \text{ are of opposite sign} \\ \text{case} - 2\ S_{x2}, S_{y2} = -1 \text{ and} (S_{x1} \& S_{y1}) \text{ are of opposite sign} \end{cases} \\ -1 \quad \text{if} \quad S_{x1}, S_{y1}, S_{x2} \quad \text{and} \quad S_{y2} = -1 \end{cases} \tag{1}$$

In (1), there are two switching functions 'S_{xj}' and 'S_{yj}' defined as generalized function for each leg of CHB-MLI. As shown in Fig. 2 'x' denotes the first leg and 'y' denotes the second leg. Furthermore, j = '1' for upper bridge and j = '2' for a lower bridge. The different switching states along with different voltage levels are shown in Table 1.

The symbols R_1 and R_2 represent the equivalent power losses in the upper and lower bridge, and C_{dc1} and C_{dc2} denote the dc-link capacitors; I_{pxj} and I_{pyj} represent current in leg 'x' and leg 'y', respectively.

3.1 Designing of (interfacing) inductor and (dc link) capacitors and dc reference value

The design of inductor and capacitor is discussed here.

3.1.1 Design of DC reference value

The importance of reactive power plays a very important role in selecting the dc-link reference voltage. Usually, for healthy and effective compensation, its value is selected higher than the supply voltage. The magnitude of dc bus capacitor voltage is calculated as [32]:

$$V_{\text{dc - ref}} = \frac{\sqrt{2}V_s}{\sqrt{3}m_f} \tag{2}$$

Here, $V_s = 110$ V is the source voltage, and $m_f = 0.8$ is the modulation index. The dc reference value is calculated as 112.25 V for each H-bridge. A combined value of 200 V is taken as the total dc reference voltage for this system.

3.1.2 Design of (interfacing) inductor

The value of L_f is calculated as [32]:

$$L_f = \frac{V_{dc,ref}}{12\,(m-1)\,f_{sw}\;\Delta I_{cr\,-\,pp}} \tag{3}$$

$V_{dc\text{-}ref} = 200$ V, $f_{sw} = 10$ kHz is the switching frequency, $m = 5$ represents voltage levels, $\Delta I_{cr\text{-}pp} = 0.2$ peak-to-peak ripple current. The computed value is 1.6mH and 3mH is selected.

3.1.3 Design of (DC link) capacitor

The proper designing of DC-link capacitors is needed which is calculated using the following relation [33]. Its expression can be given as:

$$C_{dc} = \frac{K_L V_s I_c T}{2(V_{dc\,-\,ref}^2 - V_{dc}^2)} \tag{4}$$

where V_s is taken as 110 V AC, $I_c = 4$A, $T = 0.02$ s, $V_{dc\text{-}ref} =$ 200 V, $V_{dc} = 190$ V, considering a voltage drop of 10 V, C_{dc} $= 2707.69$ μF is calculated, and for this system, it is taken as 3000 μF in experimentation and 2500 μF for simulation studies.

4 Mathematical formulation

4.1 Designing of control technique

The block diagram of the proposed algorithm is shown in Fig. 3. It is a fact that nonlinear load current can be depicted as:

$$i_L(t) = \sum\nolimits_{h=1}^{\infty} i_{Lh} \sin(h\omega t + \varphi_h) \tag{5}$$

$$i_L(t) = i_{L1} \sin(\omega t + \varphi_1) + \sum\nolimits_{h=2}^{\infty} i_{Lh} \sin(h\omega t + \varphi_h) \tag{6}$$

where h is the harmonic order, ϕ_h is the harmonic phase angle, and ω is the angular frequency of the fundamental harmonics.

The main objective of SAPF is to suppress all unwanted harmonics by generating compensation current which is equal and in phase opposition to the harmonic load current. The compensating current (i_c) can be represented as:

$$i_c = -\sum\nolimits_{h=2}^{\infty} i_{ch} = -(i_{c3} + i_{c5} + i_{c7} + i_{c9} + \cdots\cdots\cdots + i_{cn}) \tag{7}$$

For effective compensation, SAPF is designed and controlled in such a manner it would cancel all the harmonics and suppress these so that the grid current is predominantly sinusoidal. Hence, the sinusoidal current is represented as:

$$i_s(t) = i_{L1} + \sum_{h=2}^{\infty} i_{Lh} - \sum_{h=2}^{\infty} i_{ch} \tag{8}$$

Consequently, the sinusoidal source current in phase with supply voltage is given as:

$$i_s(t) = i_1 \sin(\omega t + \varphi_1) \tag{9}$$

Let $\omega(n)$ be the estimated weight vector and $x(n)$ denote the input signal vector and the error can be represented as:

$$e(n) = i_L - i_{Lf} \tag{10}$$

i_{Lf} is the fundamental component of load current extracted from NHuber algorithm and represented as:

$$i_{Lf} = x^T(n)\omega(n-1) \tag{11}$$

Putting the value of (11) in (10) gives

$$e(n) = i_{L:} - x^T(n)\omega(n-1) \tag{12}$$

where $x(n) = [x(n)\ x(n-1)\ \ldots\ldots\ldots\ldots x(n-M+1)]^T$ $(Mx1)$ denotes the input signal vector, and $w(n) = [w_o\ w_1$ $\ldots\ldots\ldots\ w_{(M-1)}]$ is the weighted vector of NHuber filter at instant n, and M indicates the size of matrix vector.

The objective of the algorithm [34] is to reduce the Huber cost function; $J(n)$ is defined as:

$$J(n) = \begin{cases} \frac{1}{2}e^2(n), & |e(n)| \le b \\ b|e(n)| - \frac{1}{2}b^2 & |e(n)| > b \end{cases} \tag{13}$$

There is a trade-off to choose the value of b. The appropriate value of b is to be chosen such that to avoid any steady-state misadjustment and the speed of convergence of adaptation. In addition, the stability of the NHuber algorithm is also affected by the eigenvalues of the autocorrelation matrix of input signals. Thus, the design of NHuber cost function is given by

$$J(n) = \frac{1}{2}\left|d(n) - w^T(n+1)x(n)\right|^2 + \frac{1}{2}\|x(n)\|^2\,\|w(n+1) - w(n)\|^2 \tag{14}$$

where $d(n)$ is the desired response, based on (14)

$$\frac{\partial J(n)}{\partial w(n+1)} = -e(n)x(n) + \|x(n)\|^2[w(n+1) - w(n)] = 0 \tag{15}$$

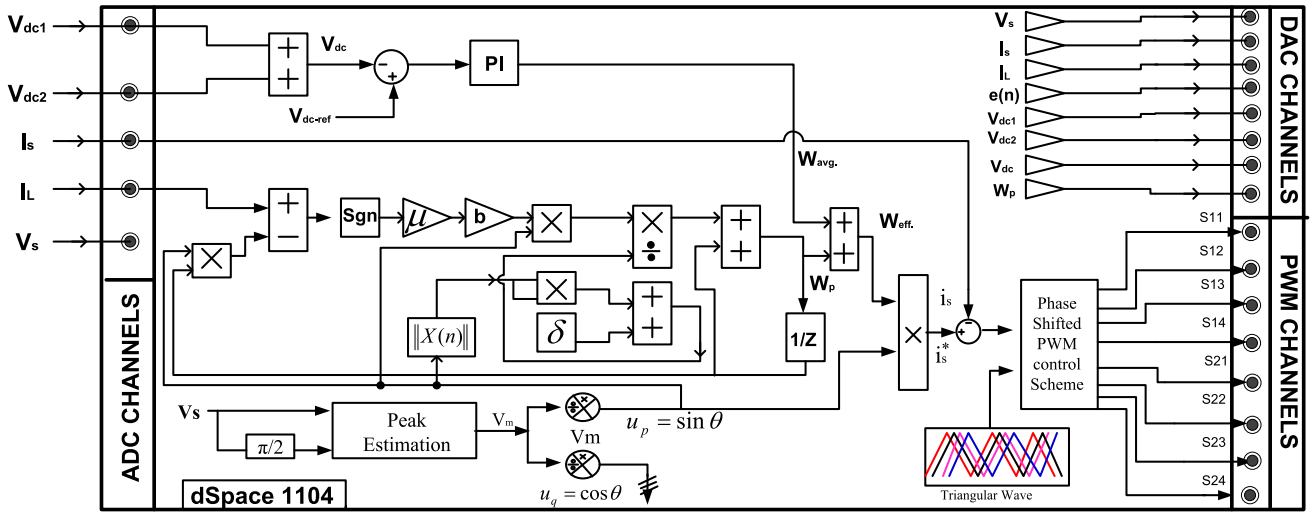


Fig. 3 Reference supply current estimation using N-Huber algorithm using dSpace 1104 under normal grid condition

Thus, weight updating (15) is as follows:

$$w(n+1) = w(n) + \mu \frac{e(n)x(n)}{\|x(n)\|^2} \tag{16}$$

where μ is the step size. If $|e(n)|>b$, then

$$J(n) = b\left|d(n) - w^{\mathrm{T}}(n+1)x(n)\right| - \frac{1}{2}b^2 + \frac{1}{2}\|x(n)\|^2 \|w(n+1)w(n)\|^2 \tag{17}$$

Thus, based on (17)

$$\frac{\partial J(n)}{\partial w(n+1)} = -b\,\mathrm{sgn}(e(n)\,x(n) + \|x(n)\|^2 [w(n+1) - w(n)] = 0 \tag{18}$$

Then, the final weight updating (18) is given as:

$$w(n+1) = w(n) + \mu \frac{b\,\mathrm{sgn}\,(e(n))x(n)}{\|x(n)\|^2} \tag{19}$$

To avoid division by small number, a very small positive constant δ is added to (16) and (19). Thus, the final weight updating equations for NHuber are defined as follows:

$$w(n+1) = \begin{cases} w(n) + \mu \frac{e(n)x(n)}{\delta + \|x(n)\|^2} & |e(n)| \le b \\ w(n) + \mu \frac{b\,\mathrm{sgn}(e(n))x(n)}{\delta + \|x(n)\|^2} & |e(n)| > b \end{cases} \tag{20}$$

In this paper, the weight updation is carried out for the fundamental active power weight (w_{p}) component of load current.

Figure 4 shows all vectors are column vectors, where input vector signals are represented by $x_n \in \Re^{Mx1} = [x[n] \quad x[n-1] \ldots\ldots\ldots\ldots x[n-M+1]]^{\mathrm{T}}$ and weight

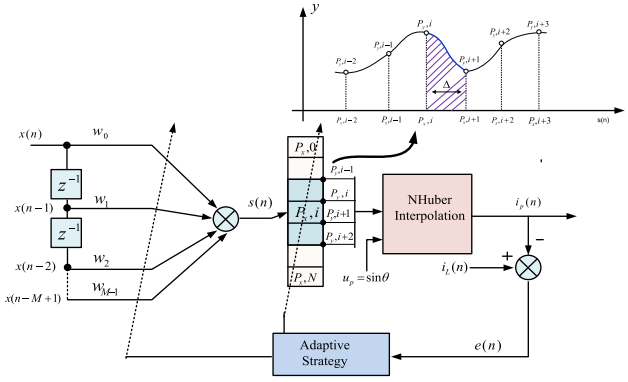


Fig. 4 Interpolation structure of NHuber algorithm

vector NHuber algorithm is depicted as $w \in \Re^{Mx1} = [w[0] \quad w[1] \ldots\ldots\ldots\ldots w[M-1]][1] \ldots\ldots\ldots\ldots w[M-1]]$.

M denotes the total number of matrix vector. For interpolation, the first step is to assign the interpolation intervals and find the interval index i corresponding to input signal $s(n)$. The look-up table (LUT) control points are denoted by $P_N \in \Re^{Px1} = [P_1 \quad P_2 \quad P_3 \ldots\ldots\ldots\ldots P_N]^{\mathrm{T}}$, where P_i denotes the ith control point, while ith span is represented as $P_N \in \Re^{4x1} = [P_{i-1} \quad P_i \quad P_{i+1} \quad P_{i+2}]$. Further, the output of NHuber interpolation is subtracted from actual load current $i_L(n)$ in order to generate an error $e(n)$. This adaptive process is continued till the steady state is achieved after load dynamics.

4.2 Estimation of unit vector templates

The grid supply voltage (V_{s}) is shifted to T/4, its shifted quadrature component (V_{qs}), and input component (V_{ps}) is used to estimate the peak value of supply which is further

used to generate the unit vector template, mathematically given as:

$$V_{sp} = V_m \, \mathrm{Sin}\, \omega t \tag{21}$$

$$V_{sq} = V_m \ \sin(\omega t + \pi/4) \tag{22}$$

$$V_t = \sqrt{V_{sp}^2 + V_{sq}^2} = V_m \tag{23}$$

The unit vector template can be calculated as:

$$U_p = \frac{V_{sp}}{V_m} = \mathrm{Sin}\omega t \tag{24}$$

4.3 Estimation of active loss component and reference grid current

The closed-loop operation of the proposed algorithm must include estimation of switching losses ($w_{avg.}$), which are calculated and supplied from the grid.

The error generated by the summation of sensed dc link voltages and $V_{dc\text{-}ref}$ is processed through the PI controller to meet the active switching loss component of the CHB-MLI and maintain both the dc-link voltages to their constant reference values; it is given as:

$$e_{dc}(n) = V_{dc} - V_{dc\text{ - }ref}^{*} \tag{25}$$

$$w_{loss}(n) = w_{loss}(n-1) + k_p\{e_{dc}(n) - e_{dc}(n-1)\} + k_i\, e_{dc}(n) \tag{26}$$

where k_p and k_i denote the proportional and integral gains of the PI controller and the effective weight component is given by

$$w_{eff} = w_p + w_{avg.} \tag{27}$$

4.4 Generation of firing pulses

NHuber algorithm is used to control SAPF. The reference current (i_s*) is calculated by multiplying the effective weight signal with unit vector template (U_p) given as:

$$i_s^{*} = w_{eff} * U_P \tag{28}$$

This current is compared with source current (i_s), and a phase-shifted (PS) PWM technique is employed for pulse generation. Furthermore, switching pulses are generated for controlling the insulated gate bipolar transistors (IGBTs) of CHB-MLI through necessary isolation and buffer circuits. Figure 5a, b shows the waveforms of the 5-level PS-PWM scheme and output phase voltage of CHB-MLI.

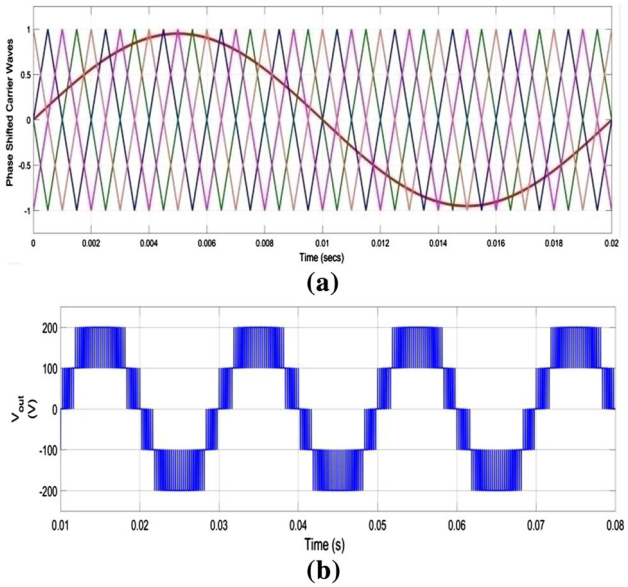


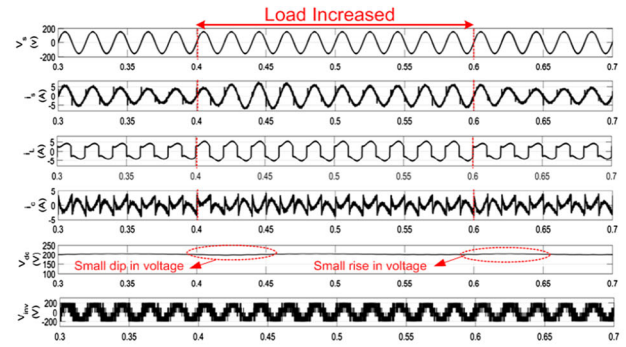
Fig. 5 **a**, **b** Phase-shifted scheme for 5-level CHB-MLI and output voltage of 5-level CHB-MLI. **a** Simulated waveforms of V_s, i_s, i_L, i_c, V_{dc}, V_{inv} during sudden load variation at $t = 1$ and $t = 1.4$ s. **b** Simulation results of i_L, V_{dc1}, V_{dc2}, V_{dc} during sudden variation of load at $t = 1$ s and $t = 1.4$ s

5 Simulation results

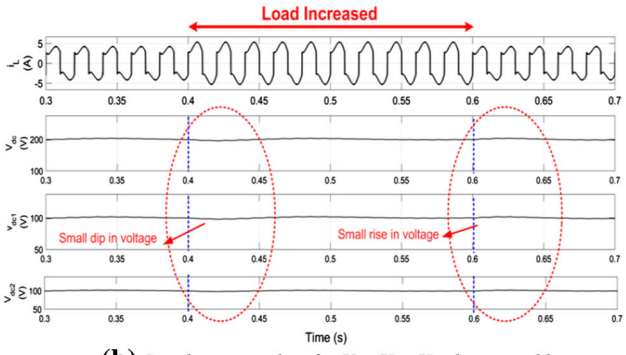
A Simulink model of the proposed system is developed in MATLAB. The performance of the system is tested under normal and distorted grid conditions. The CHB-MLI unit is modelled and controlled for power factor correction, dc-link voltage control, harmonics mitigation, and reactive power compensation. Appendix shows the simulation parameters. Figure 6a, b depicts the simulation results, and Fig. 6c shows the simulation results of the intermediate signals.

An uncontrolled rectifier is connected at the dc side of the system which is considered as nonlinear load. Figure 6a shows the observed quantities like source voltage (V_s), source current (i_s), load current (i_L), compensating current (i_C), total dc link voltage (V_{dc}), and inverter output voltage (V_{inv}). Figure 6b depicts the load current (i_L), total dc link voltages (V_{dc}) and dc-link voltage across capacitors (V_{dc1} and V_{dc2}). It is seen from Fig. 6a the load is changed at $t =$ 1 s, and the corresponding load current increases from 5 to 8A and is decreased back to original value at $t = 1.4$ s. The following observations are noticed:

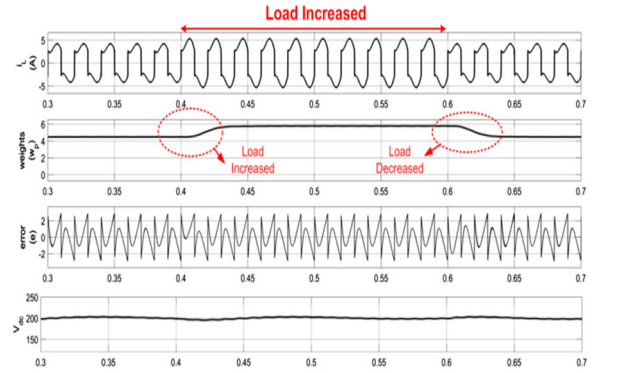
1. During dynamic load changes simulated at $t = 1$ s and $t = 1.4$ s, the total dc-link voltage (V_{dc}) sets itself to its reference value of 200 V, i.e. self-regulation is achieved within 1–2 cycles.
2. The CHB-MLI is suitably controlled so that the compensator current (i_c) cancels the effect of disturbances generated by the load at all times, even during load



(a) *Simulated waveforms of V_s,i_s,i_L,i_c,V_{dc},V_{inv} during sudden load variation at t=1 and t=1.4 s*



(b) *Simulation results of i_L,V_{dc1},V_{dc2},V_{dc} during sudden variation of load at t=1s and t=1.4 s*



(c) *Simulation waveforms of i_L, weights (w_p),error|(e) and V_{dc} during sudden variation of load at t=1s and t=1.4 s* estimated component of load current varies accordingly.

Fig. 6 **a** Simulated waveforms of V_s, i_s, i_L, i_c, V_{dc}, V_{inv} during sudden load variation at $t = 1$ and $t = 1.4$ s. **b** Simulation results of i_L, V_{dc1}, V_{dc2}, V_{dc} during sudden variation of load at $t = 1$ s and $t = 1.4$ s. **c** Simulation waveforms of i_L, weights (w_p), error|(e) and V_{dc} during sudden variation of load at $t = 1$ s and $t = 1.4$ s

changes. The resulting source current is smoothly controlled, and it is in phase with source voltage.

3. As observed in Fig. 6a, b, the performance of the PI controller is quite smooth. It is seen clearly that the PI controller quickly regulates the dc-link voltages of each capacitor.
4. The power and displacement factor are almost unity in steady-state and transient state conditions.

Figure 6c shows the waveforms of load current (i_L), weight (w_p), error (e) and V_{dc}. The control algorithm senses the parameters and calculates the updated weights corresponding to the fundamental component of load current. Furthermore, the reference current is generated to generate PWM pulses of CHB-MLI.

As depicted in Fig. 6b, additional load is added during 1–1.4 s and the voltages across capacitors V_{dc1} and $V_{dc2}$are self-regulated to 100 V each. Another advantage of proposed controller is that there is no need of complex dc link voltage balancing algorithms.

Figure 6c shows the intermediate results of the proposed algorithm such as error (e) and NHuber weights (w_p). As the nonlinear load increases, the active power and estimated component of load current vary accordingly.

It is seen from the results that the proposed algorithm effectively estimates the fundamental component of load current under steady-state as well as dynamic load conditions. The estimated values are computed fast requiring only 1–2 cycles for achieving convergence as shown in Fig. 6c. Figure 7a–c shows the harmonic spectrum of source voltage (V_s), source current (i_s) and load current (i_L). It is observed that THD of i_L is reduced from 30.11 to 2.67% in i_s after compensation.

The proposed system is also tested in distorted grid conditions. The control algorithm block diagram is modified and shown in Fig. 8. An additional block in the form of the second-order generalized integrator (SOGI)-based phase-locked loop (PLL) is now required and implemented as shown in Fig. 8. SOGI generates two orthogonal template signals from the input voltage signal [35].

The dynamic performance of the proposed system under the distorted grid is shown in Fig. 9. From $t = 0.85$ s to 1 s, distorted grid is considered and from $t = 1.3$ s to 1.4 s, load is increased from 2.4 to 3A. It is observed that the dc link voltage is maintained to its reference value of 200 V due to PI controller action within 1–2 cycles.

Figure 10 shows the dynamic results under distorted supply voltage (V_s), phase angle (θ), estimated sine template ($\sin\theta$) and estimated cosine template ($\cos\theta$). A distorted grid is considered from $t = 0.15$ s to 0.3 s. The SOGI filter extracts undistorted fundamental voltage template, and hence, it is possible to realize the control algorithm under distorted grid conditions.

6 Experimental results

The performance results obtained using Simulink model are now verified experimentally. Simulation and experimental parameters are given in the Appendix. A single-phase 110 V, 50 Hz system feeds a nonlinear load. This load is configured by connecting a series R–L branch and variable resistance

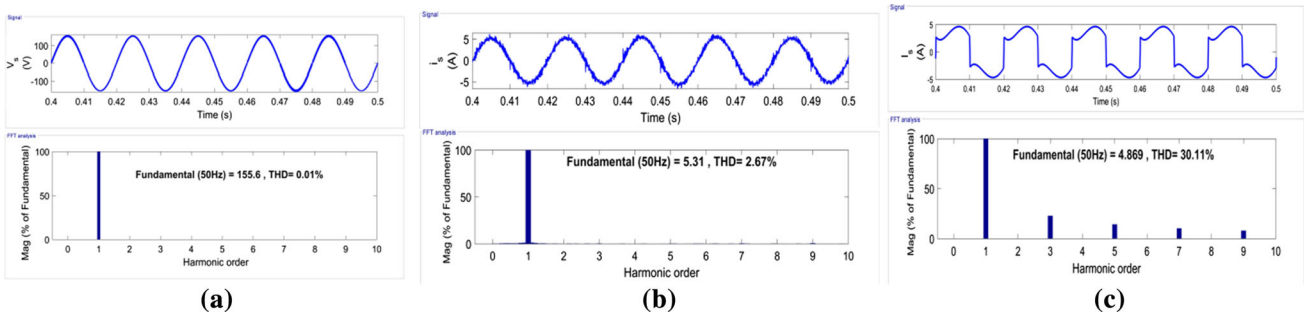


Fig. 7 **a**–**c** Harmonic spectra of **a** PCC voltage, V_s, **b** grid current is, **c** load current, i_L in PFC mode

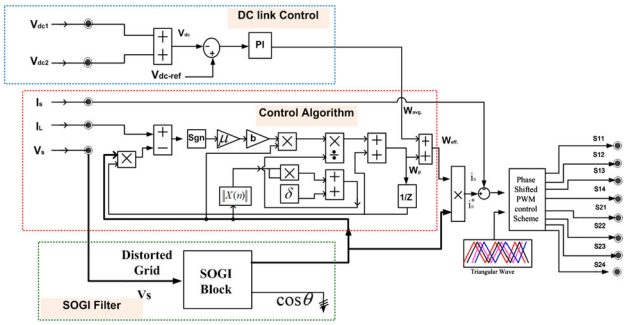


Fig. 8 Block diagram of NHuber controller under distorted grid

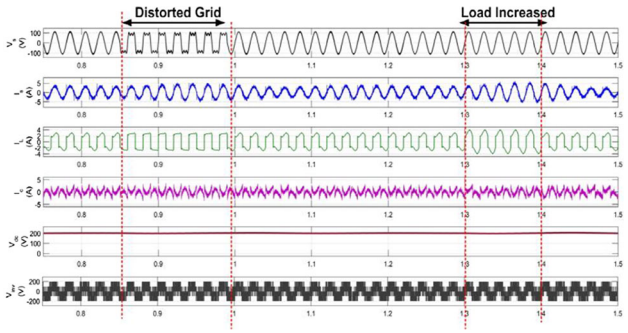


Fig. 9 Dynamic results under distorted grid condition **a** V_s **b** i_s **c** i_L **d** i_c **e** V_{dc} **f** V_{inv}

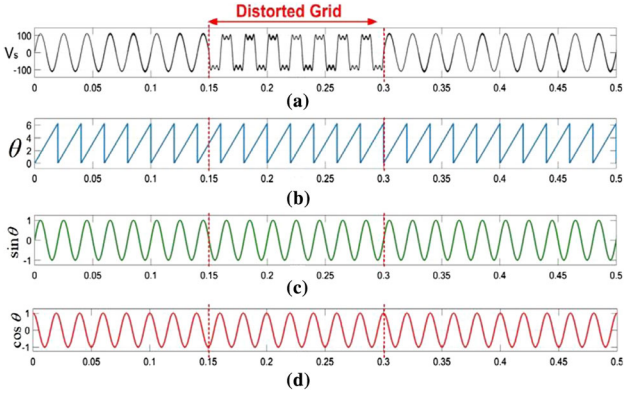


Fig. 10 Dynamic results under distorted grid condition **a** supply voltage (V_s); **b** phase angle (θ); **c** estimated sine template ($\sin\theta$); **d** estimated cosine template ($\cos\theta$)

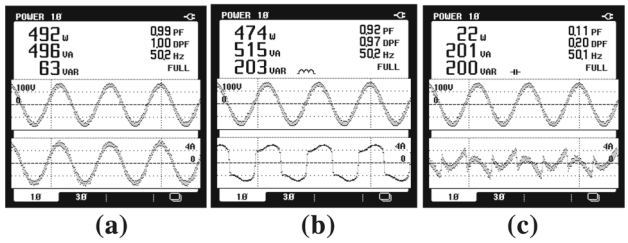


Fig. 11 **a** Supply power (P_s), **b** load power (P_L) and **c** power (P_c) of the system

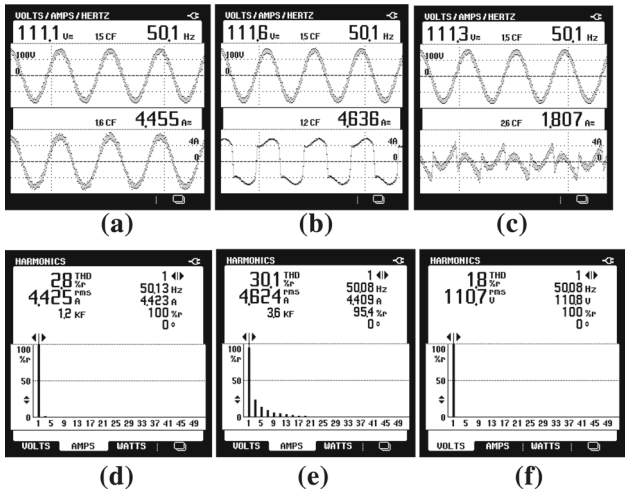
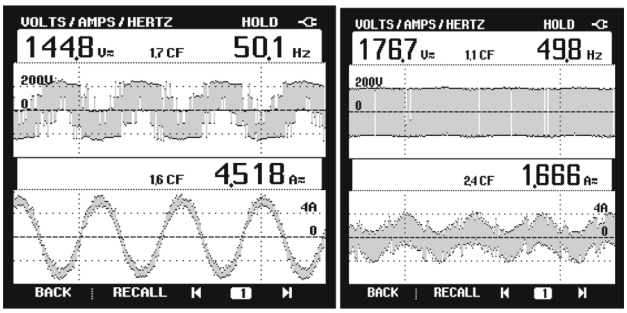


Fig. 12 **a**–**f** Steady-state waveforms of source current (i_s), load current (i_L) and compensating current (i_c) wrt PCC voltage (V_s) THD of 2.8%, 30.1% and 1.8% in i_s, i_l, V_s

unit at the ends of a diode rectifier. The voltage signals such as V_s, V_{dc1}, V_{dc2} are sensed using LV-25 V voltage sensors, and current signals i_L, i_s are sensed using LA-25P current sensors.

The proposed control algorithm generates the reference current and further generates PWM pulses buffer and isolation circuits to trigger IGBTs of CHB-MLI. The digital signal processor dSpace-1104 is used to process the control algorithm and brings PWM signals out.



(a) (b)

Fig. 13 **a**, **b** The output voltage waveform of 5-level MLI w.r.t i_s and output voltage of 2-level inverter w.r.t i_c

Figure 11a–c shows the steady-state results. It is evident that the load demands an active power of 0.474 kW and reactive power of 0.203 kvar. The supply active power demanded by the system is 0.492 kW which satisfies not only the load demand but also meets the switching losses of CHB-MLI. The compensator supplies 0.200 k VAR of the reactive power required by the load, as desired by the controller. The desired power flow is maintained between source, load, and compensator, and it is observed that the source current is sinusoidal and perfectly in phase with the supply voltage.

Figure 12a–f depicts the steady-state waveforms of i_c, i_L and i_s with respect to supply voltage V_s, followed by its %THD spectrum. It is observed that after compensation, current (i_s) THD is reduced to 2.8% and the THD of the supply voltage is 1.8%. These results show the highly satisfactory performance of the proposed control algorithm to achieve unity power at the source side. The proposed system also achieves the stipulated IEEE standards, thereby meeting PQ standards effectively.

Figure 13a, b shows the steady-state output voltage waveforms of 2-level and 5-level inverter. MLI possesses low switching losses; reduced harmonic is injected at PCC and can produce a good output voltage. The five-level inverter voltage is visible and has lower harmonic content as compared to two-level inverter.

Figure 14a–d shows the dynamic performance of the system under varying load conditions. The parameters V_s, i_s, i_L, V_{dc1}, V_{dc2}, V_{dc}, i_c and weights during sudden load variations are shown. It is observed from Fig. 14a that the source current is almost sinusoidal and in phase with voltage after

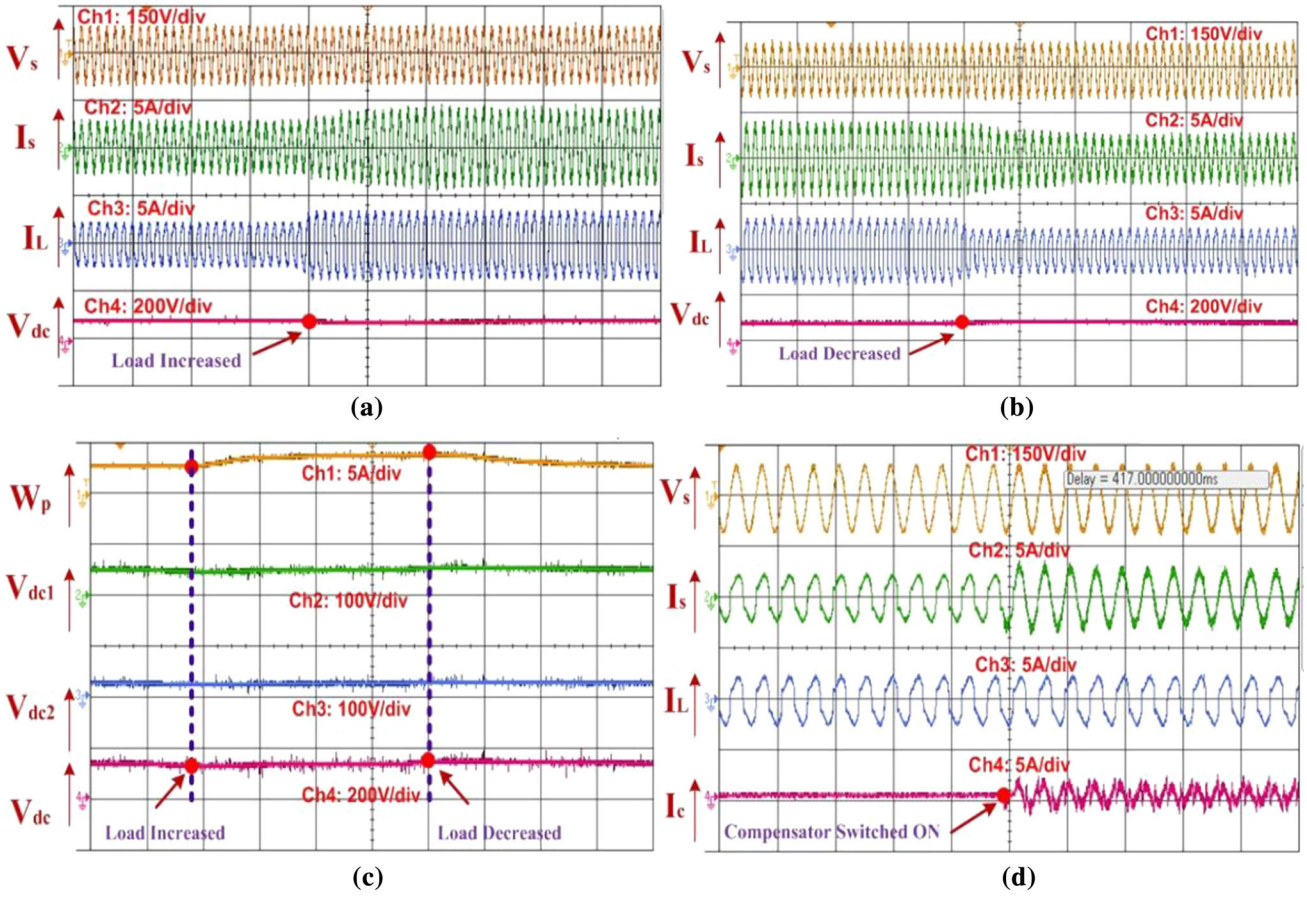


Fig. 14 **a–d** Weights (w_p), dc link voltage of capacitor 1 (V_{dc1}), dc link voltage of capacitor-2 (V_{dc2}) and total dc link voltage (V_{dc}), source voltage (V_s), source current (i_s), load current (i_L) and compensating current (i_c)

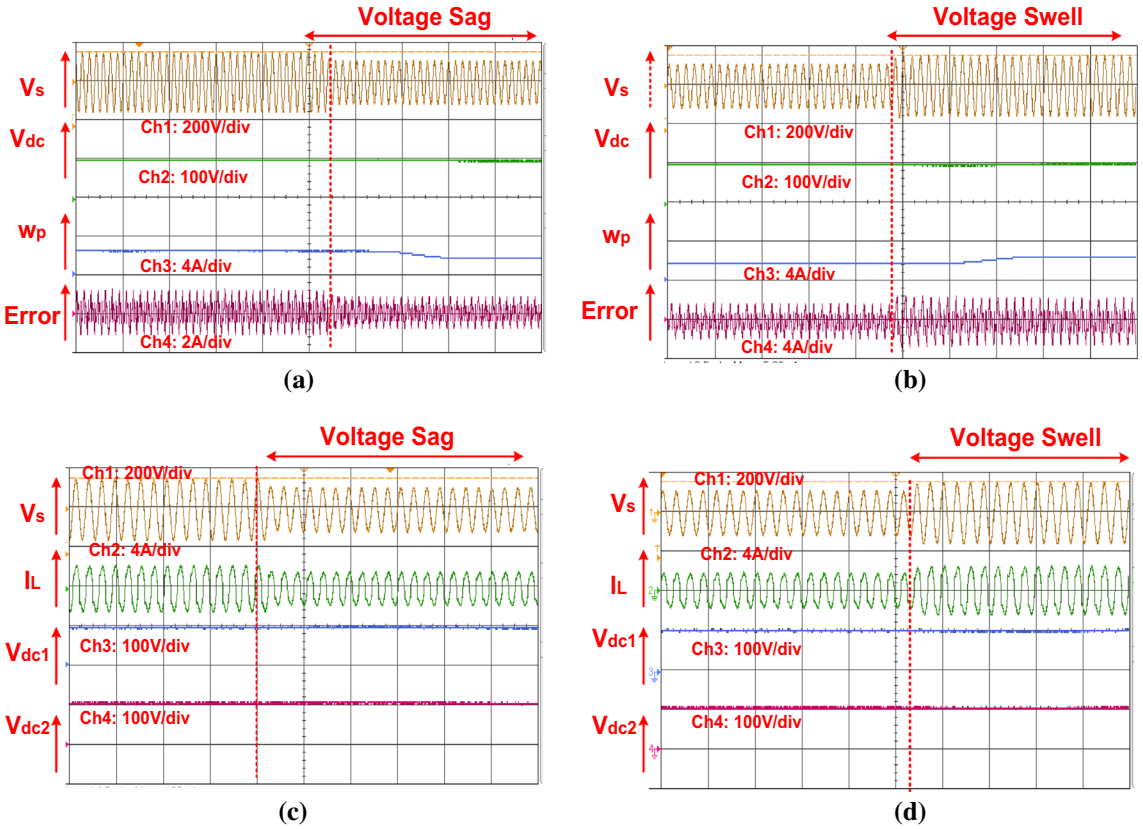


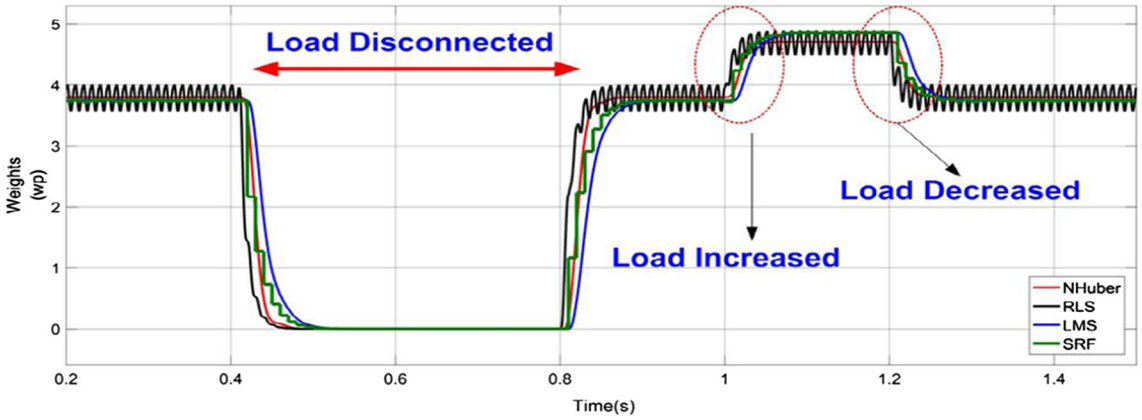
Fig. 15 **a–d** Experimental results under voltage sag and voltage swell conditions (**a**) V_s, I_L, V_{dc1} and V_{dc2}; **b** V_s, I_L, V_{dc1} and V_{dc2}; **c** V_s, V_{dc}, w_p (A) and error; (d) V_s, V_{dc}, $w_p(A)$ and error

Table 2 Comparative performance of algorithms

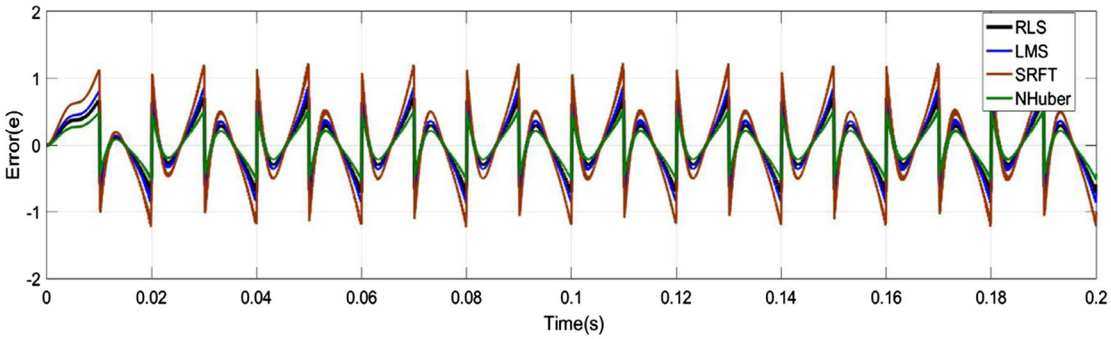
S. no	Parameters	NHuber	RLS	LMS	SRFT
1	Type of filter	Adaptive	Adaptive	Adaptive	Non-adaptive
2	Estimation of weights	Low overshoot of ± 2A	Moderate overshoot of ± 3A	High overshoot of ± 5A	Very high overshoot
3	THD of supply current (A, rms)	2.8% at 4.25A	3.9% at 4.12A	4.2% at 4A	4.3% 4.3A
4	Steady state error	Less (± 0.2%)	Higher than N-Huber (± 0.65%)	Higher than RLS (± 0.95%)	Not applicable
5	Convergence	Faster (1–2 cycles)	Moderate (3–4 cycles)	Lower (4–6 cycles)	Lower (6–7 cycles)
6	Complexity	Low	Moderate	High	High
7	Sampling time	50 µs	65 µs	70 µs	70 µs

compensation. DC link capacitors self-regulate the voltages within few cycles processed through the PI controller. Intermediate weights get converged quickly during sudden load variations.

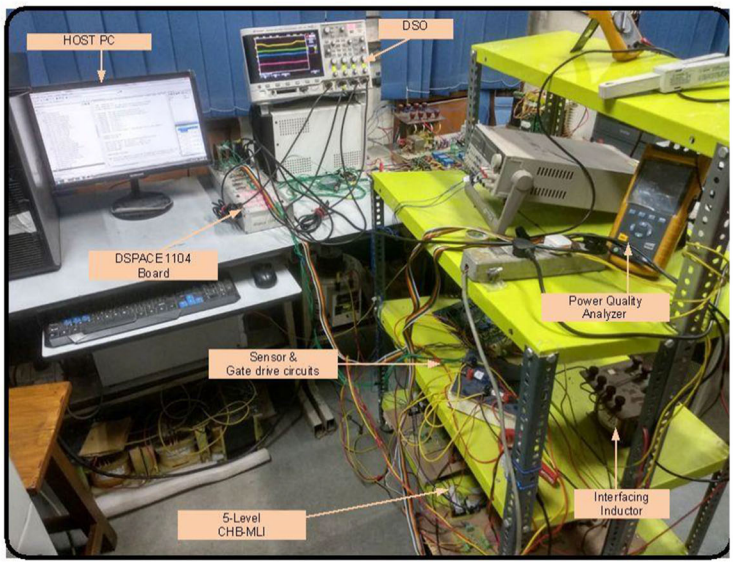
The proposed system is also tested under voltage sag and voltage swell conditions. Figure 15a–d shows the experimental results under voltage sag and swells conditions. The plots of PCC voltage (V_s), load current (i_L), dc link voltage across capacitor-1 (V_{dc1}), dc-link voltage across capacitor-2 (V_{dc2}), total dc-link voltage (V_{dc}), estimated fundamental weights (w_p), error (e) are shown. It is observed from these figures that even during sudden voltage sag or swell of magnitude ± 10 V, the dc-link voltage is restored to reference value. These results confirm the satisfactory closed-loop performance of the developed NHuber algorithm under supply voltage variations as well as load dynamics.



(a) *Weight convergence analysis of NHuber, LMS, RLS and SRFT algorithms under varying load conditions*



(b) *Error Comparison for SRFT, LMS, RLS and NHuber algorithms*



(c) *Experimental prototype developed in laboratory*

Fig. 16 **a** Weight convergence analysis of NHuber, LMS, RLS and SRFT algorithms under varying load conditions. **b** Error comparison for SRFT, LMS, RLS and NHuber algorithms. **c** Experimental prototype developed in laboratory

7 Comparative performance with conventional LMS, RLS and SRFT algorithms

Table 2 shows the comparative performance of NHuber with conventional LMS, RLS and SRFT algorithms. The first three are adaptive algorithms; however, the proposed algorithm fares better in terms of weight convergence, lower oscillations, and less sampling time required for processing. Figure 16a shows the dynamic weights under sudden variation during $t = 0.4$ to 0.8 s and $t = 1$ s to 1.2 s. It is concluded from waveforms that NHuber algorithm shows the least oscillations with reduced amplitude as compared to LMS, RLS and SRFT algorithms. Figure 16b shows the comparison of error between desired and estimated signals. Under dynamic loads weight, convergence is faster with low undershoot and low overshoot than conventional algorithms. RLS algorithm shows slightly fast convergence than LMS, but it is achieved at the expense of computational complexity. The proposed system works satisfactorily for load compensation. The THD of source current shows harmonic content of 4.3%, 4.2% and 3.9% for SRFT, LMS and RLS, respectively, while 2.8% THD is achieved using the proposed algorithm. Faster convergence of weights ensures better performance and closed-loop stability of the proposed algorithm under varying load conditions. A scaled-down prototype model is developed in the laboratory shown in Fig. 16c.

8 Conclusion

This paper primarily focuses on the application of the NHuber adaptive control algorithm implemented for CHB-MLI-based shunt compensation unit to mitigate various PQ problems. It is depicted from the simulation and experimental results that the proposed system works satisfactorily under steady-state as well as transient load varying conditions. When compared to conventional adaptive algorithms, the proposed algorithm requires less convergence time, shows lower oscillations in weights and error, less overshoot, and is quite robust. A scaled model is developed in the laboratory to verify the results. Hardware results are in complete agreement with the simulation results. The CHB-MLI is realized and controlled as a SAPF. The paper also justifies the numerous advantages of a 5-level inverter over a 2-level inverter. A lower THD is obtained at low magnitudes of source current, and also IEEE-519 standards are adequately maintained. Effective load compensation is achieved, and power quality problems are mitigated using the developed NHuber algorithm.

Acknowledgement The authors are thankful to Department of Science and Technology, India.

Appendix

Simulation parameters

Grid supply voltage, $V_s = 110$ V, nonlinear load of resistance and inductance, $R = 10\Omega$ and $L = 50$mH, dc bus capacitance $= 2500$ μF, $k_p = 1.2$, $k_i = 0.1$, $V_{dc\text{-}ref} = 200$ V, interfacing inductor $L_f = 3$mH, switching frequency $f_{sw} = 10$ kHz, sampling time $= 4$ μs.

Experimental parameters

Grid supply voltage, $V_s = 110$ V, 1-phase 2-leg diode rectifier as nonlinear load of resistance and inductance, $R = 40\Omega$ and $L = 8$mH, dc bus capacitance $= 3000$ μF, $k_p = 1.2$, $k_i = 0.1$, $V_{dc\text{-}ref} = 200$ V, interfacing inductor, $L_f = 3$mH, switching frequency $f_{sw} = 10$ kHz, sampling time $= 50$ μs.

References

1. Rodriguez J, Franquelo LG, Kouro S et al (2009) Multilevel converters: an enabling technology for high-power applications. Proc IEEE 97(11):1786–1817. https://doi.org/10.1109/JPROC.2009.2030235
2. Franquelo LG, Rodriguez J, Leon JI, Kouro S, Portillo R, Prats MAM (2008) The age of multilevel converters arrives. Ind Electron Mag IEEE 2(2):28–39. https://doi.org/10.1109/MIE.2008.923519s
3. Nabae A, Takahashi I, Akagi H (1980) Neutral-point-clamped PWM inverter. In: Conference on record IEEE IAS annual meeting, Cincinnati, OH, vol 3, pp 761–766. https://doi.org/10.1109/TIA.1981.4503992
4. Meynard TA, Foch H (1992) Multi-level conversion: high voltage choppers and voltage-source inverters. In: Proceedings of the IEEE power electronics specialist conference, pp 397–403. https://doi.org/10.1109/PESC.1992.254717
5. Lavieville JP, Carrere P, Meynard T (1997) Electronic circuit for converting electrical energy and a power supply installation making use thereof, U.S. Patent 5,668,711
6. Malinowski M, Gopakumar K, Rodriguez J, Perez MA (2010) A survey on cascaded multilevel inverters. IEEE Trans Ind Electron 57(7):2197–2206. https://doi.org/10.1109/TIE.2009.2030767
7. Tsunoda A, Hinago Y, Koizumi H (2014) Level-and phase-shifted PWM for seven-level switched-capacitor inverter using series/parallel conversion. IEEE Trans Ind Electron 61:4011–4021. https://doi.org/10.1109/TIE.2013.2286559
8. Akagi H, Watanabe EH, Aredes M (2017) Instantaneous power theory and applications to power conditioning. Wiley, NJ
9. Gonzalez-Espin F, Figueres E, Garcera G (2012) An adaptive synchronous reference-frame phase-locked loop for power quality improvement in a polluted utility grid. IEEE Trans Ind Electron 59(6):2718–2731. https://doi.org/10.1109/TIE.2011.2166236
10. Herrera RS, Salmerón P, Kim H (2008) Instantaneous reactive power theory applied to active power filter compensation: different approaches, assessment, and experimental results. IEEE Trans Ind Electron 55(1):184–196. https://doi.org/10.1109/TIE.2007.905959
11. Singh A, Singh B, Mittal AP (2010) Performance and control of converter interfaced distribution resources. In: Proceedings of the IEEE IICPE conference, pp 1–6. https://doi.org/10.1109/IICPE.2011.5728124

12. Pouresmaeil E, Montesinos-Miracle D, Gomis-Bellmunt O, Sudrià-Andreu A (2011) Instantaneous active and reactive current control technique of shunt active power filter based on the three-level NPC inverter. Eur Trans Electr Power 21(7):2007–2022. https://doi.org/10.1002/etep.536
13. Singh B, Arya S, Verma P (2014) An implementation of double-frequency oscillation cancellation technique in control of DSTATCOM. Int Trans Electr Energy Syst 24:796–807. https://doi.org/10.1002/etep.1735
14. Singh B, Arya S (2013) Composite observer-based control algorithm for distribution static compensator in four-wire supply system. IET Power Electron 6(2):251–260. https://doi.org/10.1049/iet-pel.2012.0412
15. Valdez-Fernandez A, Martinez P, Escobar G, Limones-Pozos C, Sosa J (2013) A model-based controller for the cascade H-bridge multilevel converter used as a shunt active filter. IEEE Trans Ind Electron 60(11):5019–5028. https://doi.org/10.1109/TIE.2012.2218558
16. Arya S, Singh B (2015) Implementation of kernel incremental learning algorithm in distribution static compensator. IEEE Trans Power Electron 30(3):1157–1169. https://doi.org/10.1109/TPEL.2014.2315495
17. Chittora P, Singh A, Singh M (2016) Performance evaluation of a new Kalman filter based least mean square algorithm for power quality improvement. In: International conference on power electronics, intelligent control and energy systems (ICPEICES), Delhi, India, pp 1–5. https://doi.org/10.1109/ICPEICES.2016.7853382
18. Badoni M, Singh A, Singh B (2016) Adaptive neuro fuzzy inference system least-mean-square-based control algorithm for DSTATCOM. IEEE Trans Ind Inf 12(2):483–492. https://doi.org/10.1109/TII.2016.2516823
19. Chittora P, Singh A, Singh M (2017) Simple and efficient control of DSTATCOM in three phase four wire polluted grid system using MCCF-SOGI based controller. IET Gener Trans Distrib 12(5):1–10. https://doi.org/10.1049/ietgtd.2017.0901
20. Badoni M, Singh A, Singh B (2016) An implementation of variable step-size least-mean-square based control algorithm for DSTATCOM. Int Trans Electr Energ Syst 26:1540–1554. https://doi.org/10.1049/iet-gtd.2017.0901
21. Naidu TA, Arya SR, Al-Durra A, El-Fouly THM (2020) Comparative performance of dynamic voltage restorer using adaptive control algorithms with optimized error regulator gains. Int Trans Electr Energy Syst 2020:e12696. https://doi.org/10.1002/2050-7038.12696
22. Alam SJ, Arya SR (2020) Control of UPQC based on steady state linear Kalman filter for compensation of power quality problems. Chin J Electr Eng 6(2):52–65. https://doi.org/10.23919/CJEE.2020.000011
23. Patel SK, Arya SR, Maurya R (2019) Optimal step LMS-based control algorithm for DSTATCOM in distribution system. Electr Power Components Syst 47(8):675–691. https://doi.org/10.1080/15325008.2019.1602797
24. Badoni M, Alka S, Bhim S, Hemant A (2020) Real-time implementation of active shunt compensator with adaptive SRLMMN control technique for power quality improvement in the distribution system. IET Gener Transm Distrib 14(8):1598–1606. https://doi.org/10.1049/iet-gtd.2019.0929,10.1049/iet-gtd.2019.0929
25. Badoni M, Singh A, Singh B (2020) Power quality enhancement using Euclidean direction search based control technique. IEEE Trans Ind Electron 67(3):2231–2240. https://doi.org/10.1109/TIE.2019.2905835
26. Shyu K, Yang M, Chen Y, Lin Y (2008) Model reference adaptive control design for a shunt active-power-filter system. IEEE Trans Ind Electron 55(1):97–106. https://doi.org/10.1109/IECON.2006.347931
27. Modi G, Kumar S, Singh B (2020) Improved Widrow-Hoff based adaptive control of multi objective PV-DSTATCOM system. IEEE Trans Ind Appl 56(2):1930–1939. https://doi.org/10.1109/TIA.2019.2960732
28. Jayasankar VN, Vinatha U (2020) Backstepping controller with dual self-tuning filter for single-phase shunt active power filters under distorted grid voltage condition. IEEE Trans Ind Appl 56(6):7176–7184. https://doi.org/10.1109/TIA.2020.3025520
29. Trilochan P, Mrutyunjaya M, Kumar PA, Kumar SS (2016) Sparse LMS control algorithm for fuel cell based SAPF. In: 2016 IEEE Uttar Pradesh section international conference on electrical, computer and electronics engineering (UPCON), Varanasi, pp 72–77. https://doi.org/10.1109/UPCON.2016.7894627
30. Lee H-S, Yim S-H, Song W-J (2017) z^2-proportionate diffusion LMS algorithm with mean square performance analysis. Sig Process 131:154–160. https://doi.org/10.1016/j.sigpro.2016.06.011
31. Li Z, Guan S (2018) Diffusion normalized Huber adaptive filtering algorithm. J Frankl Inst 355(8):3812–3825. https://doi.org/10.1016/j.jfranklin.2018.03.001
32. Waware M, Agarwal P (2012) Hardware realization of multilevel inverter-based active power filter. IETE J Res Sep 58(5):356–366. https://doi.org/10.4103/0377-2063.104152
33. Ray S, Gupta N, Gupta RA (2018) Hardware realization of proportional-resonant regulator based advanced current control strategy for cascaded H-bridge inverter-based shunt active power filter. Int Trans Electr Energy Syst 66:e2714. https://doi.org/10.1002/etep.2714
34. Petrus P (1999) Robust Huber adaptive filter. IEEE Trans Signal Process 47(4):1129–1133. https://doi.org/10.1109/78.752610
35. Ciobotaru M, Teodorescu R, Blaabjerg F (2006) A new single-phase PLL structure based on second order generalized integrator. In: 2006 37th IEEE power electronics specialists conference, Jeju, Korea (South), pp 1–6. https://doi.org/10.1109/pesc.2006.1711988

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/355786082

Optimizing the Machining performance of CNC tools inserts coated with Diamond like Carbon Coatings under the dry cutting environment

Article *in* Journal of Engineering Research · October 2021

DOI: 10.36909/jer.ICARI.15327

CITATIONS
0

READS
28

4 authors, including:



Ramakant Rana
Maharaja Agarsain Institute of Technology
40 PUBLICATIONS **75** CITATIONS

SEE PROFILE



Qasim Murtaza
Delhi Technological University
106 PUBLICATIONS **1,279** CITATIONS

SEE PROFILE



R. S. Walia
Delhi Technological University
167 PUBLICATIONS **1,216** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Study of dry wear behavior and morphological characteristic of 60%Mo-20%NiCr-10%CrC-10%Mo+Fe based alloy coating by atmospheric plasma spray technique View project

conjugate heat transfer within cavity View project

All content following this page was uploaded by Ramakant Rana on 31 October 2021.

The user has requested enhancement of the downloaded file.

Journal of Engg. Research ICARI Special Issue pp. 142-152 DOI : 10.36909/jer.ICARI.15327

Optimizing the Machining performance of CNC tools inserts coated with Diamond like Carbon Coatings under the dry cutting environment

Ramakant Rana*,****, Lucky Krishnia**, Qasim Murtaza* and R.S. Walia***

** Department of Mechanical Engineering, Delhi Technological University, Delhi, India*
*** Amity Center of Nanotechnology, Amity University Haryana -122413, India*
**** Department of Production and Industrial Engineering, Punjab Engineering College, Chandigarh, India*
***** Corresponding Author : ramakantrana@gmail.com*

ABSTRACT

In this research work we have coated WC (tungsten carbide) tool inserts with diamond like carbon coatings. The deposition was done using the thermal Chemical Vapour Deposition (CVD) method. We have used bagasse of sugarcane which is an agriculture waste as a carbon precursor for developing the diamond like carbon coatings. Surface finish, cutting temperature and cutting forces of as-developed diamond like carbon coatings on the WC inserts were examined and analyzed in this research work. For confirming the presence of diamond like carbon coatings on the surface of coated tungsten carbide (WC) tool inserts substrate we have used Raman spectroscopy, X-ray diffraction (XRD) and Field Emission Scanning Electron Microscope (Fe-SEM). The increase in the hardness on the developed coated samples were inspected by performing the hardness test on the coated substrates. The average evaluated Vickers Hardness number were found to be 1455.24 HV and 950.65 HV for both coated as well as un-coated samples respectively. This shows 53% rise in the hardness of the un-coated WC inserts.

Key words: Tungsten carbide; Diamond coating; Wear; Friction; Tool inserts.

INTRODUCTION

A vast requirement of tool has emerged now a days with better tool life in the manufacturing industry. The tool are basic requirement in the manufacturing processes, but the requirements of manufacturing industries are getting changed day by day in very rapid pace. These rapid changes have increased the demand of tools for machining hard materials used in aerospace & automobile industries (Tanaka and Akasawa, 1999, Boyer et al., 2015). It is getting difficult for machinist to get high tolerances while producing the high volumes of machined parts and material. These hard to cut materials with high precision require something other than conventional cutting tools (Rana et al., 2014, Ullen et al., 2020).

Coated tungsten carbide (WC) tools are ideal for the use in the field of manufacturing industries for machining the hard difficult metals. These tools offer good resistance to the wear and friction while producing the better surface fining while machining. Moreover, these WC tools have an established track record of being productive in producing large amount of parts (Fraga et al., 2016, Chandran and Hoffman, 2016).

An allotrope of carbon which is also known as diamond has drawn a lot of positive attention from the researcher, machinists, worksman, etc. from the field of technology and engineering because of its high underlying properties (Gracio et al., 2010, Srinivasan et al., 2016). As we already know that the diamond is the hardest and least compressible

material. It has one of the best thermal conductivity when measured at the room temperature. Technique of thermal chemical vapour deposition (CVD) is used to grow the diamond film on the substrate material like tungsten carbide (WC). This technique has pulled considerable attention of many researchers and scientist for many years now because of it's possible usage in the area of technology and manufacturing (Asmussen and Reinhard 2002, Schneider et al., 2010, Najar et al., 2018). The major source of agricultural waste produced is the sugarcane bagasse (SBg). Indian is the 2nd largest agricultural waste producing country, as sugarcane industries of India produces 90 million ton of sugarcane bagasse. This waste is just normally burned as any other normal waste material produced (Madurwar et al., 2015, Jeyapandiarajan and Xavior 2019). One cannot reduce the burnt sugarcane bagasse any further (Lima et al., 2012).

There are several advantages to thermal chemical vapour deposition (CVD) technique, like ease of operation, adaptability for all difficult geometries (Urakami et al., 2019). The diamond after being coated on the tungsten carbide (WC) substrate tool material produces a very hard cutting tool. These qualities of diamond coated tungsten carbide (WC) tools produces better results in machining operations (Sein et al., 2003, Bobylyov and Storozhenko, 2019). These self-developed diamond coated tungsten carbide (WC) tools have good surface finish because of which they have good applications (Chandran et al., 2013, Bansal et al., 2021). The industry focusing on the cutting or in other words machining application uses diamond coated tungsten carbide (WC) tools, these tools have splendid properties like better thermal conductivity and extravagant hardness (Haubner and Lux, 1996, Polini et al., 2012).

This paper, reports the parametric optimization of the developed diamond like carbon coated tungsten carbide (WC) tool insert for machining processes using TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) (Rana et al., 2021b). XRD along with Fe-SEM and Raman spectroscopy were also utilized to characterize the as-developed diamond like carbon coatings.

MATERIALS, METHODS AND CHARACTERIZATION

Coating Method

In this work, diamond like carbon coatings were coated on the WC inserts using thermal CVD technique. Pyrolyzed sugarcane bagasse (p-SBg) was produced by SBg, which then is used to produce activated carbon (Tyagi et al., 2019, Krishnia and Tyagi, 2018). Bagasse of sugarcane was used to grow the diamond like carbon coatings. As stated, that sugarcane bagasse has been the major waste produced from sugar industries (Madurwar et al., 2015) and burning this waste is not an environment-friendly solution (Lima et al., 2012). The deposition rate for diamond like carbon coating was found ~55 μm/h (Krishnia et al., 2018, Rana et al., 2021a).

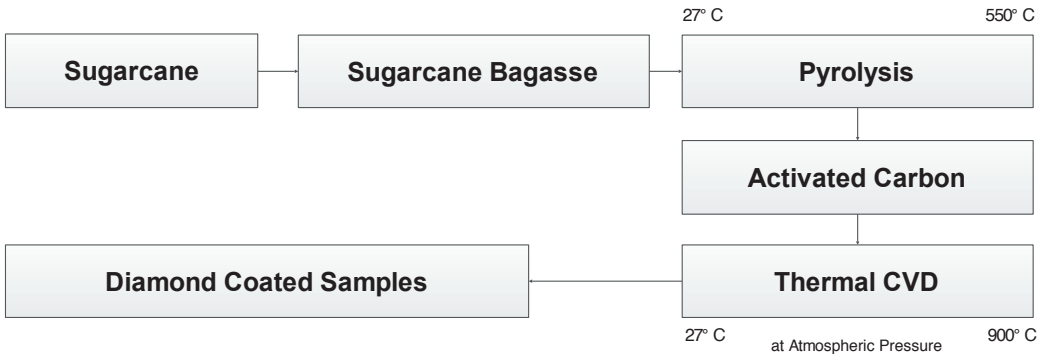


Figure 1. Flow chart of the deposition process for coating diamond like carbon on the samples (Tyagi et al., 2019, Krishnia and Tyagi, 2018, Krishnia et al., 2018)

CHARACTERIZATION

FESEM image of the worn out diamond like carbon coated sample is shown in Figure 2. Raman spectroscopy method was used to affirm the growth of the diamond like carbon coating and the graph of the same is shown in Figure 3. The obtained Raman spectra depicted that D and G band lies at 1337 cm^{-1} and 1575.4 cm^{-1} respectively. An upshift was observed on the peak at 5 cm^{-1}, because of the developed stresses on the coating (Sein et al., 2003, Krishnia and Tyagi, 2018, Krishnia et al., 2018). The SEM pictures shown in Figure 11 are in correlation with above results. Micro-hardness test were performed on the on the coated and un-coated samples. After several measuring processes the measured Vickers Hardness number were found to be 1455.24 HV and 950.65 HV for both coated as well as un-coated samples respectively. These same has been plotted as a graph in Figure 10 (Sarangi et al., 2008, Ghadai et al., 2020). FESEM machine was used for studying the growth of the diamond like carbon coating. The quality of the coating on the sample is mostly similar on the developed surface. The existence of the diamond like carbon coating is shown in Figure 11 (Rana et al., 2020, Knight and White, 1989).

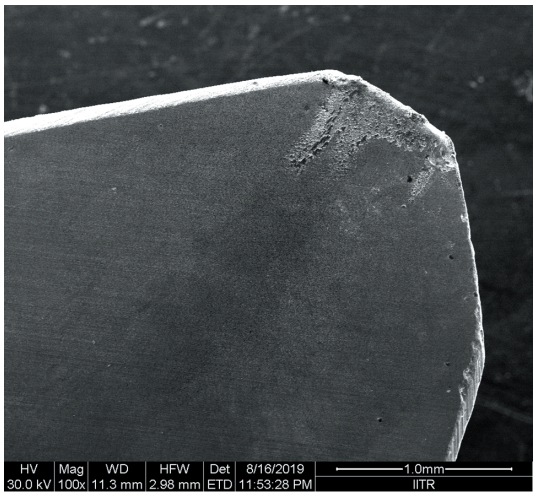


Figure 2. Fe-SEM image of the tool wear of the diamond like carbon coated sample.

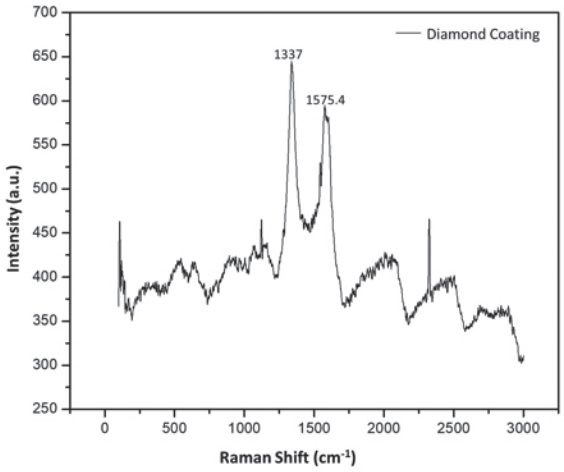


Figure 3. Raman spectra of the diamond like carbon coated sample.

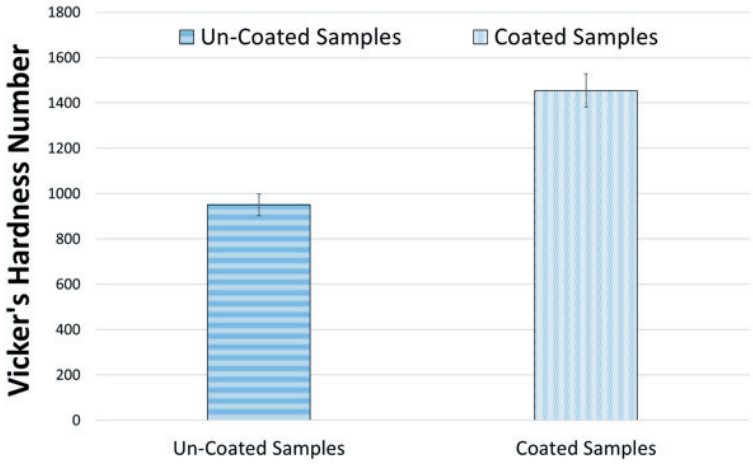


Figure 4. Graph of the Vickers hardness numbers of coated and un-coated samples.

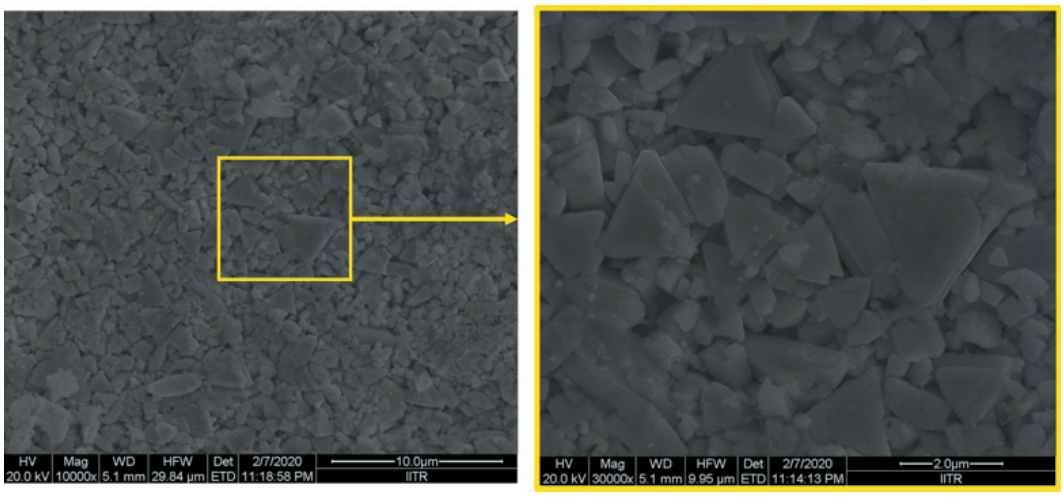


Figure 5. FESEM images depicting the Existence of coating.

RESULTS AND DISCUSSION

Selection of Orthogonal Array

With the aim to find the optimal combination of the effect of input parameters viz. cutting depth of cut (DOC), speed (V_c) and feed (*f*) on the output variables like resultant cutting temperature, surface roughness and cutting force. The orthogonal array (OA) was selected in such a way that we have least number of experimental runs. In this paper, we have used the L9 orthogonal array shown in Table 1.

Entropy Weight Method of Topsis

A statistical tool named entropy weight method was used to enlarge the distance from a negative ideal solution. In addition this statistical tool uses the optimum result from the ideal solution (Shanian and Savadogo, 2006).

Decision Matrix

The decision matrix is evaluated for all the values of responses of the variables [42].

An initial decision matrix, consisting of the response variables is also shown in Table 1.

$$A = (a_{ij})_{mxn} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad (1)$$

Table 1. Used orthogonal array (OA) along with the initial decision matrices.

Order of Exp.	Experiment No.	Parameters of Input			Variables of Response		
		DOC	Vc	f	F	Tc	Ra
8	1	0.3	60	0.15	85.49	285.50	0.44
5	2	0.1	60	0.1	51.49	139.70	0.36
1	3	0.1	90	0.15	62.06	154.00	0.37
9	4	0.1	30	0.05	47.84	144.30	0.34
6	5	0.5	60	0.05	109.63	447.40	0.52
7	6	0.3	30	0.1	78.47	277.50	0.44
4	7	0.5	30	0.15	101.18	589.90	0.46
2	8	0.3	90	0.05	88.44	277.90	0.53
3	9	0.5	90	0.1	127.96	423.30	0.66

Standardization

For having uniform units for all the selected responses we have to standardization of the decision matrix (Tiwari et al., 2019). We used “lower the better” as a criteria for the responses. The response variables are shown in Table 2.

a_{ij} = the response variables “lower the better”.

$$(r_{ij})` = a_{ij} / max\ (a_{ij}),\ i=1,2,3….m;j = 1,2,3….,n \quad (2)$$

$$(r_{ij})` = (r_{ij})` / \sum_{i=1}^{m}(\mathbf{r_{ij}})`,\ i=1,2,3….m;j = 1,2,3….,n \quad (3)$$

Method of Entropy

We calculated the weight for performance indicator using this technique. The calculations for the performance indicators are:

$$e = -k\sum_{i}^{m} \mathrm{r}_{ij\, lnr_{ij}}, j = 1,2,\ldots..,n\ and \tag{4}$$

$$k = \frac{1}{\ln(m)} \tag{5}$$

$$E = \sum_{j=1}^{n} e - j = -\frac{1}{\ln(m)} * \sum_{j=1}^{n} \quad \sum_{i=1}^{m} \mathrm{r}_{ij}\, ln(r_{ij}) \tag{6}$$

$$d_j = |1 - e_j|,\ j = 1,2,3,\ldots n \tag{7}$$

Table 2. Response variables in the form of obtained standardized decision matrix

Run No.	Variables of Response		
	F	Tc	Ra
1	0.3261	0.2817	0.3111
2	0.1964	0.1378	0.2567
3	0.2368	0.1520	0.2631
4	0.1825	0.1424	0.2460
5	0.4182	0.4415	0.3704
6	0.2994	0.2738	0.3132
7	0.3860	0.5821	0.3275
8	0.3374	0.2742	0.3790
9	0.4882	0.4177	0.4705

Whensoever d_j will be greater, we propagate r_{ij}, which in turns will make performance indicator the important one. Calculations are:

$$w_j = d_j / \sum_{j=1}^{n} d_j \tag{8}$$

Table 3 is showing the weight of evaluated responses.

Weight Determination

w_j = Calculated weight of the evaluated responses,
R = Standardized decision matrix = $(r_{ij})_{m\times n}$,
Following matrix was used to calculate the response variables:

$$v = \begin{bmatrix} w_1 r_{11} & w_2 r_{12} \cdots & w_n r_{1n} \\ \vdots & \ddots & \vdots \\ w_1 r_{m1} & w_2 r_{m2} \cdots & w_n r_{mn} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} \cdots & v_{mn} \end{bmatrix} \quad (9)$$

Table 4 is showing the Weighted and standardized decision matrices of Performance indicators.

Table 3. Weights of the performance indicators.

F	Tc	Ra
0.3481	0.2739	0.3781

Table 4. Weighted and standardized decision matrices of Performance indicators.

Run No.	F	Tc	Ra
1	0.1135	0.0771	0.1176
2	0.0684	0.0377	0.0971
3	0.0824	0.0416	0.0995
4	0.0635	0.0390	0.0930
5	0.1456	0.1209	0.1400
6	0.1042	0.0750	0.1184
7	0.1344	0.1594	0.1238
8	0.1174	0.0751	0.1433
9	0.1699	0.1144	0.1779

Positive and Negative Ideal Solutions:

The formula for finding the ideal solution is:

$$V^+ = \{v_1^+, v_2^+, \cdots\cdots, v_n^+\} = \{(max\ v_{ij}|j\varepsilon J_1), (min\ v_{ij}|j\varepsilon J_2)|\mathrm{i} = 1,2, \cdots\cdots, \mathrm{m}\} \quad (10)$$

$$V^- = \{v_1^-, v_2^-, \cdots\cdots, v_n^-\} = \{(min\ v_{ij}|j\varepsilon J_1), (max\ v_{ij}|j\varepsilon J_2)|\mathrm{i} = 1,2, \cdots\cdots, \mathrm{m}\} \quad (12)$$

Relative Distance

$$S_i^+ = \sqrt{\sum_{i=1}^{n}\left(v_{ij} - v_j^+\right)^2}, i = 1, 2, 3, \ldots\ldots, m \quad (13)$$

$$S_i^- = \sqrt{\sum_{i=1}^{n}\left(v_{ij} - v_j^-\right)^2}, i = 1, 2, 3, \ldots\ldots, m \quad (14)$$

Table 5 is showing the , and calculated "C_j".

(15)

Table 5. Weighted and standardized performance indicators.

Run No.	S_i^+	S_i^-	C_j	Rank
1	0.0682	0.1165	0.6307	5
2	0.0063	0.1779	0.9657	2
3	0.0203	0.1664	0.8911	3
4	0.0012	0.1817	0.9932	1
5	0.1259	0.0592	0.3199	7
6	0.0607	0.1224	0.6684	4
7	0.1441	0.0647	0.3099	8
8	0.0826	0.1052	0.5600	6
9	0.1562	0.0450	0.2237	9

The ranks obtained from the table 5 shows that the experiment run no. 4 produces the best optimum solution. The TOPSIS (technique for order of preference by similarity to ideal solution) method helped us in identifying the optimal parameter for the machining process while using the diamond like carbon coated samples.

CONCLUSION

The hardness of the coated sampled increased by 53% in as comparison to the uncoated samples. In the Raman Spectra, we obtained the peak of D band at the position of 1337 cm^{-1} which supported the presence of diamond on the coated samples. These results were in correlation with the obtained FESEM images. For optimizing the process parameters of machining processes for diamond like carbon coated WC samples, using TOPSIS technique was the main aim of this research. Optimum values of process parameters are 0.05 mm/rev, 0.1 mm and 30 m/min for feed rate, depth of cut and cutting speed respectively. The TOPSIS (technique for order of preference by similarity to ideal

solution) method has been successfully applied as a computational technique for optimizing the process parameters for the turning process using the developed diamond like carbon coated samples.

ACKNOWLEDGEMENT

Authors would like to thank Indian Institute Technology, Delhi, India and Indian Institute Technology, Roorkee, India for allowing us to use their facilities for material characterization. Authors also extend their regards to thank Precision Manufacturing Lab and Centre for Advanced Production and Industrial Engineering Research (CAPIER) of Delhi Technological University, New Delhi, India, for providing the facilities for the completion of this work.

REFERENCES

Asmussen, J. and Reinhard, D. eds., 2002. Diamond films handbook. CRC Press.

Bansal S., Saraf A., Rana R. and Lal R., 2021. Effect of Picosecond Laser Texture Surface on Tribological Properties on High-Chromium Steel Under Non-lubricated Conditions. Advances in Manufacturing and Industrial Engineering, pp. 257-267. Springer, Singapore.

Bobylyov, E.E. and Storozhenko, I.D., 2019. December. Diffusion metallization of carbide cutting tools as a way to improve the surface treatment quality. In Journal of Physics: Conference Series (Vol. 1399, No. 4, p. 044084). IOP Publishing.

Boyer, R.R., Cotton, J.D., Mohaghegh, M. and Schafrik, R.E., 2015. Materials considerations for aerospace applications. MRS Bulletin, 40(12): 1055-1066.

Chandran, M. and Hoffman, A., 2016. Diamond film deposition on WC–Co and steel substrates with a CrN interlayer for tribological applications. Journal of Physics D: Applied Physics, 49(21): 213002.

Chandran, M., Kumaran, C.R., Gowthama, S., Shanmugam, P., Natarajan, R., Bhattacharya, S.S. and Rao, M.R., 2013. Chemical vapor deposition of diamond coatings on tungsten carbide (WC–Co) riveting inserts. International Journal of Refractory Metals and Hard Materials, 37, pp.117-120.

Fraga, M.A., Contin, A., Rodríguez, L.A.A., Vieira, J., Campos, R.A., Corat, E.J. and Airoldi, V.T., 2016. Nano- and microcrystalline diamond deposition on pretreated WC–Co substrates: structural properties and adhesion. Materials Research Express, 3(2): 025601.

Ghadai, R.K., Das, S., Kalita, K., Swain, B.P. and Davim, J.P., 2020. Structural and Mechanical Analysis of APCVD Deposited Diamond-Like Carbon Thin Films. Silicon, pp.1-10.

Gracio, J.J., Fan, Q.H. and Madaleno, J.C., 2010. Diamond growth by chemical vapour deposition. Journal of Physics D: Applied Physics, 43(37): 374017.

Haubner, R. and Lux, B., 1996. On the formation of diamond coatings on WC/Co hard metal tools. International Journal of Refractory Metals and Hard Materials, 14(1-3): 111-118.

Jeyapandiarajan, P. and Xavior, A., 2019. Influence of cutting condition on machinability aspects of Inconel 718. Journal of Engineering Research, 7(2): 315-332.

Knight, D.S. and White, W.B., 1989. Characterization of diamond films by Raman spectroscopy. Journal of Materials Research, 4(2), pp.385-393.

Krishnia, L. and Tyagi, P.K., 2018. Growth and characterization of polycrystalline diamond films on silicon using sugarcane bagasse as carbon precursor at atmospheric pressure by thermal chemical vapor deposition. Diamond and Related Materials, 87: 18-26.

Krishnia, L., Yadav, B.S., Palnitkar, U., Satyam, P.V., Gupta, B.K., Koratkar, N.A. and Tyagi, P.K., 2018. As-pyrolyzed sugarcane bagasse possessing exotic field emission properties. Applied Surface Science, 443: 184-190.

Lima, S.A., Varum, H., Sales, A. and Neto, V.F., 2012. Analysis of the mechanical properties of compressed earth block masonry using the sugarcane bagasse ash. Construction and Building Materials, 35: 829-837.

Madurwar, M.V., Mandavgane, S.A. and Ralegaonkar, R.V., 2015. Development and feasibility analysis of bagasse ash bricks. Journal of Energy Engineering, 141(3): 04014022.

Najar, K.A., Sheikh, N.A., Butt, M.M. and Shah, M.A., 2018. Enhancing the wear resistance of WC–Co cutting inserts using synthetic diamond coatings. Industrial Lubrication and Tribology, 70 (7): 1224-1233.

Polini, R., Barletta, M., Rubino, G. and Vesco, S., 2012. Recent advances in the deposition of diamond coatings on Co-cemented tungsten carbides. Advances in Materials Science and Engineering, 2012: 151629.

Rana, R., Murtaza, Q. and Walia, R.S., 2020. GA based optimization of tri-bological behaviour of diamond coated tungsten carbide. World Journal of Engineering, vol. 17, no. 3, pp. 335–346.

Rana, R., Murtaza, Q. and Walia, R.S., 2021a. Optimization using genetic algorithm of tribological behaviour of WC tool material. Indian Journal of Engineering and Materials Sciences (IJEMS), 27(4): 889-896.

Rana, R., Rajput, K., Saini, R. and Lal, R., 2014. Optimization of tool wear: a review. Int J Mod Eng Res, 4(11): 35-42.

Rana, R., Walia, R. S., & Murtaza, Q., 2021b. Characterization and Parametric Optimization of Performance Parameters of DLC-Coated Tungsten Carbide (WC) Tool Using TOPSIS. Coatings, 11(7), 760.

Sarangi, S.K., Chattopadhyay, A. and Chattopadhyay, A.K., 2008. Effect of pretreatment methods and chamber pressure on morphology, quality and adhesion of HFCVD diamond coating on cemented carbide inserts. Applied Surface Science, 254(13): 3721-3733.

Schneider, A., Steinmueller-Nethl, D., Roy, M. and Franek, F., 2010. Enhanced tribological performances of nanocrystalline diamond film. International Journal of Refractory Metals and Hard Materials, 28(1): 40-50.

Sein, H., Ahmed, W., Rego, C.A., Jones, A.N., Amar, M., Jackson, M. and Polini, R., 2003. Chemical vapour deposition diamond coating on tungsten carbide dental cutting tools. Journal of Physics: Condensed Matter, 15(39): 2961-2967.

Shanian, A. and Savadogo, O., 2006. TOPSIS multiple-criteria decision support analysis for material selection of metallic bipolar plates for polymer electrolyte fuel cell. Journal of Power Sources, 159(2), pp.1095-1104.

Srinivasan, B., Rao, M.R. and Rao, B.C., 2016. On the development of a dual-layered diamond-coated tool for the effective machining of titanium Ti-6Al-4V alloy. Journal of Physics D: Applied Physics, 50(1): 015302.

Tanaka, T. and Akasawa, T., 1999. Machinability of hypereutectic silicon-aluminum alloys. Journal of Materials Engineering and Performance, 8(4): 463-468.

Tiwari, D., Sherwani, A.F., Muqeem, M. and Goyal, A., 2019. Parametric optimization of organic Rankine cycle using TOPSIS integrated with entropy weight method. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, pp.1-18.

Tyagi, A., Walia, R.S. and Murtaza, Q., 2019. Tribological behavior of temperature dependent environment friendly thermal CVD diamond coating. Diamond and Related Materials, 96: 148-159.

Ullen, N.B., Hasak, S.M.A. and Dirikolu, M.H., 2020. Factors influencing the machinability during turning sinter-hardened Cu-Ni-Mo based steel: Dependency on cutting speed, feed rate and cutting depth. Journal of Engineering Research, 8(4): 236-257.

Urakami, N., Kosaka, M. and Hashimoto, Y., 2018. Thermal chemical vapor deposition and luminescence property of graphitic carbon nitride film for carbon-based semiconductor systems. Japanese Journal of Applied Physics, 58(1): 010907.

View publication stats

People Lie, Actions Don't! Modeling Infodemic Proliferation Predictors among Social Media Users

Chahat Raj[1], Priyanka Meel[2]
[1]chahatraj58@gmail.com, [2]priyankameel86@gmail.com
Department of Information Technology, Delhi Technological University, India

Abstract: Social media is interactive, and interaction brings misinformation. With the growing amount of user-generated data, fake news on online platforms has become much frequent since the arrival of social networks. Now and then, an event occurs and becomes the topic of discussion, generating and propagating false information. Existing literature studying fake news primarily elaborates on fake news classification models. Approaches exploring fake news characteristics and ways to distinguish it from real news are minimal. Not many researches have focused on statistical testing and generating new factor discoveries. This study assumes fourteen hypotheses to identify factors exhibiting a relationship with fake news. We perform the experiments on two real-world COVID-19 datasets using qualitative and quantitative testing methods. This study concludes that sentiment polarity and gender can significantly identify fake news. Dependence on the presence of visual media is, however, inconclusive. Additionally, Twitter-specific factors like followers count, friends count, and retweet count significantly differ in fake and real news. Though, the contribution of status count and favorites count is disputed. This study identifies practical factors to be conjunctly utilized in the development of fake news detection algorithms.

Keywords: Factor identification, Fake News, COVID-19, Misinformation, Infodemic, Modeling predictors

1 Introduction

COVID-19 spread worldwide even faster than a human brain could imagine. Humans hardly even heard about it than before it turned to be the most fatal. After facing the catastrophic results only, many people became aware of it and started to ponder it. Talks about COVID-19 were everywhere and on everybody's minds and lips. Interactions about the hot topic have overwhelmed social networking platforms. Social media has now established its feet to feed information to people in the easiest way. The internet has been flooded with various types of information. But not everything that is on the internet is not reliable. Information that roams around on social media has not been validated and is merely people's ideas. Gradually these talks turned to be all sorts of fake news. With the feasibility of posting, sharing, and accessing the information on the web, its users can be quickly confounded with fake news. Fake news consists of every type of misinformation and disinformation. From the desks of politicians and public figures made the maiden attempt in spreading fake news worldwide, misleading people at large. It was the result of fake news that 5G towers in the UK turned into ruins. Fake news oozed out deadly political, social, religious, technological, environmental changes around the globe. It generated a sense of distrust among the people of the world. Enmity started grasping its enclosures. People claimed China to be the most causative element in spreading the disease. Detection of all sorts of talks that tend to be getting converted as fake news was the greatest need to lead the world into another mass destruction-like situation.

Fake news about the pandemic sprawled amongst various dimensions of society. One of these is the claiming of the remedial part of COVID-19. Enormous remedial approaches and suggestions started their part to play in contributing to fake news. "A pinch of turmeric or a drop of garlic juice could cure the fatal" was amongst the most prevailing unauthentic fake remedies. Poor

perceptions, unproven methods, illogical claims, false figures, and alarming news overwhelmed the global information scenario. Social media platforms are well known for the spread of misinformation and denial of scientific literature [1]. False social media posts have also tricked users into relying on harmful and poisonous substances like weed, cannabis, and ethanol intake [2]. The rapid evolution of the COVID-19 pandemic has not permitted immediate and specific scientific data [3]. COVID-19 is not the only fake news generating event. In the past, there have been many instances that led to colossal misinformation spread on online social networks, such as the 2016 US presidential elections, Pizzagate, hurricane Harvey, etc. [4]. COVID-19, whereas, is one major event generating misinformation on a scale larger than any other events. This led the World Health Organization into coining the term "Infodemic," referring to the mass propagation of false news revolving around the pandemic.

Previous research has contributed variously to solving the fake news problem. Researchers from behavioral sciences have covered the factors involved in sharing and accepting fake news [5, 6, 7]. Others have investigated several factors like user demographics and background information [8]. Many studies have developed fake news detection algorithms [9, 10]. Such algorithms widely utilize news content, such as linguistic features, visual features, and network features. However, there is an absence of ideal classifiers, and most of the fake news characteristics are unidentified. In this paper, we identify several key factors associated with fake and real news on Twitter. We formulate fourteen hypotheses on the key elements and their direct and mediating relationship with fake news. These hypotheses are evaluated on two real-world datasets which contain tweets about the COVID-19 pandemic. MediaEval 2020 [11] is a benchmark dataset containing tweets pertaining to coronavirus and 5G conspiracy. CovidHeRA [12] is a collection of tweets associated

with spreading health-related misinformation amidst the pandemic. The contribution of this paper is the analysis of characteristics that differentiate between fake and real news. We identify the following key factors: sentiment polarity, gender, media usage, follower count, friends count, status count, retweet count, and favorites count. Interdependence of factors like sentiment polarity, gender, and media usage are studied intensely. The relationship between fake news and these factors has not been studied in past research. We also extend the work of Parikh et al. [13] by demonstrating the relationship between fake news and particular sentiment polarities. This paper comes up with exciting outcomes suggesting important features demonstrating fake news dependence. The research bridges existing gaps in the literature and forms the basis for a new direction in fake news analysis. Our hypotheses shall be helpful in developing efficient fake news detection algorithms covering a wide range of fake news components.

The organization of this paper is as follows: Section 2 studies the existing literature in fake news and COVID-19 Infodemic. The survey provides insights into existing hypotheses and conclusions drawn upon fake news. Section 3 presents the research methodology explaining the datasets used and characteristics assumed for this study. Section 4 covers the results obtained by performing statistical tests on the datasets. Section 5 describes the insights drawn from the results and summarizes the acceptance/rejection of formulated hypotheses. Section 6 concludes the paper by discussing future directions.

2 Literature Review

This section discusses the progress in fake news and hypothesis domain presented by fellow researchers so far.

Fake News: The menace of fake news has been a challenging problem for information consumers. It has constantly been a topic of concern in the research society. Various studies have discussed the identification and detection of fake news on online social networks [14]. Past studies have focused on a vast dimension of fake news ranging between its origin, propagation, consumption, and impact [15]. In the recent era, various solutions have been proposed to detect fake news by the help of exploiting its textual [16], visual [17], and nodal features [18]. In contrast, studies pertaining to hypothesis formulation and testing are very few. There is limited literature available discussing the latest trends in online social networks highlighting vulnerabilities in fake news propagation and consumption. It is essential to formulate and discover dependent dimensions of fake news. Some studies have proposed important insights beneficial for fake news detection. For instance, Parikh et al. proposed hypotheses discussing the origin, proliferation, and tone of fake information [13]. They concluded that such misleading information is published more on lesser-known websites than the popular ones. In terms of proliferation or sharing, unverified users are more often shared on social media than by verified accounts. They also demonstrated that fake news has a specific tone or sentiment (positive, negative, or neutral) but did not conclude which type of particular tone is fake news mostly related to. Their study provides ways to form additional hypotheses, which is also a motivation for our work. Demographics and culture form the basis of theories proposed by Rampersad and Althiyabi [8]. They identified the established relationships between age and acceptance of fake news. It was noted that other demographics like gender and education played a more minor role in fake news acceptance. Another notable hypothesis confirmed that educated people are less likely to accept fake news. It was also observed that culture indirectly impacts the acceptance of fake news significantly. Works have highlighted the

connection between Third Person Effect (TPE) and fake news sharing [19, 20]. Brewer et al. have drawn several conclusions towards readers' reactions to consuming fake news [21]. Horne et al. have distinguished between real and fake news based on stylistic and physiological features of the text [22]. In another work by Silverman and SingerVine, it was identified that 75% of the US adults accepted fake news as true [23]. Similarly, Bovet and Makse studied the fake news propagation on Twitter during the 2016 US presidential elections and explored its influence [24]. Altay et al. hypothesized the relation between users' reputation and fake news sharing [5]. They studied that very few people were indulged in sharing fake news and identified the causes of such behavior. They arrived at the conclusion that sharing fake news harmed people's reputations and resulted in trust issues, which is a significant reason for very few people being indulged in sharing fake news. Osatuyi and Hughes figured that the amount of information available on fake news platforms is lesser than real news [25]. Exploring the role of comments in identifying and rejecting fake news shows that users are less likely to accept fake news if they come across critical comments about the content [26].

Infodemic: With the outbreak of the COVID-19 pandemic, social media communication and interactions rose at a level greater than before. Global concerns about the disease brought the world together to share information on online social networks. Such large-scale propagation gave rise to a phenomenon- "Infodemic." In an early response, researchers approached this problem by analyzing various concerns and suggesting solutions to the issue. Moscadelli et al. [27] have investigated the topics about the pandemic most polluted with fake news. Calvillo et al. [28] have analyzed political associations with the discerning of fake news. Hypotheses linking the fake news belief structure to its acceptance, Kim and Kim [29] proposed that factors like source credibility,

quality of information, receiver's ability, perceived benefit, trust, and knowledge decrease people's belief in fake news. Contrastingly, heuristic information, perceived risk, and stigma strengthen the confidence in fake news. Greene and Murphy [30] have discussed the likeliness of people sharing true or false stories on social media, establishing the association with their knowledge concerns. Another study that links conscience and ideology with infodemic sharing behavior is provided by Lawson and Kakkar [31]. Montesi [32] spreads light on the nature of infodemic and suggests that the harm caused by fake news is not health-related but more of a moral sort. Society, politics, and society are identified as the dominant infodemic themes. Building constructs over the Third Person Effect (TPE), Lui and Huang [33] have facts regarding the susceptibility and perception of fake news in the pandemic era. Similarly, Laato et al. [34] discuss the factors such as information sharing, information overload, and cyberchondria aiding fake news propagation. Experimenting on a Nigerian sample, Sulaiman [35] proposed no relationship between information evaluation and fake news sharing. With many hypotheses, Alvi and Saraswat [36] explored connections amongst various heuristic and systematic factors such as Sharing Motivation, Social Media Fatigue, Feel Good Factor. Fear Of Missing Out. News Characteristics, Extraversion, Conscientiousness, Agreeableness, Neuroticism, Trust, and Openness. As observed from the existing literature, past studies revolve around identifying psychological and behavioral factors that demonstrate any relationship with fake news. There is a research gap in characterizing features that could aid in distinguishing false information from real and serve as contributing factors to build fake news detection algorithms.

3 Research Methodology

3.1 Data

This study uses two publicly available benchmark datasets, MediaEval 2020 [11] and CovidHeRA [12]. MediaEval 2020 issued a benchmark dataset for its fake news detection task. The dataset consists of 5842 tweets classified into three classes: 5G coronavirus conspiracy, other conspiracy, and non-conspiracy. The tweets contain real and false information revolving around the COVID-19 pandemic. For this study, we classify these tweets into two coarse classes, with non-conspiracy tweets as real and the remaining tweets as fake. CovidHeRA is another benchmark dataset containing false tweets related to coronavirus and health. These tweets are a collection of fake remedies, preventive measures, treatments, and other health-related information spread across Twitter amidst the pandemic. Originally, the datasets consisted of tweet ids. To procure various characteristics of the tweets, the python library Tweepy is utilized. This scraping results in providing various information of the tweet and user content. This contextual information forms the basis of this study. To obtain the gender information of Twitter users, a gender predictor algorithm by Sap et al. [37] is used. Sentiments on the dataset are extracted using Microsoft's Text Analytics service. Sentiment scores are returned as values in the range of 0.0 to 0.1. A score between 0.0 to 0.3 signifies negative, 0.3 to 0.7 represents neutral and 0.7 to 1.0 represents positive sentiment. For media usage, we utilize the 'extended_entities' column from the scraped datasets. Sizes of both the datasets pertaining to each category are provided in tables 1, 2, and 3.

Table 1: Count of fake and real items with gender as a category

	CovidHeRA			Mediaeval		
Label	Male	Female	Total	Male	Female	Total
Fake	1532	772	2304	929	837	1766
Real	42683	40104	82787	2011	2065	4076
Total	44215	40876	85091	2940	2902	5842

Table 2: Count of fake and real items with sentiment polarity as a category

	CovidHeRA				Mediaeval			
Label	Negative	Neutral	Positive	Total	Negative	Neutral	Positive	Total
Fake	1292	391	621	2304	1042	346	378	1766
Real	31004	24638	27145	82787	2320	690	1066	4076
Total	32296	25029	27766	85091	3362	1036	1444	5842

Table 3: Count of fake and real items with media usage as a category

	CovidHeRA			Mediaeval		
Label	With	W/o Media	Total	With	W/o Media	Total
Fake	150	2154	2304	289	1477	1766
Real	17700	65087	82787	791	3285	4076
Total	17850	67241	85091	1080	4762	5842

3.2 Research Hypotheses

To identify characteristics that distinguish fake news and real news and consequently identify fake news based on these characteristics, we have formed fourteen hypotheses based on the qualitative and quantitative variables present in the dataset. Past research to determine factors related to fake news is limited. To identify the dependence of social media misinformation, we identify and analyze eight key elements: sentiment polarity, gender, media usage, follower count, friends count, status count, retweet count, and favorite count. We assume that fake news characterization, propagation, and acceptance have a relationship with these factors, which can be consequently utilized in fake news detection. For a better understanding, each tweet labeled as fake/real in the datasets has specific characteristics mentioned above. It is crucial to examine which feasible aspects demonstrate a relationship with false tweets. We also aim to study if there are any significantly different factors between real and fake tweets. By establishing such relationships, we tend to describe certain features useful for real and fake tweet classification. As evident from the existing literature, very few features have been exploited by fake news detection algorithms. Now

examining the stated features, we propose to add more of such contributing characteristics. Qualitative hypotheses $H_{A,}$ $H_{B,}$ and $H_{C,}$ are tested to scrutinize the direct relationships between sentiment, gender, and media usage with fake news, respectively. Further, it is vital to analyze if the bias of one independent variable influences the bias of another independent variable. For example, to test whether or not it is the higher proportion of one categorical variable contributing to the higher proportion of another categorical variable. To do so, we construct six more qualitative hypotheses, H_D, H_E, H_F, $H_{G,}$ H_H, and H_I. These nine hypotheses are tested using the Chi-square test of independence. The relationship is demonstrated in figure 1. To study quantitative variables, we formulate hypotheses H_J to H_N and perform Analysis of Means on each one of, also and calculate intervals. Figure 2 demonstrates the quantitative relationships.

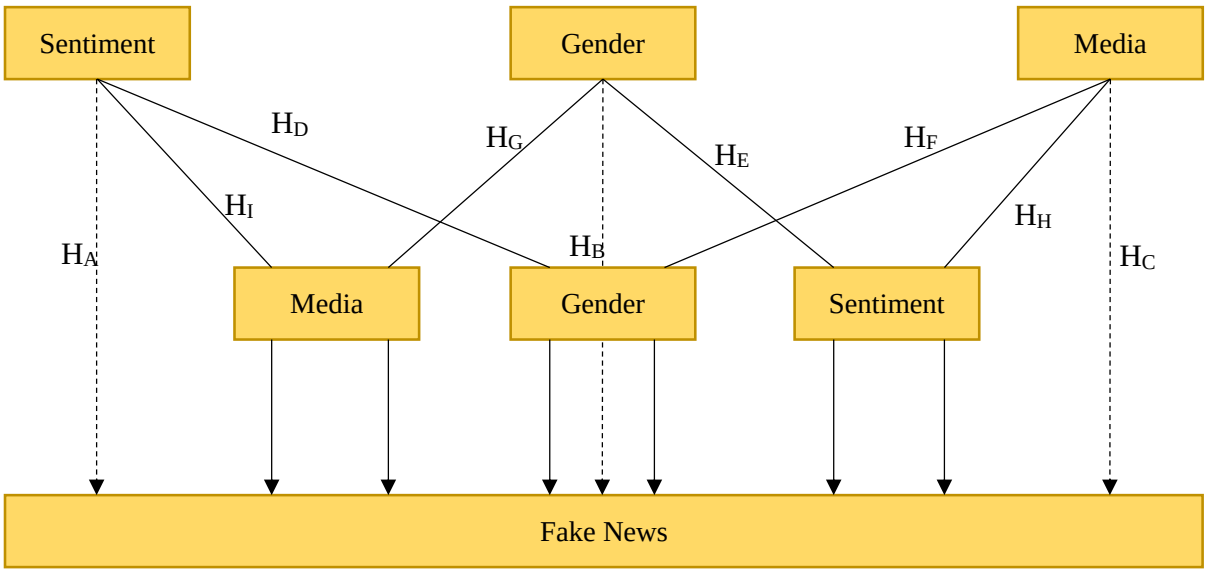


Figure 1: Factors determining fake news (qualitative hypotheses)

3.2.1 Qualitative Hypotheses and Factors

Sentiment: According to Parikh et al., it is widely assumed that most of the news spreading online is negative in terms of its linguistic tone. However, it has not been proven that fake news has a

higher negative polarity than neutral or positive polarities. Parikh et al. noted that it was inconclusive to say if fake news had a bias towards a particular polarity. Following their assumption, H_A forms the primary hypothesis to test if fake news has a tendency towards a specific sentiment polarity.

H_{A0}: There is no bias in the proportion of different sentiments between fake news and real news.
H_{A1}: There is a significant bias in the proportion of different sentiments between fake news and real news.

Gender: Rampersad and Althiyabi, examining a sample of Saudi Arabia, observed that gender has a weakly positive effect on the acceptance of fake news by people. The sample is specific to a particular demographic region. In our research, datasets consist of tweets from Twitter users across the globe. This helps to examine the assumptions on a universal scale. We test this hypothesis by using HB's statement to verify if there is a significant relationship between gender and false information.

H_{B0}: There is no bias of the gender of users involved in fake news with respect to real news.
H_{B1}: There is a significant bias of gender of users involved in fake news with respect to real news.

Media: Several fake news detection algorithms have been designed that detect whether a visual media in a piece of fake information is credible or not. We, hereby, analyze whether it can be stated solely based on the presence of visual media that a post/message is false. We categorize the datasets into two modalities: without and with visual media (pictures/videos). We try to analyze what data modality of social media posts contribute more/demonstrate bias towards misinformation using the statement H_C.

H_{C0}: There is no bias of media usage in fake news with respect to real news.
H_{C1}: There is a significant bias of media usage in fake news with respect to real news.

Based on the above three univariate hypotheses, we decide the mediating relationships among these factors and formulate multivariate hypotheses (H_D to H_I) to determine whether bias in one of the above proportions is due to bias in proportions of the other variable.

H_{D0}: There is no influence of bias in the proportion of a particular gender of the user on the bias in the proportion of sentiments in fake news with respect to real news.
H_{D1}: There is significant influence of bias in the proportion of a particular gender of the user on the bias in the proportion of sentiments in fake news with respect to real news.

H_{E0}: There is no bias in the proportion of a particular sentiment used in fake news between different gender of users.
H_{E1}: There is a significant bias in the proportion of a particular sentiment used in fake news between different gender of users.

H_{F0}: There is no bias in inducing a particular sentiment with media usage in fake news.
H_{F1}: There is a significant bias in inducing a particular sentiment with media usage in fake news.

H_{G0}: There is no bias in the usage of media amongst different sentiments used in fake news.
H_{G1}: There is a significant bias in media usage amongst different sentiments used in fake news.

H_{H0}: There is no relationship between a particular gender and media usage in fake news.
H_{H1}: There is a significant relationship between a particular gender and media usage in fake news.

H_{I0}: There is no bias in and media usage in fake news between different gender of users.
H_{I1}: There is a significant bias in and usage of media in fake news between different gender of users.

3.2.2 Quantitative Hypotheses and Factors

Using the data scraped from Twitter, we decided on testing our hypotheses on five key factors, which can be categorized into three user/profile-specific features, i.e., the number of followers, friends, and statuses and two post-specific features, i.e., retweets count and favorites

count. In our novel approach, we assume that these factors can be utilized in identifying the credibility of tweets, or in other words, labeling of tweets. Moreover, we assume that these factors impose an effect on fake news sharing and acceptance.

Followers and friends count determine the extent of reachability of a particular post or message within the user's social network who created it. A retweet is an action of sharing a particular tweet on one's timeline, which is done mainly by the follower of the user who created it and is visible to other Twitter users who turn the followers of the user who retweeted it. Retweet count determines the propagation and acceptance behavior of a fake post by checking the social reach. It is similar to the action "Share" on other social networks. It spreads a particular post to the user's social network. The larger the retweet count, the more likely the people reading the post will believe that particular piece of information and further spread it across the web. Status count corresponds to the number of total posts/retweets a specific user has posted since the creation of his account. Favorites are user markings made on a post a user would like to save for the future.

We determine the relationship between these quantitative variables and the label of the post, i.e., the relationship between the number of retweets and favorites of the post and the followers, friends, and status of the user who posted it, and it being real or fake. Since the source of misinformation can range from a random regular user to a credible account such as commercial news channels, journalists, or celebrities, it becomes difficult to assume any specific range for the count of these quantitative variables. Hence, we test based on a characteristic whether there is a significantly distinguishable bias in the values attributed to the mean and a confidence interval around it for each of these variables. In other words, the probability with which a post or a piece

of information under examination can be labeled as fake or real based on its values of the above-mentioned quantitative variables.

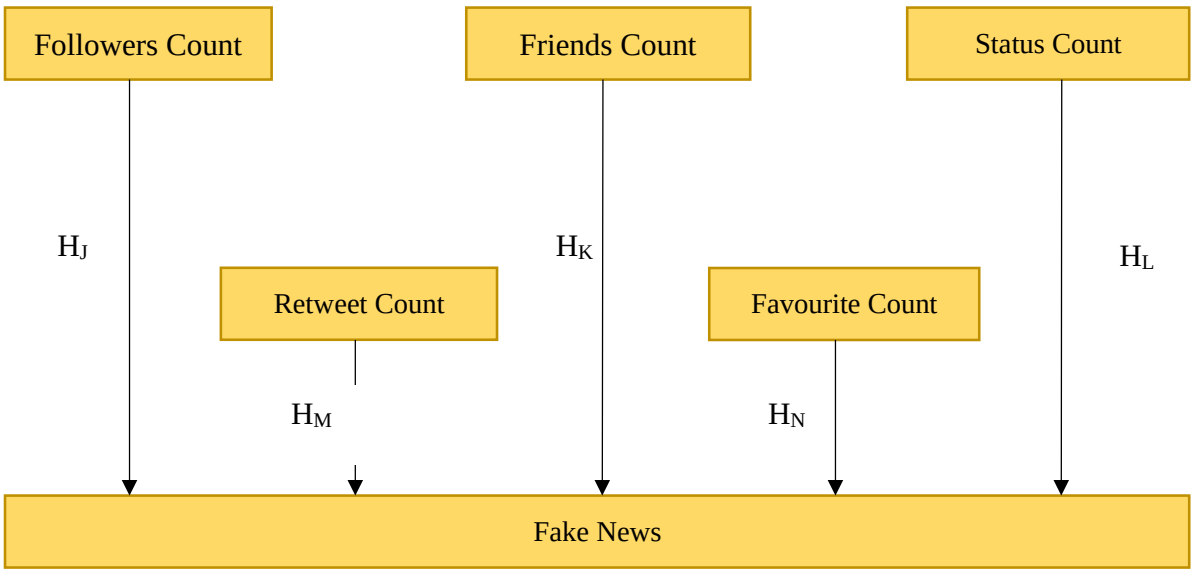


Figure 2: Factors determining fake news (quantitative hypotheses)

H_{J0}: There is no bias of follower count in fake news.
H_{J1}: There is a significantly distinguishable bias of follower count in fake news.

H_{K0}: There is no bias of friends count in fake news.
H_{K1}: There is a significantly distinguishable bias of friends count in fake news.

H_{L0}: There is no bias of status count in fake news.
H_{L1}: There is a significantly distinguishable bias of status count in fake news.

H_{M0}: There is no bias of retweet count in fake news.
H_{M1}: There is a significantly distinguishable bias of retweet count in fake news.

H_{N0}: There is no bias of favorite count in fake news.
H_{N1}: There is a significantly distinguishable bias of favorite count in fake news.

4 Results

To test on the nine hypotheses H_A to H_I, which are formed upon the categorical variables, we use the Chi-Square test of independence alongside computing "Cramer's V," "Pearson's r," and "spearman's rho" values. Cramer's V value provides us with the strength of association between the nominal categorical variables for the conclusion arrived using the Chi-Square test. Its values range between 0 and 1. Pearson's r value signifies both the strength of association and the direction of the association between two continuous variables. Here direction indicates if one variable would increase or decrease with respect to change in another variable. Its values range from -1 to +1, where the value of -1 means that as one variable increases, the other decreases, and +1 means that as one variable increases, the other increases too. A value of 0 indicates no strength of association. Spearman's rho values differ from the outcomes of Pearson's r values by a feature that they can describe the correlation even when the variables do not have a linear association. It is also proof from the long tail of outlier values as it uses the ranks of the values of the variable. The values in the table 7 include degrees of freedom as df, Chi-Square test value as χ^2, probability value as p-value and Cremer's V value, Pearson's r-value, and Spearman's rho. The first column in this table indicates the hypothesis to which the variables and their values belong to. From the first row of the same table, we observe that χ^2 values for testing hypothesis H_A with 2 degrees of freedom (df) for both CovidHeRA and MediaEval datasets are 352.963 and 17.103, respectively, and are more significant than critical value $\chi^2_c = 5.991$ with $p < 0.001$ (Significance level $\alpha = 0.05 = p_c$, critical p-value). This implies that there is a significant difference in proportions of sentiments used between Fake and Real news. But despite there being a substantial difference in ratios, low values of Cramer's V (less than 0.2), Pearson's r (between -0.20 and +0.20), and Spearman's rho (between -0.20 and +0.20) indicate weak association of label (news being fake or real) and the

sentiment (sentiment being negative or neutral or positive). These values (Cramer's V, Person's r, and Spearman's rho) are low for all the hypotheses tested. Therefore, we rely on comparing Actual values from tables 1,2 and 3 with Expected values in tables 4,5 and 6, respectively, to determine the association between an independent and a categorical dependent variable, or in other words, the bias of fake news towards a specific or a group of categorical variables. On comparing table 2 and table 5, we observe that in both CovidHeRA and MediaEval datasets, Fake news with Negative sentiment has a higher Actual count (1292, 1042) with respect to Expected count (874.5, 1016.3) and Fake news with Positive sentiment has less Actual count (621, 378) with respect to Expected count (751.8, 436.5). Count of Neutral sentiment varies inversely in both datasets, with CovidHeRA showing reduced count and MediaEval showing an Increase. Similarly, we observe from the same tables that the Actual count of Real news with Negative sentiment is less than that of the Expected count in both datasets. The Actual count of Real news with Positive sentiment is greater than that of the Expected count in both the datasets. Therefore, we reject the Null hypothesis (H_0) of H_A and observe that Fake news propagation during CoVID-19 has had a proportional bias towards Negative sentiment.

Table 4: Expected count of fake and real items with gender as a category

	CovidHeRA			Mediaeval		
Label	Male	Female	Total	Male	Female	Total
Fake	1197	1107	2304	888.7	877.3	1766
Real	43018	39769	82787	2051.3	2024.7	4076
Total	44215	40876	85091	2940	2902	5842

Table 5: Expected count of fake and real items with sentiment polarity as a category

	CovidHeRA				Mediaeval			
Label	Negative	Neutral	Positive	Total	Negative	Neutral	Positive	Total
Fake	874.5	677.7	751.8	2304	1016.3	313.2	436.5	1766
Real	31421.5	24351.3	27014.2	82787	2345.7	722.8	1007.5	4076

Total	32296	25029	27766	85091	3362	1036	1444	5842

Table 6: Expected count of fake and real items with media usage as a category

	CovidHeRA			Mediaeval		
Label	With	W/o Media	Total	With	W/o Media	Total
Fake	483	1821	2304	326.5	1439.5	1766
Real	17367	65420	82787	753.5	3322.5	4076
Total	17850	67241	85091	1080	4762	5842

From the second row of the table 7, we observe that χ^2 values for testing hypothesis H_B for both CovidHeRA and MediaEval datasets are 200.321 and 5.261, respectively, and are greater than critical value $\chi^2_c = 3.841$ with $p < 0.001$ and $p = 0.022$, respectively, both less than $\alpha = 0.05$. This implies a significant difference in proportions of the gender of users between Fake and Real news. By comparing Actual values with Expected values from table 1 and table 4 respectively, we observe that the Male gender has a greater Actual proportion in Fake news than the Expected proportion, and the Female gender has a higher Actual proportion involved in Real news than Expected Proportion, in both datasets. Therefore, we reject the Null hypothesis (H_0) for H_B and observe a significant bias in the gender of users involved in CoVID-19 Fake news propagation.

To test for Hypothesis H_C, from third row of table 7, we observe that χ^2 values for both CovidHeRA and MediaEval datasets are 298.995 and 7.765, respectively, and are greater than critical value $\chi^2_c = 3.841$ with $p < 0.001$ and $p = 0.006$ respectively, both less than $\alpha = 0.05$. For both the datasets, comparing the values of Actual and Expected Media usage from table 3 and table 6 respectively shows that Actual values for Fake news with media used is less than Expected values and the same is more in the case of Real news. Therefore, there is a significant difference in the proportion of Fake news and Real news propagation with media usage than the expected proportion, which leads us to reject the Null Hypothesis (H_0) for H_C.

The test for hypothesis H_D involves comparing values from row four and row five of the table 7. From row 4, the χ^2 values of Male gender from datasets CovidHeRA and MediaEval are 217.67 and 13.342, respectively, both higher than $\chi^2_c = 5.991$ and p values being $p < 0.001$ and $p = 0.001$, respectively, both less than $\alpha = 0.05$. From row 5, the χ^2 value for Female gender from CovidHeRA dataset is 169.979, greater than the critical value $\chi^2_c = 5.991$ and the value of $p < 0.001$ is less than $\alpha = 0.05$. But for the same gender in the MediaEval dataset, the χ^2 value turns out to be 5.503, which is less than $\chi^2_c = 5.991$, and the p-value of $p = 0.064 > \alpha = 0.05$ suggests contradictory inference from these two datasets. But since the MediaEval dataset gave both the χ^2 and p values close to their respective critical values for female gender, we reject Null Hypothesis (H_0) for H_D and conclude that there is a significant bias in proportion of sentiments used by users of both the gender and the bias in proportion of the user gender has no influence on the bias of proportion of sentiments.

Further, to identify towards which sentiment is the bias more by the users of both genders, we use results from rows six, seven, and eight of table 7 for testing hypothesis H_E. For the CovidHeRA dataset, the three rows mentioned above have χ^2 value of 78.005, 13.65, and 146.509 for negative, neutral, and positive sentiment, respectively, which are all greater than $\chi^2_c = 3.841$ and their respective p values being $p < 0.001$ for all three, is less than $\alpha = 0.05$. Results from this dataset do not indicate the specific sentiment towards which the bias is more. However, we can infer that there is a significant difference in the proportion of each sentiment when compared to real news. Observing results from these three rows for the MediaEval dataset, we obtain χ^2 value of 1.702, 4.82, and 0.411, for negative, neutral, and positive sentiments, respectively, where χ^2 values for Negative and Positive sentiments are both less than $\chi^2_c = 3.841$ and for Neutral sentiment, the χ^2

value is higher than χ^2_c. The p values for these corresponding χ^2 values are p = 0.192, p = 0.028 and p = 0.521, respectively. This shows no significant bias of the user gender on Negative and Positive sentiment as p values (0.192 and 0.521) obtained are greater than $\alpha = 0.05$. But for Neutral sentiment, we observe a bias as the p-value of 0.028 is less than $\alpha = 0.05$. Therefore, we reject the Null hypothesis (H_0) for H_E and conclude that Fake news is more biased towards being sentiment Neutral, followed by being sentiment Negative, and show no significant difference in proportions of Real news towards being sentiment Positive.

For testing Hypothesis, H_F, the bias of usage of media to induce a particular sentiment in the propagation of COVID-19 Fake news, from table 7, the values from rows nine, ten and eleven for CovidHeRA dataset indicate χ^2 values of 97.382, 59.615, and 124.61 for Negative, Neutral and Positive sentiment, respectively, with all of them being greater than $\chi^2_c = 5.991$ and with a p-value for each of them being $p < 0.001$, less than $\alpha = 0.05$ indicate rejection of Null Hypothesis (H_0) for H_F. For the MediaEval dataset, however, the χ^2 values of 0.235, 2.399, and 20.077 for Negative, Neutral and Positive sentiments, respectively, with the former two being less than $\chi^2_c = 3.841$ and the latter being more excellent, and their respective p values being p = 0.628, p = 0.121 and $p < 0.001$ indicate that only for Positive sentiment, there is a significant difference of proportion in the usage of media for Fake news with respect to real news. From the contradictory results from the two datasets for Negative and Neutral sentiments, we understand that there is a bias produced by usage of media for only positive sentiment. Hence, we reject the Null Hypothesis (H_0) for H_F.

	Datasets →	CovidHeRA						MediaEval					
(H)	variables ↓	df	χ^2	p value	V^2	r	rho	df	χ^2	p value	V^2	r	rho
H_A	Label vs Sentiment	2	352.963	p < 0.001	0.004	0.047	0.048	2	17.10	p < 0.001	0.003	0.037	0.032
H_B	Label vs Gender	1	200.321	p < 0.001	0.002	0.048	0.048	1	5.261	p = 0.022	0.001	0.03	0.03
H_C	Label vs Media usage	1	298.995	p < 0.001	0.003	0.059	0.059	1	7.565	p = 0.006	0.001	0.035	0.035
H_D	Label vs Sentiment (Gender - male)	2	217.67	p < 0.001	0.004	0.039	0.041	2	13.342	p = 0.001	0.004	0.035	0.027
H_D	Label vs Sentiment (Gender - female)	2	169.979	p < 0.001	0.004	0.058	0.058	2	5.503	p = 0.064	0.001	0.038	0.035
H_E	Label vs Gender (Sentiment - Negative)	1	78.005	p < 0.001	0.002	0.049	0.049	1	1.702	p = 0.192	0.001	0.023	0.023
H_E	Label vs Gender (Sentiment - Neutral)	1	13.65	p < 0.001	0.001	0.023	0.023	1	4.82	p = 0.028	0.004	0.068	0.068
H_E	Label vs Gender (Sentiment - Positive)	1	146.509	p < 0.001	0.005	0.072	0.072	1	0.411	p = 0.521	0	0.016	0.016
H_F	Label vs Media usage (Sentiment - Negative)	1	97.382	p < 0.001	0.003	0.055	0.055	1	0.235	p = 0.628	0	-0.008	-0.008
H_F	Label vs Media usage (Sentiment - Neutral)	1	59.615	p < 0.001	0.002	0.048	0.048	1	2.399	p = 0.121	0.002	0.048	0.048
H_F	Label vs Media usage (Sentiment - Positive)	1	124.61	p < 0.001	0.004	0.066	0.066	1	20.077	p < 0.001	0.013	0.117	0.117
H_G	Label vs Sentiment (Media not used)	2	267.346	p < 0.001	0.003	0.042	0.044	2	7.223	p = 0.027	0.001	0.01	0.005
H_G	Label vs Sentiment (Media used)	2	61.585	p < 0.001	0.003	0.045	0.042	2	25.154	p < 0.001	0.023	0.145	0.141
H_H	Label vs Gender (Media not used)	1	193.333	p < 0.001	0.003	0.053	0.053	1	5.561	p = 0.018	0.001	0.034	0.034
H_H	Label vs Gender (media used)	1	5.472	p = 0.019	0	0.017	0.017	1	0.345	p = 0.557	0	0.017	0.017
H_I	Label vs Media usage (Gender - male)	1	209.649	p < 0.001	0.005	0.069	0.069	1	5.664	p = 0.017	0.001	0.043	0.043
H_I	Label vs Media usage (Gender - female)	1	87.831	p < 0.001	0.002	0.046	0.046	1	2.529	p = 0.112	0.001	0.03	0.03

Table 7: Chi-square test on qualitative hypotheses

From rows twelve and thirteen of table 7, we test for Hypothesis H_G to observe a bias of proportion of sentiment caused when media is used and when it is not used, respectively. For CovidHeRA dataset, for with usage of media (row 12) and without the usage of media row (13), χ^2 values of 267.346 and 61.585, respectively, both less than $\chi^2_c = 5.991$ and their respective p values of $p < 0.001$ each for both being less than $\alpha = 0.05$, suggest that there is a difference in the proportion of sentiment used in Fake news with respect to Real news. Similar inference can be obtained from MediaEval dataset, in which, with the usage of media (row 12) and without the use of media row (13) have χ^2 values of 7.223 and 25.15, respectively, both less than $\chi^2_c = 5.991$ and their respective p values of $p = 0.027$ and $p < 0.001$, both being less than $\alpha = 0.05$. Hence, there is a bias induced in the proportions of sentiment in Fake news with respect to Real news by usage and non-usage of media, and therefore we reject the Null Hypothesis (H_0) for H_G.

Further, from rows fourteen and fifteen of table 7, we test for Hypothesis H_H to check for bias in proportion of gender of Fake news with respect to Real news is influenced by bias in usage of media. For CovidHeRA, we obtain χ^2 values of 193.333 and 5.472 for "media used" and "media not used", respectively, both greater than $\chi^2_c = 3.84$ with their respective p values being $p < 0.001$ and $p = 0.019$, both less than $\alpha = 0.05$. For the MediaEval dataset, for the same rows, we obtain χ^2 values of 5.561 and 0.345 and p values of $p = 0.018$ and $p = 0.557$ for "media used" and "media not used," respectively. We observe that for "media not used," the test shows the opposite result with that compared from CovidHeRA dataset, meaning that there is no difference in the proportion of user's gender when media is not used in Fake news propagation, with respect to Real news propagation. These contradictory results make Hypothesis H_H inconclusive.

Statistics	Followers		Friends		Retweets		Status		Favorites	
	Fake	**Real**	**Fake**	**Real**	**Fake**	**Real**	**Fake**	**Real**	**Fake**	**Real**
Mean (C)	5421.657	63656.21	3181.374	2293.652	154.132	628.718	56262.98	46189.4	2.238	7.766
Standard Error (C)	445.735	3847.024	149.150	35.658	38.207	17.616	2269.618	497.010	0.255	0.490
Median (C)	1742.5	21733	953	605	52	173	17180	9238	1	2
Mode (C)	706	7810	775	209	18	28	3145	1760	0	0
Standard Deviation (C)	21395.32	1106894	7159.233	10259.94	1833.94	5068.56	108941.7	143003.4	12.261	141.126
Sample Variance (C)	4.58E+08	1.23E+12	51254619	1.05E+08	3.36E+06	2.57E+07	1.19E+10	2.04E+10	150.344	19916.57
Count (C)	2304	82787	2304	82787	2304	82787	2304	82787	2304	82787
Confidence Level(95.0%) (C)	874.085	7540.138	292.483	69.890	74.886	34.527	4450.709	974.136	0.500	0.961
Mean (M)	23255.37	99511.34	3012.989	1999.394	260.701	644.781	38369.96	55846.16	679.669	2244.092
Standard Error (M)	9979.619	11302.57	302.646	159.182	71.410	61.498	1787.307	1744.372	201.229	224.780
Median (M)	4711.5	37160.5	733	609	60	155	12955	18305.5	48	292
Mode (M)	1180	12548	650	138	77	92	547	446	48	177
Standard Deviation (M)	419381.5	721596.4	12718.35	10162.77	3000.955	3926.28	75109.43	111366.9	8456.42	14350.77
Sample Variance (M)	1.76E+11	5.21E+11	1.62E+08	1.03E+08	9005729	15415676	5.64E+09	1.24E+10	71511039	2.06E+08
Count (M)	1766	4076	1766	4076	1766	4076	1766	4076	1766	4076
Confidence Level(95.0%) (M)	19560.05	22153.04	593.583	311.998	140.058	120.570	3505.461	3419.921	394.672	440.692

Table 8: Descriptive statistics of CovidHeRA(C) and MediaEval dataset(M)

From the values in rows 16 and 17 in table 7, for the CovidHeRA dataset, both genders show that there is a difference in the proportion of media used for Fake news propagation with respect to Real news. This can be observed as the χ^2 values of 209.649 and 87.831 for the male and female gender, respectively, are both greater than $\chi^2_c = 3.84$, and their respective p values, both $p < 0.001$ is more diminutive than $\alpha = 0.05$. In the MediaEval dataset, we observe from rows 16 and 17 of table 7 that while users of Male gender with χ^2 value of 5.664 and $p = 0.017$ show difference in the proportion of media used for Fake news with respect to Real news, but for Female gender, indifference in proportions of usage of media in Fake news with respect to real news is observed as the χ^2 value of 2.529 is less than $\chi^2_c = 3.84$ and its p-value of $p = 0.112$ is more remarkable than $\alpha = 0.05$. Therefore, for Hypothesis H_I, we cannot come to any conclusive decision.

For the Quantitative variables, we plot the data distribution around the mean with a 95% confidence interval. This will distinguish the central values of the variables and help us determine the strength of the distinguishment, i.e., the smaller the upper and lower bound distance from the mean, the more the reliance on these values representing the true mean value of the population. Table 8 provides descriptive statistics of both the datasets. From fig 3 and fig 4, for CovidHeRA and MediaEval dataset, respectively, we observe that users who propagated Fake news have a smaller number of followers than the users with Real news. The mean values for Fake news for these datasets are 5421.65 and 23255.37 in the order mentioned above. These are distinct from the mean number of followers of Real news 63656.21 and 99511.34 for the two datasets. We also observe that there is a significant bias in the number of followers of the users of Fake news and Real news as the range of 95% CI for mean do not overlap for Fake news and Real news and

therefore attributing a label to a piece of information on Twitter by comparing, the number of followers of the user who shared it with the mean range of these plots can be done more accurately.

From the plots of the number of Friends in fig 3 and fig 4 for CovidHeRA and MediaEval datasets, respectively, the previously mentioned inference becomes much more robust as not only the 95% CI bounds remain distinct for Fake news and real news, but also the closer proximity of the value of mean for a particular label in both datasets shows the repeatability of the trend. The mean value for Fake news in CovidHeRA and MediaEval dataset is 3181.374 and 3012.989, respectively, and the same for Real news in these datasets is 2293.652 and 1999.394, respectively. There is a significant bias in the mean number of friends for users who propagated Fake news compared to the number of friends of users who propagated real news.

The plots from fig 4 and fig 4 for the number of retweets have similar mean value for fake news and real news. For Fake news, the mean values of 154.132 and 260.701 for the two datasets, and Real news, the mean values are 628.718, and 644.781 show the closeness within the label and distinction between the labels. Therefore, this bias can prove helpful to label a piece of information based on its proximity to one of the mean values 95% CI interval.

For the “number of statuses” variable, the 95% CI interval for mean and the mean value for Fake news and Real news alternate between the two datasets. Therefore, we cannot come to any specific conclusion using the information of this variable of a particular information sample despite there being bias in the mean values between the Labels. The same conclusion can be drawn for “number of favorites” as the ranges in both the datasets are significantly different amongst the

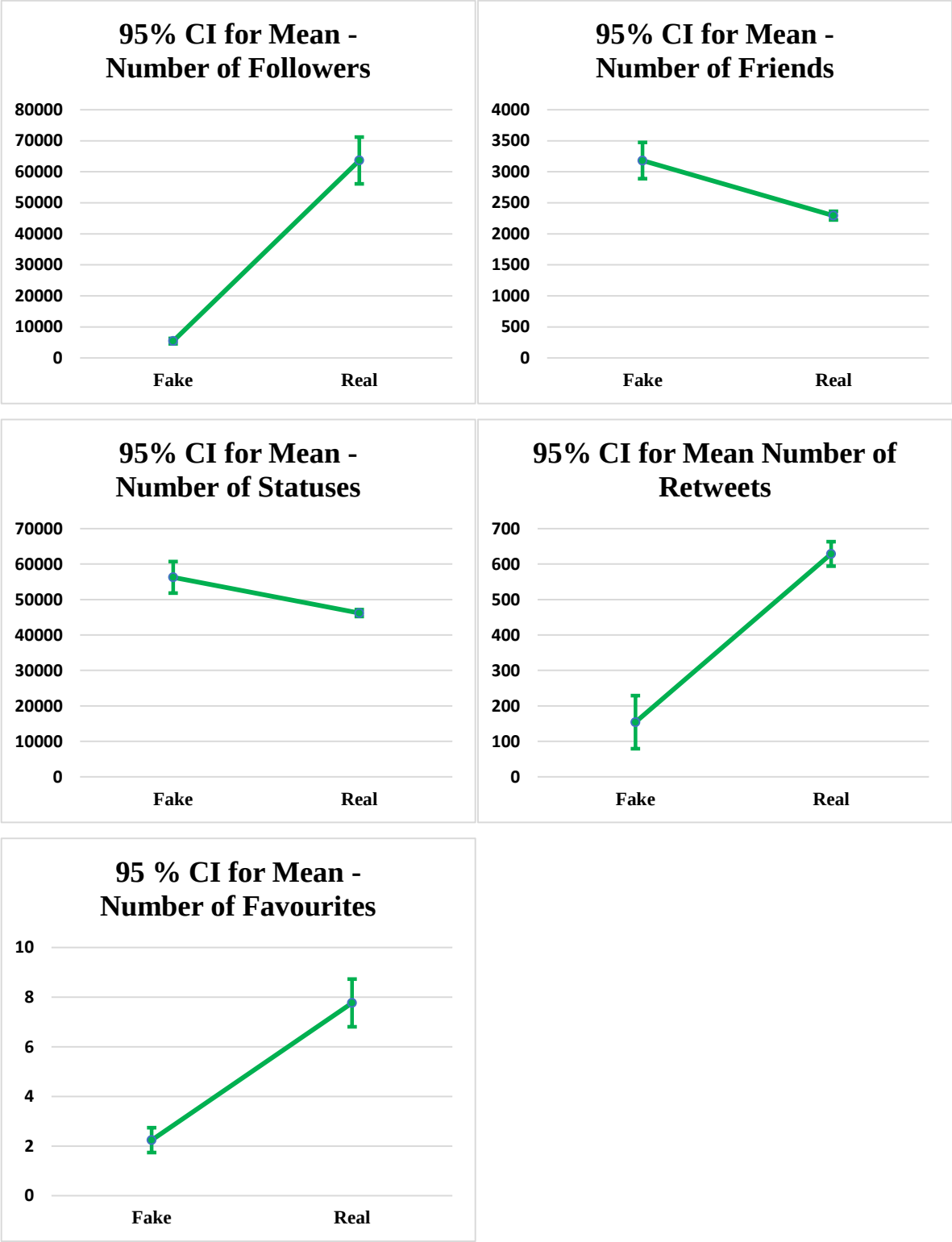


Figure 3: 95% Confidence Interval for quantitative factors on CovidHeRA dataset

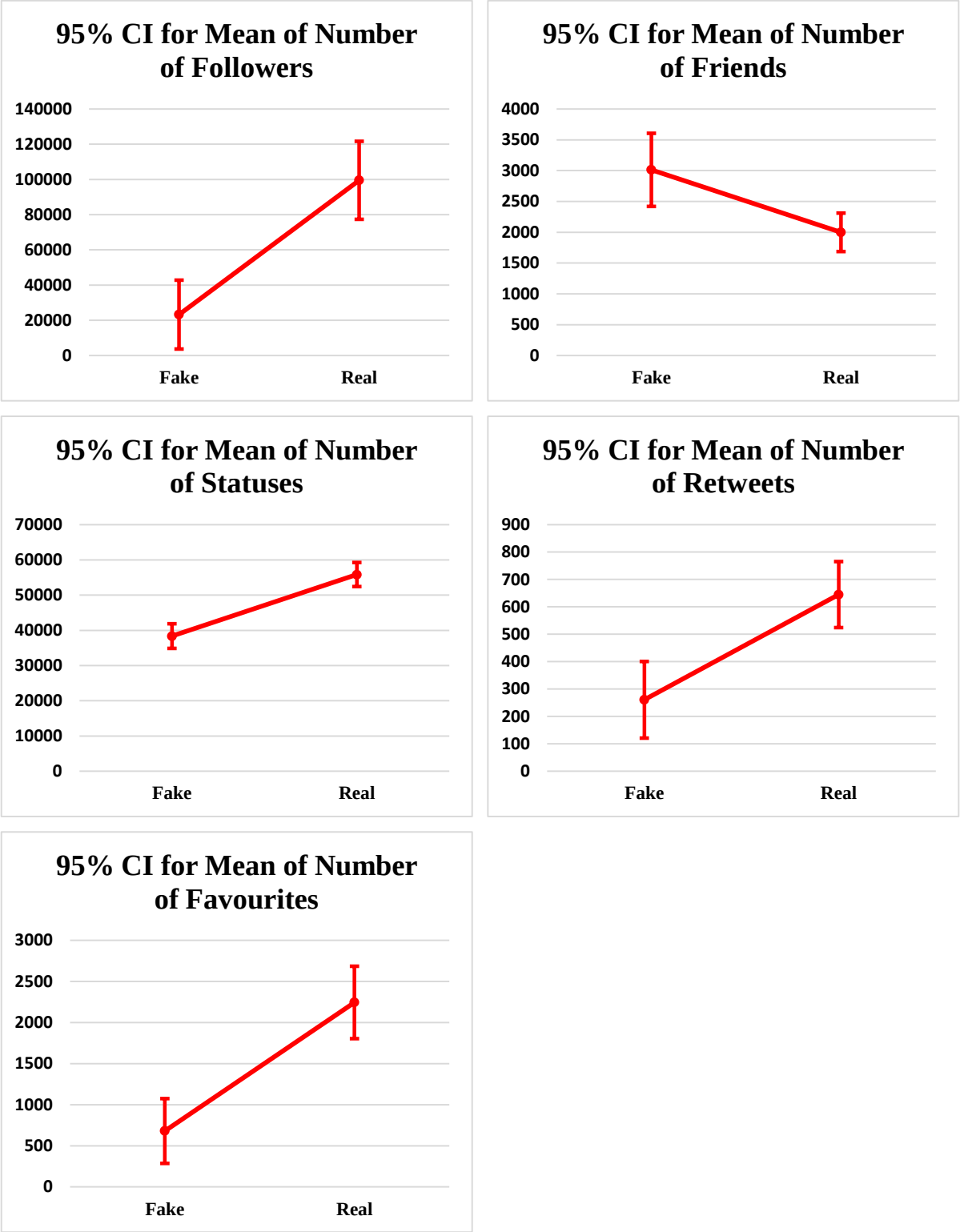


Figure 4: 95% Confidence Interval for quantitative factors on MediaEval dataset

same variable. Hence, any information about this variable in a sample information under test cannot be determined as Fake or Real.

5 Discussion

Fake news on social media is a menace hard to identify and characterize. It is unclear which factors are helpful in distinguishing between real and fake news. Past literature has identified several psychological and behavioral features associated with fake news propagation and acceptance. Little research has been done in identifying key factors characterizing fake news. This study delves deep into factor analysis and their interdependence. We examine how certain factors influence fake news detection and propagation on Twitter. Table 9 summarizes the results of all hypotheses considered.

Table 9: Summary Table

Hypotheses	Results
H_A: Bias of sentiment in fake news with respect to real news.	Reject Null Hypothesis
H_B: Bias of the gender of users involved in fake news with respect to real	Reject Null Hypothesis
H_C: Bias of media usage in fake news with respect to real news.	Reject Null Hypothesis
H_D: Bias in the proportion of a particular gender of the user on the bias in the proportion of sentiments in fake news with respect to real news.	Reject Null Hypothesis
H_E: Bias in the proportion of a particular sentiment used in fake news between different gender of users.	Reject Null Hypothesis
H_F: Bias of inducing a particular sentiment with the usage of media in fake	Reject Null Hypothesis
H_G: Bias in the usage of media amongst different sentiments used in fake	Reject Null Hypothesis
H_H: Relationship between a particular gender and media usage in fake news.	Inconclusive
H_I: Bias in and usage of media in fake news between different gender of users.	Inconclusive
H_J: Significantly distinguishable bias of “follower” count in fake news.	Reject Null Hypothesis
H_K: Significantly distinguishable bias of “friends” count in fake news.	Reject Null Hypothesis
H_L: Significantly distinguishable bias of “status” count in fake news.	Fail to Reject Null Hypotheses
H_M: Significantly distinguishable bias of “retweet” count in fake news.	Reject Null Hypothesis
H_N: Significantly distinguishable bias of “favorite” count in fake news.	Fail to Reject Null Hypotheses

In our qualitative hypotheses H_A, it is assumed that there is a bias in the proportions of sentiment (linguistic tone) in fake news. Although, the central polarity of bias was unclear. With

our study on two COVID-19 specific datasets, we found a strong bias of fake news towards neutral sentiment followed by negative sentiment with respect to real news, which is proved by the results of our first hypothesis. In the second hypothesis, H_B, we tested the bias in the proportion of gender in fake news. The results predicted that there is a strong bias of the male gender towards fake news propagation with respect to real news. Now the influence of the gender ratio of Twitter users is not taken into account as the test is performed to distinguish characteristics of real news and fake news. Any sort of this influence is assumed to affect both types of news equally and nullify its effect. In other words, the speculated gender ratio of 6.85:3.15 should be observed in any random sample collection of tweets. Hence, we directly compare the actual ratio from the dataset without considering the deviation from the speculated ratio. In our datasets, the proportion of tweets (both real and fake) with media is more minor than tweets without media. From the chi-square test results on hypothesis H_C, we find that the proportion of fake news with media is significantly less than expected and substantially more than anticipated for real news with media. Further, we explore if the bias in proportions of one category amongst sentiment, gender, and media usage, is significantly influenced by the bias in proportions of these categories. From the test for Hypothesis H_D, we find that Fake news shared by both male and female gender show bias in proportion of sentiment. The result for hypothesis H_E indicates that this bias is towards fake news being sentiment Neutral, followed by sentiment negative, with respect to real news. This supports our Hypothesis H_A. Further, from the results of testing Hypothesis H_F and H_G, H_G concludes that there is a bias of sentiment in both "with" and "without" media usage. From $H_{F,}$ we conclude that this bias in fake news propagation is proportional to using positive sentiment. For the remaining combination of gender and media usage, from the results of hypotheses H_H and H_I, it cannot be

concluded if there is a mutual influence of Media usage and gender of the user in the bias observed in Hypothesis H_B and H_C due to the contradictory results from the two datasets. In Hypothesis H_H, the contradictory results for "media used" and for H_I, the contradictory results for "Female" gender.

From the quantitative variables, we observe a significant distinguishable difference in the mean number of followers, friends, and retweets for fake and real news. The smaller value of mean for followers can be attributed to why most Real news sources are official media channels and celebrity users who share information on Twitter. In contrast, fake news comes mostly from regular Twitter users who do not have such a huge following. Similar reasons can be attributed to a smaller mean value for retweets of fake news. For the larger value of mean for the number of friends, we understand that the users who propagate fake news are involved in more mutual social connections. Understandably, celebrities and official media sources, when compared to active regular Twitter users, do not have many mutual connections that Twitter classifies as "friends" and, therefore, the resulting smaller value of the mean. The confidence interval for mean for each of these plots acts as a range for true mean for fake and real news and can be used to identify any sample of data by comparing its mean to the 95% CI for the mean of these plots. The non-distinguishable mean value and reverse in the plotted trend for the number of statuses posted by the users who propagated fake and real news and the difference of range for the mean of the number of users who favorited the tweet between the two datasets make these variables unsuitable for classification of the label for the tweet.

6 Conclusion

Fake information on social platforms has constantly been increasing. In the state of the COVID-19 pandemic, this problem has grown at an exponential rate globally. The pandemic is one major event generating misinformation and promoting its consumption through social networks worldwide. In the absence of a holistic fake news detection model, it is unclear what factors can be used to identify misinformation. Very few past works are dedicated to identifying such factors. In this work, we examined several factors from two Twitter datasets, MediaEval 2020 and CovidHeRA, using fourteen hypotheses H_A to H_N. The study uses Chi-square tests for nine qualitative theories (H_A to H_I), whereas for five quantitative tests (H_J to H_N), we have calculated Confidence Intervals using Analysis of Means. Observations from this study unravel specific characteristics to distinguish fake news from real news. These new findings pave the way for future research and development of fake news detection algorithms. We motivate fellow researchers to design algorithms that utilize the discovered dependencies using their combined decisions. Also, we encourage to discover more identifiers that can characterize false information present online ubiquitously. This study provides a new dimension to the existing literature in the fake news domain.

Acknowledgments

We thank Mihir P Mehta (Indian Institute of Management Raipur) for his support and feedback in the experimental setup.

References

[1] Rosenberg, H., Syed, S., & Rezaie, S. (2020). The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. Canadian journal of emergency medicine, 22(4), 418-421.

[2] Soltaninejad, K. (2020). Methanol mass poisoning outbreak, a consequence of COVID-19 pandemic and misleading messages on social media. The international journal of occupational and environmental medicine, 11(3), 148.
[3] Tagliabue, F., Galassi, L., & Mariani, P. (2020). The "pandemic" of disinformation in COVID-19. SN comprehensive clinical medicine, 2(9), 1287-1289.
[4] Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringini, G., & Blackburn, J. (2018, April). What is gab: A bastion of free speech or an alt-right echo chamber. In Companion Proceedings of the The Web Conference 2018 (pp. 1007-1014).
[5] Altay, S., Hacquin, A. S., & Mercier, H. (2019). Why do so few people share fake news? It hurts their reputation. new media & society, 1461444820969893.
[6] Ardèvol-Abreu, A., Delponti, P., & Rodríguez-Wangüemert, C. (2020). Intentional or inadvertent fake news sharing? Fact-checking warnings and users' interaction with social media content. Profesional de la Información, 29(5).
[7] Talwar, S., Dhir, A., Kaur, P., Zafar, N., & Alrasheedy, M. (2019). Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior. Journal of Retailing and Consumer Services, 51, 72-82.
[8] Rampersad, G., & Althiyabi, T. (2020). Fake news: Acceptance by demographics and culture on social media. Journal of Information Technology & Politics, 17(1), 1-11.
[9] Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Systems with Applications, 153, 112986.
[10] Vishwakarma, D. K., & Jain, C. (2020, June). Recent State-of-the-art of Fake News Detection: A Review. In 2020 International Conference for Emerging Technology (INCET) (pp. 1-6). IEEE.
[11] Pogorelov, K., Schroeder, D. T., Burchard, L., Moe, J., Brenner, S., Filkukova, P., & Langguth, J. (2020). Fakenews: Corona virus and 5g conspiracy task at mediaeval 2020. In MediaEval 2020 Workshop.
[12] Dharawat, A., Lourentzou, I., Morales, A., & Zhai, C. (2020). Drink bleach or do what now? Covid-HeRA: A dataset for risk-informed health decision making in the presence of COVID19 misinformation. arXiv preprint arXiv:2010.08743.
[13] Parikh, S. B., Patil, V., & Atrey, P. K. (2019, March). On the origin, proliferation and tone of fake news. In 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 135-140). IEEE.
[14] Lozano, M. G., Brynielsson, J., Franke, U., Rosell, M., Tjörnhammar, E., Varga, S., & Vlassov, V. (2020). Veracity assessment of online data. Decision Support Systems, 129, 113132.
[15] Narwal, B. (2018, October). Fake news in digital media. In 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (pp. 977-981). IEEE.
[16] Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. arXiv preprint arXiv:1811.00770.
[17] Parikh, S. B., & Atrey, P. K. (2018, April). Media-rich fake news detection: A survey. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 436-441). IEEE.
[18] Zhou, X., & Zafarani, R. (2019). Network-based fake news detection: A pattern-driven approach. ACM SIGKDD Explorations Newsletter, 21(2), 48-60.
[19] Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. Computers in human behavior, 80, 295-302.
[20] Talwar, S., Dhir, A., Singh, D., Virk, G. S., & Salo, J. (2020). Sharing of fake news on social media: Application of the honeycomb framework and the third-person effect hypothesis. Journal of Retailing and Consumer Services, 57, 102197.
[21] Brewer, P. R., Young, D. G., & Morreale, M. (2013). The impact of real news about "fake news": Intertextual processes and political satire. International Journal of Public Opinion Research, 25(3), 323-343.
[22] Horne, B., & Adali, S. (2017, May). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 11, No. 1).
[23] C. Silverman and J. Singer-Vine, "Most americans who see fake news believe it, new survey says," BuzzFeed News, 2016.

[24] A. Bovet and H. Makse, “Influence of fake news in Twitter during the 2016 US presidential election,” Nature communications, vol. 10, no. 1, p. 7, 2019.
[25] Osatuyi, B., & Hughes, J. (2018, January). A tale of two internet news platforms-real vs. fake: An elaboration likelihood model perspective. In Proceedings of the 51st Hawaii International Conference on System Sciences.
[26] Colliander, J. (2019). “This is fake news”: Investigating the role of conformity to other users’ views when commenting on and spreading disinformation in social media. Computers in Human Behavior, 97, 202-215.
[27] Moscadelli, A., Albora, G., Biamonte, M. A., Giorgetti, D., Innocenzio, M., Paoli, S., ... & Bonaccorsi, G. (2020). Fake news and covid-19 in Italy: Results of a quantitative observational study. International journal of environmental research and public health, 17(16), 5850.
[28] Calvillo, D. P., Ross, B. J., Garcia, R. J., Smelter, T. J., & Rutchick, A. M. (2020). Political ideology predicts perceptions of the threat of covid-19 (and susceptibility to fake news about it). Social Psychological and Personality Science, 11(8), 1119-1128.
[29] Kim, S., & Kim, S. (2020). The Crisis of public health and infodemic: Analyzing belief structure of fake news about COVID-19 pandemic. Sustainability, 12(23), 9904.
[30] Greene, C. M., & Murphy, G. (2020). Individual differences in susceptibility to false memories for COVID-19 fake news. Cognitive research: principles and implications, 5(1), 1-8.
[31] Lawson, A., & Kakkar, H. (2020). Of pandemics, politics, and personality: The role of conscientiousness and political ideology in sharing of fake news.
[32] Montesi, M. (2020). Understanding fake news during the Covid-19 health crisis from the perspective of information behaviour: The case of Spain. Journal of Librarianship and Information Science, 0961000620949653.
[33] Liu, P. L., & Huang, L. V. (2020). Digital disinformation about COVID-19 and the third-person effect: examining the channel differences and negative emotional outcomes. Cyberpsychology, Behavior, and Social Networking, 23(11), 789-793.
[34] Laato, S., Islam, A. N., Islam, M. N., & Whelan, E. (2020). What drives unverified information sharing and cyberchondria during the COVID-19 pandemic?. European Journal of Information Systems, 29(3), 288-305.
[35] Sulaiman, K. A., Adeyemi, I. O., & Ayegun, I. (2020). Information Sharing and Evaluation as Determinants of Spread of Fake News on Social Media among Nigerian Youths: Experience from COVID-19 Pandemic. International Journal of Knowledge Content Development & Technology, 10(4), 65-82.
[36] Alvi, I., & Saraswat, N. Information Processing–Heuristic vs Systematic and Susceptibility of Sharing COVID 19-Related Fake News on Social Media.
[37] Sap, M., Park, G., Eichstaedt, J., Kern, M., Stillwell, D., Kosinski, M., ... & Schwartz, H. A. (2014). Developing age and gender predictive lexica over social media. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1146-1151).

International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934



Performance evaluation of NH_3H_2O vapour absorption refrigeration system with three cascaded vapour compression refrigeration systems using HFO refrigerants for ultra-low temperature applications

R. S. Mishra

Department of Mechanical Engineering, Delhi Technological University Delhi, India

Abstract

The potential alternative refrigeration technology due to the utilization of energy resources is the vapour compression–absorption refrigeration system. The main issues observed with NH3-H2O vapour absorption refrigeration system are low overall COP and large size of system which is mainly due to the reduced thermodynamic performance of absorption systems. It was also found that at low generator temperatures, the ejector-assisted absorption portion provides better thermodynamic performance than the conventional absorption system adapted in vapour compression–absorption hybrid refrigeration system. The overall performance of system can be improved by increasing the evaporator and generator temperature and by decreasing the absorber, cascade condenser and condenser temperature. Simulation results show that the overall COP of integrated cascaded system is improved by 56 % to 94 %. whereas the improvement of exergetic efficiency is ranging from 25.2% is observed in the absorption section. The total irreversibility of the system is reduced by 7.55% to 28.25%.
©2020 ijrei.com. All rights reserved

Keywords: HFO refrigerants, Thermodynamic performance evaluation, NH_3-H_2O vapour absorption refrigeration, HFO refrigerants

1. Introduction

The potential alternative refrigeration technology due to the utilization of energy resources is the vapour compression–absorption refrigeration system. The main issues observed with VARS is low overall COP and large size of system which is mainly due to the reduced thermodynamic performance of absorption systems. It was also found that at low generator temperatures, the ejector-assisted absorption portion provides better thermodynamic performance than the conventional absorption system adapted in vapour compression–absorption hybrid refrigeration system. The overall performance of system can be improved by increasing the evaporator and generator temperature and by decreasing the absorber, cascade condenser and condenser temperature. In the present study, the performance of vapour compression–absorption cascaded refrigeration system (VCACRS) is studied and the results are compared with the exergy evaluation. The quantitative difference in two approaches is presented which necessitates the need to employ modified approach. The expressions are formulated for effective temperature and real irreversible loss for different components of VCACRS.

2. Integrated Vapour compression absorption refrigeration systems

Canan et.al. [1] carried out studied on LiBr–H2O absorption refrigeration section in compression–absorption cascade refrigeration cycle using with using different (R134a, R-410A) refrigerants in the compression refrigeration cycle and and found electrical energy consumption in the cascade refrigeration cycle is 48–51% lower than classical vapour compression refrigeration cycles that use R134a and R-410A as working fluids under the same operating conditions at evaporator temperature of 263 K and a condenser temperature of 313 K while using R134a, in VCRS the thermal energy consumption of cascade refrigeration cycle could be reduced by 35% and also overall coefficient of performance ($COP_{cyclegen}$) could be improved by 33%.

Corresponding author: R.S. Mishra
Email Address: hod.mechanical.rsm@dtu.ac.in
https://doi.org/10.36037/IJREI.2020.4605

Gulshan sachdeva et.al [2] studied , vapor compression–absorption cascade system using Ammonia–water is considered as the working fluid in absorption cycle and R407C as the working fluid in vapour compression cycle section and carried out second law analysis has been done for the pair in absorption section and determined the exergy destruction (irreversibility rate) in each components of cascade system, for a wide range of cooling capacity by considering a variable speed reciprocating compressor and found the significant reduction in electrical energy consumption. Also found strong effect of effect of varying the generator temperature, effectiveness of solution heat exchanger, inlet temperature of external fluids in evaporator/condenser and some other variables on the system using Coefficient of structural bond (CSB) analysis to quantify the effect of solution heat exchanger and concluded a great scope of improvement to reduce the irreversibility rate of the whole system.

Vaibhav Jain et.al [3] developed, a thermodynamic model for cascaded vapor compression-absorption system consists of a vapor compression refrigeration system (VCRS) coupled with single effect vapor absorption refrigeration system and carried out comparative performance analysis of cascaded system along with independent VCRS for a design capacity of 66.67 kW using first and second laws, and found that the electric power consumption in cascaded system is reduced by 61%. Also computed rational efficiency and effect of operating parameters (i.e., superheating, subcooling, cooling capacity, inlet temperature, product of effectiveness and heat capacitance of external fluids0 on the COP, total irreversibility and rational efficiency of the cascaded system using environment friendly refrigerants such as R410A, R407C and R134. Dario Colorado-Garrido and V. M. Velazquez [4] studied a compression–absorption cascade system LiBr/H_2O using absorption refrigeration cycle with different working fluids(i.e. ammonia, R134a and carbon dioxide) in vapour compression refrigeration cycle and evaluated performance using first and second laws of thermodynamic analysis and simulated the results for finding the best working fluid performance s(i.e. coefficient of performance, exergetic efficiency, irreversibility of the main components of the system, total irreversibility of the system). Also found the highest irreversibilities occurred in the cascade heat exchanger using carbon dioxide or ammonia, and observed that this value decreased by using R134a. The highest value of coefficient of performance is observed by using R134a–LiBr/ H_2O system with the minimum of irreversibility in the absorber and generator within a range of generator temperature from 339 to 345 K. M. Dixit [5] analyzed. absorption-compression cascade refrigeration, comprising of a VCR system using CO2, NH3 and R134a in low temperature stage and a VAR system at the high temperature stage using H_2O-LiBr refrigerant and developed a mathematical model using energy and exergy analysis and found the major source of exergy destruction are cascade condenser, compressor and refrigerant throttle valve

Radhey Shyam Mishra [6] carried out thermodynamic performance of the HFO refrigerants in the medium temperature compression stage between 5ºC to -50ºC and NH_3H_2O refrigerants in the absorption stage and its overall effect on the cascade system is presented. The effect of these HFO refrigerants on the intermediate temperature in the range of (-50ºC to 95ºC) using R245fa of medium temperature cycle cascade system using R32 refrigerant/ hydrocarbons in ultra-low evaporator temperature first and second law performances using the pair of NH_3–H_2O in the high temperature absorption stage and HFO refrigerants at the evaporator temperature of 223K (-50ºC) and R245fa in in the medium temperature compression cycle for evaporator temperature of -95ºC evaluated the effect of various performance parameters of multi cascade refrigeration system in which a compression system at the low temperature stage using R32 in low temperature cycle at evaporator temperature of -130ºC . It is found that R1233zd (E), R1225ye(Z) and HFO-1336mzz(z) gives better thermodynamic performances than using R1243yf. The above investigator has not studied the effect of ecofriendly HFO refrigerants (R1233zd(E), R1225ye(Z) and HFO-1336mzz(z)) in the altra low temperature range from -50ºC to -130ºC respectively. This paper mainly deals the effect of HFO ecofriendly refrigerants in the above range for cryogenic applications.

3. Results and Discussion

The following new systems have been considered for thermodynamic performance evaluations;

System-1

NH_3H_2O vapour absorption refrigeration system at -1ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1233zd(E) in medium temperature cycle at evaporator temperature =-50ºC, R-1225ye(Z) in intermediate temperature cycle at evaporator temperature = -95ºC and using HFO-1336mzz(Z) in lower temperature cycle at evaporator temperature = -150ºC.

System-2

NH_3H_2O vapour absorption refrigeration system at -1ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1233zd(E) in medium temperature cycle at evaporator temperature =-50ºC, HFO-1336mzz(Z) in intermediate temperature cycle at evaporator temperature = -95ºC and using R-1225ye(Z) in lower temperature cycle at evaporator temperature = -150ºC.

System-3

NH_3H_2O vapour absorption refrigeration system at -1ºC of evaporator temperature cascaded with three stages vapour

compression refrigeration system using R1233zd(E) in medium temperature cycle at evaporator temperature =-30ºC, HFO-1336mzz(Z) in intermediate temperature cycle at evaporator temperature = -75ºC and using R-1225ye(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-4

NH_3H_2O vapour absorption refrigeration system at -1ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1234ze(E) in medium temperature cycle at evaporator temperature =-30ºC, R1233zd(E) in intermediate temperature cycle at evaporator temperature = -75ºC and using R-1225ye(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-5

NH_3H_2O vapour absorption refrigeration system at -1ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1243zf in medium temperature cycle at evaporator temperature =-30ºC, R1233zd(E) in intermediate temperature cycle at evaporator temperature = -75ºC and using HFO-1336mzz(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-6

NH_3H_2O vapour absorption refrigeration system at -1ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1243zf in medium temperature cycle at evaporator temperature =-30ºC, R1233zd(E) in intermediate temperature cycle at evaporator temperature = -75ºC and using R-1225ye(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-7

NH_3H_2O vapour absorption refrigeration system at -1ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1243zf in medium temperature cycle at evaporator temperature =-30ºC, R-1225ye(Z) in intermediate temperature cycle at evaporator temperature = -75ºC and using HFO-1336mzz(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-8:

NH_3H_2O vapour absorption refrigeration system at -1ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1243zf in medium temperature cycle at evaporator temperature =-30ºC, HFO-1336mzz(Z) in intermediate temperature cycle at evaporator temperature = -75ºC and using R-1225ye(Z) in lower temperature cycle at evaporator temperature = -135ºC.

Following numerical values have been used for validation of code developed for Integrated NH_3-H_2O VARS using ecofriendly refrigerants.

- Generator temperature= 130ºC
- Absorber temperature=35ºC
- Condenser temperature =35ºC
- VARS evaporator temperature=-1ºC
- load on VARS Evaporator= kW
- Ambient (dead state) temperature=25ºC
- Temperature overlapping in MTC = 10ºC
- VCR evaporator temperature of MTC=-50ºC
- VCR evaporator temperature of ITC=-95ºC
- VCR evaporator temperature of LTC=-150ºC
- VCR compressor efficiency of MTC= 80%
- VCR compressor efficiency of ITC= 80%
- VCR compressor efficiency of MTC= 80%
- Refrigeration Load=35.kW.

Thermodynamic first law energy performances of integrated NH_3-H_2O VARS system cascaded with three stages vapour compression cascaded systems shown in table 1(a) respectively. It was observed that system-2 has lowest thermodynamic performances than system-1, however first law thermodynamic performance improvement is less. Similarly, Thermodynamic second law exergy performances of integrated NH_3-H_2O VARS system cascaded with three stages vapour compression cascaded systems shown in table-1(b) respectively. It was observed that system-2 has higher thermodynamic exergetic performances than system-1. However, second law thermodynamic exergetic performance improvement for -150ºC evaporator temperature is less. It means by putting HFO-1336mzz(Z) in lower temperature cycle in system-1 gives lower first and second law thermodynamic performances than using R-1225ye(Z) in low temperature cycle. However, first law(energy) and second law exegetic performances of double stage cascaded integrated system-1 is higher than double stage cascaded integrated system-2 at evaporator temperature of -100ºC.It means by putting HFO-1336mzz(Z) in intermediate temperature cycle gives lower first and second law thermodynamic performances than using R-1225ye(Z)

Table- 1: Thermodynamic first (Energy) law performance of three stages cascade vapour compression refrigeration system cascade with NH_3-H_2O VARS ($T_{_EVA_LTC}$ =-150ºC, $T_{_EVA_ITC}$=-95ºC, $T_{_EVA_MTC}$=-50ºC)

Integrate system	First law (energetic) Efficiency $COP_{_VARS}$	First law (energetic) Efficiency $COP_{_MTC}$	First law (energetic) Efficiency $COP_{_ITC}$	First law (energetic) Efficiency $COP_{_LTC}$	% Improvement in $COP_{_MTC}$	%Improve-ment in $COP_{_ITC}$	%Improve-ment in $COP_{_LTC}$
System-1	0.322	0.5022	0.6028	0.6251	55.96	87.21	94.13
System-2	0.322	0.5022	0.6011	0.6259	55.98	86.69	94.39

Table- 1(b): Thermodynamic exergy performance of three stages cascade vapour compression refrigeration system cascade with NH_3-H_2O VARS ($T_{_EVA_LTC}$ =-150ºC, $T_{_EVA_ITC}$=-95ºC, $T_{_EVA_MTC}$=-50ºC)

Integrated system	Second law (exergetic) Efficiency $ETA_{_VARS}$	Second law (exergetic) Efficiency $ETA_{_MTC}$	Second law (exergetic) Efficiency $ETA_{_ITC}$	Second law (exergetic) Efficiency $ETA_{_LTC}$	%Improvement in $ETA_{_MTC}$	%Improvement in $ETA_{_ITC}$	%Improvement in $ETA_{_LTC}$
System-1	0.2475	0.3056	0.3211	0.2662	23.5	29.74	7.552
System-2	0.2475	0.3056	0.3174	0.2720	23.5	28.25	9.901

Thermodynamic first law energy performances of integrated NH_3-H_2O VARS system cascaded with three stages vapour compression cascaded systems shown in table 2(a) respectively. It was observed that system-6 has highest thermodynamic performances than other integrated systems. However, first law thermodynamic performance improvement is higher of system-5 and lowest for system-7. Similarly, Thermodynamic first law exergy performances of integrated NH_3-H_2O VARS system cascaded with three stages vapour compression cascaded systems shown in table-2(b) respectively. It was observed that system-8 has lowest thermodynamic exergetic performances. Similarly, second law thermodynamic exergetic performance improvement is less of system-8 as compared to other systems at -75ºC of evaporator temperature.

Table- 2(a): Thermodynamic first law (Energy) performance of three stages cascade vapour compression refrigeration system cascade with NH_3-H_2O VARS for ultra-low temperature applications ($T_{_EVA_LTC}$ =-135ºC, $T_{_EVA_ITC}$=-75ºC, $T_{_EVA_MTC}$=-30ºC)

Integrated System	First law (energetic) Efficiency $COP_{_VARS}$	First law (energetic) Efficiency $COP_{_MTC}$	First law (energetic) Efficiency $COP_{_ITC}$	First law (energetic) Efficiency $COP_{_LTC}$	%Improve-ment in $COP_{_MTC}$	%Improve-ment in $COP_{_ITC}$	%Improve-ment in $COP_{_LTC}$
System-3	0.322	0.5485	0.6712	0.7087	70.36	108.4	120.10
System-4	0.322	0.5485	0.6712	0.7115	70.36	108.4	121.0
System-5	0.322	0.5486	0.6712	0.7088	70.37	108.5	120.1
System-6	0.322	0.5486	0.6712	0.7116	70.37	108.5	121.0
System-7	0.322	0.5486	0.6709	0.7085	70.37	108.4	120.0
System-8	0.322	0.5486	0.6698	0.7109	70.37	108.0	120.5

Table- 2(b): Thermodynamic first law (exergy) performance of three stages cascade vapour compression refrigeration system cascade with NH_3-H_2O VARS for ultra-low temperature applications ($T_{_EVA_LTC}$ =-130ºC, $T_{_EVA_ITC}$=-75ºC, $T_{_EVA_MTC}$) =-30ºC

Integrated cascaded three stages VCRS System	Second law (exergetic) Efficiency $ETA_{_VARS}$	Second law (exergetic) Efficiency $ETA_{_MTC}$	Second law (exergetic) Efficiency $ETA_{_ITC}$	Second law (exergetic) Efficiency $ETA_{_LTC}$	%Improvement in $ETA_{_MTC}$	%Improvement in $ETA_{_ITC}$	%Improvement in $ETA_{_LTC}$
System-3	0.2475	0.2705	0.3117	0.2893	9.326	25.96	16.9
System-4	0.2475	0.2705	0.3117	0.2966	9.326	25.96	19.87
System-5	0.2475	0.2706	0.3118	0.2893	9.356	25.98	16.92
System-6	0.2475	0.2706	0.3118	0.2967	9.356	25.98	19.89
System-7	0.2475	0.2706	0.3112	0.2889	9.356	25.75	16.75
System-8	0.2475	0.2706	0.3092	0.2947	9.356	24.93	19.1

4. Conclusions

Following conclusions were drawn from present investigation

- In single cascading with VARS, at low temperature applications up to -30ºC evaporator temperature, HFO ecofriendly refrigerants (R-1234ze(Z), R-1234ze(E), R1233zd(E), R-1243zf, R1225ye(Z), HFO-1336mzz(Z) and R1234yf will be certainly useful for replacing HFC,

- HCFC and CFC refrigerants, while R1224yd(Z) will be suitable for -10ºC above evaporator temperature for replacing R134a.
- In single cascading with VARS, at low temperature applications up to -50ºC evaporator temperature, HFO ecofriendly refrigerants (R1225ye(Z), R1233zd(E), HFO-1336mzz(Z) and R1234yf will be certainly useful for replacing HFC, HCFC and CFC refrigerants.
- In the double cascading with VARS, at low temperature applications up to -75ºC evaporator temperature, HFO ecofriendly refrigerants (R1225ye(Z), R1233zd(E), HFO-1336mzz(Z) will be certainly useful for replacing HFC, HCFC and CFC refrigerants and can be better than replacing R32, R245fa and R134a.
- In the triple cascading with VARS, at ultra-low temperature applications up to -135ºC evaporator temperature, HFO ecofriendly refrigerants (R1225ye(Z), HFO-1336mzz(Z) will be certainly useful for replacingR32, and, HCFC and CFC refrigerants,

References

[1] Canan ei.al. [2012], Analysis of compression–absorption cascade refrigeration cycles, Applied Thermal Engineering Vol-40: page-311–317.

[2] Gulshan Sachdeva Vaibhav Jain , Surendra Singh Kachhwaha 2013 Exergy analysis of a vapor compression-vapor absorption cascade systemInternational Journal of Air-Conditioning and Refrigeration 21(04):1350026

[3] Vaibhav Jain Surendra Singh Kachhwaha, Gulshan Sachdeva & [2013] Thermodynamic performance analysis of a vapor compression-absorption cascaded refrigeration system, Energy Conversion and Management 75:685-700

[4] Dario Colorado-Garrido & V. M. Velazquez [2013] Exergy analysis of a compression–absorption cascade system for refrigeration, International Journal of Energy Research 37(14), 1851–1865

[5] M Dixit, SC Kaushik, A Arora [2017], Energy and exergy analysis of absorption-compression cascade refrigeration system, Journal of Thermal Engineering,Vol-3, issue-5, 1466-1477

[6] Radhey Shyam Mishra, Single and multiple cascading of VCRS in NH3H2O vapour absorption refrigeration systems for improving thermodynamic (energy-exergy) performances using five ecofriendly new HFOs and other low GWP refrigerants for replacing R134a, International Journal of Research in Engineering and Innovation Vol-4, Issue-2 (2020), 96-104.

Cite this article as: R.S. Mishra, Performance evaluation of NH_3H_2O vapour absorption refrigeration system with three cascaded vapour compression refrigeration systems using HFO refrigerants for ultra-low temperature applications, International Journal of Research in Engineering and Innovation Vol-4, Issue-6 (2020), 333-337. *https://doi.org/10.36037/IJREI.2020.4605*

International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934



ORIGINAL ARTICLE

Performance evaluation of single effect Li/Br-H_2O vapour absorption refrigeration system with three cascaded vapour compression refrigeration systems using HFO refrigerants for ultra-low temperature applications

R. S. Mishra

Department of Mechanical Engineering, Delhi Technological University Delhi, India

Article Information

Received: 23 August 2020
Revised: 15 October 2020
Accepted: 01 November 2020
Available online: 5 November 2020

Keywords:

HFO refrigerants;
Thermodynamic performance evaluation;
Energy-Exergy Evaluation;
Compression-Absorption; Cascade system;

Abstract

The potential alternative absorption refrigeration technology is due to the utilization of renewable and nonrenewable energy resources through vapour compression–absorption refrigeration systems for ultra-low temperature applications. The main issues observed with the vapour absorption refrigeration systems are low overall COP and large size of system which is mainly due to the reduced thermodynamic performance.
This paper mainly deals with the utility of HFO refrigerants in cascaded vapour compression cycles for low temperature and ultra-low temperature applications using single effect Li/Br-H_2O absorption unit. The percentage improvement in overall first law (energy) performance of cascaded system is 7.88% to 8.67% using R1233zd(E), 3.36% to 4.15% by using R1225ye(Z) and 4.66% to 5.45% using HFO-1336mzz(Z) by increasing the generator temperature from 90oC to 120oC, and percentage improvement in exergetic efficiency. from 26.1% to 55.67% by using R1233zd(E) and 17.43% to45.5 % by using R1225ye(Z) and17.4% to 45.49% by using HFO-1336mzz(Z) for low evaporator temperature of -65oC respectively. However, the performances decrease by increasing condenser and absorber temperatures.

©2021 ijrei.com. All rights reserved

1. Introduction

For reducing global warming and ozone depletion and maximizing the use of natural and renewable energy resources with minimal environmental impacts is key building of an energy efficient sustainable system. Although completely sustainable energy efficient system would require a fully reversible process, exploits the second law of thermodynamics which proves is not possible, which indicating that all real processes are irreversible and impact on the environment. Approaching sustainability can yield greater benefits both the current and future environment. The utilization of natural and renewable resources in a responsible fashion and by using efficient methods is important aspect when developing new technologies and analyzing current refrigeration and air conditioning systems. The conventional vapour compression refrigeration systems and vapour absorption systems utilize different working fluids. Many of the vapour compression systems use the ozone depleting chloro-fluoro-carbon refrigerants (CFCs), Hydro-chloro-fluoro-carbon refrigerants (HCFCs), Hydro-fluoro-carbon refrigerants (CFCs), because of the thermo-physical properties obtainable by them. Many industrial processes frequently produce a significant amount of thermal energy; regularly by burning fossil fuels for heat or

Corresponding author: R.S. Mishra
Email Address: hod.mechanical.rsm@dtu.ac.in
https://doi.org/10.36037/IJREI.2021.5101

steam. The opportunities to convert wastes or excess heat into useful cooling can be done by integration of absorption refrigeration systems into these systems. Szargut et al [1]. suggested that exergy methods should be considered to better realize increased efficiencies and environmental impacts because exergy is normally considered to be the measure of work potential (i.e. maximum work) that can be obtained from a system with respect to its environment. The exergy, is a non-conserved quantity, and exergy balances account for inputs, losses and wastes of a process. the exergy input and destruction rates provide an accounting of the utilized efficiency of resources used. Rosen and Dincer [2] have given links between energy, exergy and sustainable development, which shows that exergy may allow for measuring impacts on the environment
Gebreslassie et al [3] had evaluated exergy in Lithium bromide (Li/Br) absorption systems by using structural method to obtain a simplified equation to estimate the optimum heat exchanger area for absorption cooling system and assessed the relationship between heat exchange area and exergy. Also concluded that in the optimum case, the maximum exergy destruction was in the solution heat exchanger and the condenser while in all other components, the destruction rates decreasing as increased heat exchange area and presented at detailed analysis of exergy for half to triple effect absorption chillers. Bereche et al [4] analyzed single and double-effect LiBr systems using a thermo-economic analysis and exergy and concluded that the single-effect absorption refrigeration systems are suitable utilizing waste heat or operating in cogeneration systems because of their operation at lower temperatures compared to double-effect chillers, Morosuk and Tsatsaron [5] is presented an exergy analysis of the internal components of absorption refrigeration machines. And concluded that the absorber and generator destroyed about 40% of their exergy and are main (primary) candidates for improvement. Kilic and Kaynakli [6] carried out a second law(exergy) thermodynamic analysis of water and lithium bromide absorption refrigeration and found the evaporator is a major component for the exergy loss rates and concluded that the generator, absorber and evaporator were the largest sources of exergy destruction. Garimella et al [7] proposed absorption/vapour compression cascade refrigeration system driven by waste heat used in naval ship and concluded that electricity consumption is reduced by 31% than that of conventional vapour compression refrigeration system.
The large number of research studies on cascade refrigeration systems have been done. These research studies provide valuable insights about energy and exergy analysis of VCR, VAR cycles in the single, double and triple stage and compression-absorption in the form of cascade cycles. Most of the research studies considered till date emphasize on VCR and VAR cycles (single and double effect, triple and half effect in Li/Br-H_2O and NH_3H_2O). Mishra [8] carried out exergy- energy analysis of compression-absorption (combined) or cascade cycles. Though, exhaustive research has been carried out on cascade cycles, but very less consideration has been given to explore the thermodynamic performance using HFO refrigerants in the range from -30ºC to -150ºC and none of the research work is available on performance analysis of compression-absorption cascade refrigeration system using HFO refrigerants. Radhey Shyam Mishra [9-10] carried out thermodynamic performance of the HFO refrigerants in the medium temperature compression stage between 5ºC to -50ºC and NH_3H_2O, Li/Br-H_2O refrigerants in the absorption stage and its overall effect on the cascade system. The effect of these HFO refrigerants on the intermediate temperature in the range of (-50ºC to 95ºC) using R245fa of medium temperature cycle cascade system using R32 refrigerant/ hydrocarbons in ultra-low evaporator temperature first and second law performances using the pair of NH_3–H_2O in the high temperature absorption stage and HFO refrigerants at the evaporator temperature of 223K (-50ºC) and R245fa in the medium temperature compression cycle for evaporator temperature of -95ºC evaluated the effect of various performance parameters of multi cascade refrigeration system in which a compression system at the low temperature stage using R32 in low temperature cycle at evaporator temperature of -130ºC . It is found that R1233zd (E), R1225ye(Z) and HFO-1336mzz(z) gives better thermodynamic performances than using R1243yf. The studies carried out so far on the double and triple effect cycles in the use of HFO refrigerants up to evaporator temperature of -50ºC in cascaded vapour compression refrigeration systems. to some extent, the effect of HFO refrigerants in the intermediate temperature cycle at -75ºC and -95ºC is missing. Also the effect of HFO –refrigerants in ultra-low temperature cycle of compression –absorption refrigeration systems have not been investigated so far. The present study has therefore been carried out the effect of HFO refrigerants in intermediate temperature cycle(ITC) up to -95ºC and LTC evaporator of -135ºC and -150ºC.

2. Results and Discussion

Following systems have been taken for performance evaluations.

System-1

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1233zd(E) in medium temperature cycle at evaporator temperature =-50ºC, R-1225ye(Z) in intermediate temperature cycle at evaporator temperature = -95ºC and using HFO-1336mzz(Z) in lower temperature cycle at evaporator temperature = -150ºC.

System-2

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1233zd(E) in medium temperature cycle at evaporator temperature =-50ºC, HFO-1336mzz(Z) in intermediate temperature cycle at evaporator temperature = -95ºC and using R-1225ye(Z) in lower

temperature cycle at evaporator temperature = -150ºC.

System-3

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1234yf in medium temperature cycle at evaporator temperature =-50ºC, R-1225ye(Z) in intermediate temperature cycle at evaporator temperature = -95ºC and using HFO-1336mzz(Z) in lower temperature cycle at evaporator temperature = -150ºC.

System-4

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1234yf) in medium temperature cycle at evaporator temperature =-50ºC, HFO-1336mzz(Z) in intermediate temperature cycle at evaporator temperature = -95ºC and using R-1225ye(Z) in lower temperature cycle at evaporator temperature = -150ºC.
Following numerical values have been used for validation of code developed for Integrated single effect Li/Br-H_2O VARS using ecofriendly refrigerants.

- Generator temperature= 110ºC
- Absorber temperature=35ºC
- Condenser temperature =35ºC
- VARS evaporator temperature=10ºC
- load on VARS Evaporator= 175 kW
- Ambient (dead state) temperature=25ºC
- Temperature overlapping in MTC = 10ºC
- VCR evaporator temperature of MTC=-50ºC
- VCR evaporator temperature of ITC=-95ºC
- VCR evaporator temperature of LTC=-150ºC
- VCR compressor efficiency of MTC= 80%
- VCR compressor efficiency of ITC= 80%
- VCR compressor efficiency of MTC= 80%
- Ambient (dead state) temperature=25ºC

Thermodynamic first law energy performances of integrated single effect Li/Br-H_2O VARS system cascaded with three stages vapour compression cascaded systems shown in table 2(a) respectively. It was observed that system-6 has lowest thermodynamic performances than system-5, however first law thermodynamic performance improvement is less. Similarly, Thermodynamic second law exergy performances of cascaded single effect Li/Br-H_2O VARS system cascaded with three stages vapour compression cascaded systems shown in table-2(b) respectively. It was observed that system-6 has higher thermodynamic exergetic performances than system-5.

Table-1(a) Thermodynamic first law (energetic) Performances of Integrated single effect Li/Br-H_2O VARS using ecofriendly refrigerants at evaporator temperature=-150ºC

Integrate system	COP_{VARS}	COP_{MTC}	COP_{ITC}	COP_{LTC}	% Improvement in COP_{MTC}	% Improvement in COP_{ITC}	% Improvement in COP_{LTC}
System-1	0.7560	0.8759	0.9847	0.9236	15.86	30.65	22.17
System-2	0.7560	0.8759	0.9847	0.9282	15.86	30.26	22.79
System-3	0.7560	0.8102	0.9120	0.8634	7.168	20.64	14.22
System-4	0.7560	0.8102	0.9120	0.8674	7.168	20.31	14.74

Table-1(b) Thermodynamic second law (exergetic) Performances of Integrated single effect Li/Br-H_2O VARS using ecofriendly refrigerants at evaporator temperature=-150ºC

Integrate system	ETA_{VARS}	ETA_{MTC}	ETA_{ITC}	ETA_{LTC}	% Improvement in ETA_{MTC}	% Improvement in ETA_{ITC}	% Improvement in ETA_{LTC}
System-1	0.1564	0.321	0.5331	0.5539	77.98	195.3	206.8
System-2	0.1564	0.321	0.5293	0.5639	77.98	193.1	212.3
System-3	0.1564	0.2680	0.4561	0.4890	48.94	152.6	170.9
System-4	0.1564	0.2689	0.4530	0.4974	48.94	150.9	175..5

Table-1(c) Thermodynamic exergy destruction ratio of Integrated single effect Li/Br-H_2O VARS using ecofriendly refrigerants at evaporator temperature= -150ºC

Integrate system	EDR_{VARS}	EDR_{MTC}	EDR_{ITC}	$ED R_{LTC}$	% Improvement in EDR_{MTC}	% Improvement in EDR_{ITC}	% Improvement in EDR_{LTC}
System-1	4.539	2.112	0.8754	0.8053	-53.47	-80.7	-82.26
System-2	4.539	2.112	0.8894	0.7735	-53.47	-80.4	-82.96
System-3	4.539	2.719	0.8754	0.8053	-40.10	-73.72	-76.98
System-4	4.539	2.719	0.8754	0.8053	40.10	-73.40	-77.74

However, second law thermodynamic exergetic performance improvement for -150ºC evaporator temperature is less. It means

by putting HFO-1336mzz(Z) in lower temperature cycle in system-5 gives lower first and second law thermodynamic performances than using R-1225ye(Z) in low temperature cycle. However, first law(energy) and second law exegetic performances of double stage cascaded integrated system-5 is higher than double stage cascaded integrated system-6 at evaporator temperature of -95ºC.It means by putting HFO-1336mzz(Z) in intermediate temperature cycle gives lower first and second law thermodynamic performances than using R-1225ye(Z). The exergy destruction ratio of whole cascaded system is shown in Table-2(c) respectively.

System-5

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1234ze(Z) in medium temperature cycle at evaporator temperature =-30ºC, R1225ye(Z) in intermediate temperature cycle at evaporator temperature = -75ºC and using HFO-1336mzz(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-6

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1234ze(Z) in medium temperature cycle at evaporator temperature =-30ºC, HFO-1336mzz(Z) in intermediate temperature cycle at evaporator temperature = -75ºC and using R1225ye(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-7

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1234ze(E) in medium temperature cycle at evaporator temperature =-30ºC, R1225ye(Z) in intermediate temperature cycle at evaporator temperature = -75ºC and using HFO-1336mzz(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-8

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1234ze(E) in medium temperature cycle at evaporator temperature =-30ºC, HFO-1336mzz(Z) in intermediate temperature cycle at evaporator temperature = -75ºC and using R1225ye(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-9

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1243zf in medium temperature cycle at evaporator temperature =-50ºC, R-1225ye(Z) in intermediate temperature cycle at evaporator temperature = -95ºC and using HFO-1336mzz(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-10

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1243zf in medium temperature cycle at evaporator temperature =-30ºC, HFO-1336mzz(Z)in intermediate temperature cycle at evaporator temperature = -75ºC and using R-1225ye(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-11

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1233zd(E) in medium temperature cycle at evaporator temperature =-50ºC, R-1225ye(Z) in intermediate temperature cycle at evaporator temperature = -95ºC and using HFO-1336mzz(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-12

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1233zd(E) in medium temperature cycle at evaporator temperature =-50ºC, HFO-1336mzz(Z) in intermediate temperature cycle at evaporator temperature = -95ºC and using R-1225ye(Z) in lower temperature cycle at evaporator temperature = -150ºC.

System-13

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1234yf in medium temperature cycle at evaporator temperature =-50ºC, R-1225ye(Z) in intermediate temperature cycle at evaporator temperature = -95ºC and using HFO-1336mzz(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System14

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1234yf in medium temperature cycle at evaporator temperature =-50ºC, HFO-1336mzz(Z) in intermediate temperature cycle at evaporator temperature = -95ºC and using R-1225ye(Z) in lower

temperature cycle at evaporator temperature = -135ºC.

System-15

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R-1225ye(Z) in medium temperature cycle at evaporator temperature =-50ºC, R1233zd(E) in intermediate temperature cycle at evaporator temperature = -95ºC and using HFO-1336mzz(Z) in lower temperature cycle at evaporator temperature = -135ºC.

System-16

Li/Br-H_2O single effect vapour absorption refrigeration system at 10ºC of evaporator temperature cascaded with three stages vapour compression refrigeration system using R1234yf in medium temperature cycle at evaporator temperature =-50ºC, R1233zd(E) in intermediate temperature cycle at evaporator temperature = -95ºC and using R-1225ye(Z) in lower temperature cycle at evaporator temperature = -135ºC.

Following numerical values have been used for validation of code developed for Integrated single effect Li/Br-H_2O VARS using ecofriendly refrigerants

- Generator temperature= 110ºC
- Absorber temperature=35ºC
- Condenser temperature =35ºC
- VARS evaporator temperature=10ºC
- load on VARS Evaporator= 175 kW
- Ambient (dead state) temperature=25ºC
- Temperature overlapping in MTC = 10ºC
- VCR evaporator temperature of MTC=-30ºC
- VCR evaporator temperature of ITC=-75ºC
- VCR evaporator temperature of LTC=-135ºC
- VCR compressor efficiency of MTC= 80%
- VCR compressor efficiency of ITC= 80%
- VCR compressor efficiency of MTC= 80%
- Ambient (dead state) temperature=25ºC

Thermodynamic first law energy performances of integrated single effect Li/Br-H_2O VARS system cascaded with three stages vapour compression cascaded systems shown in table 2(a) respectively. It was observed that system-2 has lowest thermodynamic performances than system-1, however first law thermodynamic performance improvement is less

Table-2(a) Thermodynamic first law (energetic) Performances of Integrated single effect Li/Br-H_2O VARS using ecofriendly refrigerants at evaporator temperature=-135ºC

Integrate system	COP_{VARS}	COP_{MTC}	COP_{ITC}	COP_{LTC}	%Improvement in COP_{MTC}	%Improvement in COP_{ITC}	% Improvement in COP_{LTC}
System-5	0.7560	1.002	1.145	1.05	32.56	51.44	38.96
System-6	0.7560	1.002	1.143	1.057	32.58	51.15	39.82
System-7	0.7560	0.9662	1.104	1.020	27.81	46.032	34.92
System-8	0.7560	0.9662	1.102	1.057	27.81	45.82	35.75
System-9	0.7560	0.9698	1.108	1.057	28.29	46.62	35.35
System-10	0.7560	0.9698	1.106	1.057	28.29	46.36	36.16
System-11	0.7560	0.9961	1.138	1.057	31.76	50.53	38.28
System-12	0.7560	0.9961	1.138	1.045	31.76	50.53	38.28
System-13	0.7560	0.9431	1.078	1.0	24.76	42.62	32.31
System-14	0.7560	0.9431	1.076	1.006	24.76	42.37	32.08
System-15	0.7560	0.9634	1.102	1.018	2744	45.74	34.68
System-16	0.7560	0.9794	1.120	1.039	29.56	48.13	37.5

Table-2(b) Thermodynamic second law (exergetic) Performances of Integrated single effect Li/Br-H_2O VARS using ecofriendly refrigerants at evaporator temperature=-135ºC

Integrate system	ETA_{VARS}	ETA_{MTC}	ETA_{ITC}	ETA_{LTC}	%Improvement in ETA_{MTC}	% Improvement in ETA_{ITC}	%Improvement in ETA_{LTC}
System-5	0.1560	0.322	0.5278	0.5383	78.49	192.3	198.1
System-6	0.1560	0.3223	0.5255	0.5491	78.49	191.0	204.1
System-7	0.1560	0.2970	0.4936	0.5126	64.50	173.5	183.5
System-8	0.1560	0.2970	0.4917	0.5227	64.5	172.3	189.5
System-9	0.1560	0.2992	0.4972	0.5152	65.86	175.4	185.3
System-10	0.1560	0.2995	0.4951	0.5253	65.86	174.5	191.0
System-11	0.1560	0.3170	0.5219	0.5339	76.04	189.1	195.7
System-12	0.1560	0.3170	0.5219	0.5339	76.04	189.1	195.7
System-13	0.1560	0.2817	0.4723	0.4958	55.96	161.6	174.6
System-14	0.1560	0.2816	0.4704	0.5054	55.96	160.5	179.9
System-15	0.1560	0.2950	0.4918	0.5110	55.96	172.4	183.0
System-16	0.1560	0.3061	0.5068	0.5347	69.54	180.7	196.1

Table-2(c) Thermodynamic exergy destruction ratio of Integrated single effect Li/Br-H_2O VARS at evaporator temperature=-135°C

Integrate system	EDR_{VARS}	EDR_{MTC}	EDR_{ITC}	$ED R_{LTC}$	% Improve in EDR_{MTC}	%Improvement in EDR_{ITC}	%Improvement in EDR_{LTC}
System-5	4.539	2.103	0.9031	0.8212	-53.66	-80.10	-81.91
System-6	4.539	2.103	0.8947	0.8577	-53.66	-80.29	-81.1
System-7	4.539	2.367	1.025	0.9529	-47.85	-77.41	-79.05
System-8	4.539	2.367	1.034	0.9133	-47.82	-77.23	-79.88
System-9	4.539	2.339	1.011	0.9410	-48.46	-77.72	-79.23
System-10	4.539	2.339	1.02	0.9036	-48.46	-77.53	-80.09
System-11	4.539	2.146	0.916	0.8729	-52.71	-79.82	-80.77
System-12	4.539	2.146	0.916	0.8729	-52.71	-79.82	-80.77
System-13	4.539	2.551	1.117	1.017	-43.79	-75.39	-77.6
System-14	4.539	2.551	1.126	0.9785	-47.37	-77.23	-78.92
System-15	4.389	2.551	1.033	0.9569	-43.79	-75.19	-78.44
System-16	4.539	2.267	0.9730	0.8703	-50.06	-78.56	-80.83

Similarly, Thermodynamic second tlaw exergy performances of cascaded single effect Li/Br-H_2O VARS system cascaded with three stages vapour compression cascaded systems shown in table-2(b) respectively. It was observed that system-2 has higher thermodynamic exergetic performances than system-1. However, second law thermodynamic exergetic performance improvement for -150°C evaporator temperature is less. It means by putting HFO-1336mzz(Z) in lower temperature cycle in system-1 gives lower first and second law thermodynamic performances than using R-1225ye(Z) in low temperature cycle. However, first law(energy) and second law exegetic performances of double stage cascaded integrated system-1 is higher than double stage cascaded integrated system-2 at evaporator temperature of -95°C.It means by putting HFO-1336mzz(Z) in intermediate temperature cycle gives lower first and second law thermodynamic performances than using R-1225ye(Z). The exergy destruction ratio of whole cascaded system is shown in Table-2(c) respectively.

Thermodynamic first law energy performances of integrated single effect Li/Br-H_2O VARS system cascaded with three stages vapour compression cascaded systems with varying evaporator temperature is shown in table -3 respectively. It was observed that by increasing VAR evaporator temperature, the thermodynamic first law (energy) performance is increases while second law (exergy)efficiency is decreases and exergy destruction ratio is increases. Similarly, Thermodynamic second law exergy performances of cascaded single effect Li/Br-H_2O VARS system cascaded with three stages vapour compression cascaded systems with variation of absorber temperature is shown in table-4 respectively. It was observed thermodynamic performances
is decreases, by increasing absorber temperature.

Table-3 Thermodynamic (exergy) performances of Integrated single effect Li/Br-H2O VARS using R1233zd(E) ecofriendly refrigerants at MTC evaporator temperature=-65°C, T_Eva_VARS= 5°C, Approach=, 5°C

Evaporator temperature of VARS (°C)	COP_{VARS}	EDR	Exergetic Efficiency	$COP_{cascade}$	$EDR_{cascade}$	Exergetic $Efficiency_{cascade}$	% Improvement in $COP_{cascade}$	% Decrement in $EDR_{cascade}$	%Improv-ment in Exergetic $Efficiency_{cascade}$
3	0.7360	2.783	0.2643	0.7789	1.933	0.3410	5.03	-30.56	29.0
4	0.7410	2.964	0.2523	0.7808	1.970	0.3367	5.03	-33.55	33.48
5	0.7560	3.163	0.2402	0.7827	2.008	0.3325	5.03	-36.52	38.41
6	0.7438	3.382	0.2282	0.7848	2.046	0.3283	5.03	-39.49	43.84
7	0.7466	3.624	0.2163	0.7869	2.085	0.3241	5.03	-42.46	49.87
8	0.7496	3.894	0.2043	0.7892	2.125	0.320	5.03	-45.42	56.60
9	0.7525	4.197	0.1924	0.7916	2.166	0.3159	5.16	-48.40	64.16
10	0.7560	4.539	0.1805	0.7940	2.207	0.3119	5.03	-51.38	72.73

Table-4 Effect of Absorber temperature on thermodynamic performances of Integrated single effect Li/Br-H2O VARS using R1233zd(E) ecofriendly refrigerants at MTC evaporator temperature=-65°C, T_Eva_VARS= 5°C, Approach=, 5°C

Absorber temperature of VARS (°C)	COP_{VARS}	EDR	Exergetic Efficiency	$COP_{cascade}$	$EDR_{cascade}$	Exergetic $Efficiency_{cascade}$	% Improvement in $COP_{cascade}$	% Decrement in $EDR_{cascade}$	% Improvement in Exergetic $Efficiency_{cascade}$
30	0.7486	3.121	0.2427	0.7885	1.997	0.3337	5.323	-36.02	37.51
35	0.7410	3.163	0.2402	0.7827	2.008	0.3325	5.627	-36.52	38.41
40	0.7353	3.195	0.2384	0.7784	2.016	0.3315	5.856	-36.90	39.09
45	0.7312	3.219	0.2370	0.7752	2.022	0.3309	6.023	-37.17	39.59

Thermodynamic first law energy performances of integrated single effect Li/Br-H_2O VARS system cascaded with three stages vapour compression cascaded systems with varying Condenser temperature is shown in table -5 respectively. It was observed that by increasing VARS Condenser temperature, the thermodynamic first law (energy) performance is decreases while second law (exergy)efficiency is decreases and exergy destruction ratio is increases. Similarly, Thermodynamic performances of cascaded single effect Li/Br-H_2O VARS system cascaded with three stages vapour compression cascaded systems with variation of generator temperature is shown in table-6 respectively. It was observed thermodynamic performances is decreases, by increasing generator temperature.

Table-5 Effect of condenser temperature on thermodynamic performances of Integrated single effect Li/Br-H_2O VARS using R1233zd(E) ecofriendly refrigerants at MTC evaporator temperature=-65ºC, T_Eva_VARS= 5ºC, Approach=, 5ºC

Condenser temperature of VARS (ºC)	COP_VARS	EDR	Exergetic Efficiency	COP _ cascade	EDR_ cascade	Exergetic Efficiency_ cascade	% Improvement in COP _ cascade	% Decrement in EDR _ cascade	%Improv-ment in Exergetic Efficiency _ cascade
30	0.7472	3.129	0.2422	0.7874	2.003	0.3330	5.38	-36.11	37.61
32	0.7449	3.141	0.2415	0.7857	2.004	0.3328	5.470	-36.26	37.94
34	0.7423	3.156	0.2406	0.7837	2.006	0.3327	5.574	-36.44	38.25
35	0.7410	3.163	0.2402	0.7827	2.008	0.3325	5.627	-36.52	38.41
36	0.7398	3.170	0.2398	0.7818	2.070	0.3323	5.678	-36.61	38.56
38	0.7374	3.183	0.2390	0.780	2.013	0.3319	5.773	-36.76	38.84
40	0.7353	3.195	0.2384	0.7784	2.016	0.3315	5.856	-36.90	39.09
42	0.7336	3.205	0.2378	0.7771	2.019	0.3313	5.925	-37.01	39.30
44	0.7322	3.213	0.2373	0.7760	2.021	0.3310	5.983	-37.11	39.47
45	0.7315	3.217	0.2371	0.7755	2.022	0.3309	6.009	-37.15	39.55

Table-6 Effect of generator temperature on thermodynamic performances of Integrated single effect Li/Br-H_2O VARS using R1233zd(E) ecofriendly refrigerants at MTC evaporator temperature=-65ºC, T_Eva_VARS= 5ºC, Approach=, 5ºC

Generator temperature of VARS (ºC)	COP_VARS	EDR_VARS	Exergetic Efficiency_VARS	COP _MTC	EDR _MTC	Exergetic Efficiency _ M TC	%Improve-ment in COP _MTC	%Improve-ment in COP _ITC	%Improv-ment in COP _ LTC
90	0.7596	2.277	0.3052	0.8194	1.599	0.3849	7.881	-29.79	26.10
95	0.7532	2.510	0.2849	0.8145	1.657	0.3761	8.132	-33.98	32.10
100	0.7478	2.737	0.2676	0.8103	1.715	0.3684	8.346	-37.37	37.67
105	0.7437	2.955	0.2528	0.8070	1.770	0.3611	8.509	-40.13	42.82
110	0.7410	3.163	0.2402	0.8049	1.822	0.3544	8.618	-42.40	47.53
115	0.7388	3.364	0.2291	0.8031	1.873	0.3481	8.709	-44.34	51.93
120	0.7399	3.541	0.2202	0.8040	1.917	0.3428	8.665	-45.87	55.68

Thermodynamic first law energy performances of integrated single effect Li/Br-H_2O VARS system cascaded with three stages vapour compression cascaded systems with varying generator temperature for constant evaporator temperature of -75ºC in lower vapour compression cycle is shown in tables -7 for R1233zd(E), R1225ye(Z) and HFO-1336mzz(Z) respectively. It was observed that by increasing VARS Condenser temperature, the thermodynamic first law (energy) performance is decreases while second law (exergy)efficiency is decreases and exergy destruction ratio is increases. Similarly, Thermodynamic performances of cascaded single effect Li/Br-H_2O VARS system cascaded with three stages vapour compression cascaded systems with variation of generator temperature is shown in table-6 respectively. It was observed thermodynamic performances

Table-7(a) Effect of generator temperature on thermodynamic performances of Integrated single effect Li/Br-H_2O VARS using R1233zd(E) ecofriendly refrigerants at MTC evaporator temperature=-75ºC, T_Eva_VARS= 5ºC, Approach=, 5ºC

Generator temperature of VARS (ºC)	COP_ Cascade	EDR_ Cascade	Exergetic Efficiency_ Cascade
90	0.8194	1.599	0.3849
95	0.8145	1.657	0.3761
100	0.8103	1.715	0.3684
105	0.8070	1.770	0.3611
110	0.8049	1.822	0.3544
115	0.8031	1.873	0.3481
120	0.8040	1.917	0.3428

Table-7(b) Effect of generator temperature on improvement in thermodynamic performances of Integrated single effect Li/Br-H_2O VARS using R1233zd(E) ecofriendly refrigerants at MTC evaporator temperature=-75°C, T_Eva_VARS= 5°C, Approach=, 5°C

Generator temperature of VARS (°C)	% improve-ment in COP_ Cascade	% decre-ment in EDR_ Cascade	% improvement in Exergetic Eff_ Cascade
90	7.881	-29.79	26.10
95	8.132	-33.98	32.10
100	8.346	-37.36	37.67
105	8.509	-40.13	42.82
110	8.618	-42.40	47.53
115	8.709	-44.34	51.93
120	8.665	-45.87	55.68

Table-7(c) Effect of generator temperature on thermodynamic performances of Integrated single effect Li/Br-H_2O VARS using HFO1336mzz(Z) ecofriendly refrigerants at MTC evaporator temperature=-75°C, T_Eva_VARS= 5°C, Approach=, 5°C

Generator temperature of VARS (°C)	COP_ cascade	EDR_ cascade	Exergetic Efficiency_ cascade
90	0.7950	1.790	0.3594
95	0.7902	1.852	0.3507
100	0.7478	1.911	0.3435
105	0.7437	1.968	0.3369
110	0.7410	2.022	0.3309
115	0.7388	2.075	0.3252
120	0.7399	2.122	0.3204

Table-7(d) Effect of generator temperature on percentage improvement in thermodynamic performances of Integrated single effect Li/Br-H_2O VARS using HFO1336mzz(Z) ecofriendly refrigerants at MTC evaporator temperature=-65°C, T_Eva_VARS= 5°C, Approach=, 5°C

Generator temperature of VARS (°C)	% improvement in COP_cascade	% decrement in EDR_ cascade	% improvement in Exergetic Efficiency_ cascade
90	4.662	-21.36	17.43
95	4.914	-26.26	23.10
100	5.129	-30.18	28.38
105	5.293	-33.41	33.26
110	5.402	-36.06	37.73
115	5.493	-38.32	41.92
120	5.449	-40.09	45.49

Table-7(e) Effect of generator temperature on thermodynamic performances of Integrated single effect Li/Br-H_2O VARS using R1225ye(Z) ecofriendly refrigerants at MTC evaporator temperature=-65°C, T_Eva_VARS= 5°C, Approach=, 5°C

Generator temperature of VARS (°C)	COP_ cascade	EDR_ cascade	Exergetic Efficiency_ cascade
90	0.7851	1.874	0.3479
95	0.7532	1.936	0.3406
100	0.7478	1.997	0.3337
105	0.7437	2.055	0.3274
110	0.7410	2.11	0.3216
115	0.7388	2.163	0.3161
120	0.7399	2.211	0.3115

Table-7(f) Effect of generator temperature on percentage improvement thermodynamic performances of Integrated single effect Li/Br-H_2O VARS using R1225ye(Z) ecofriendly refrigerants at MTC evaporator temperature=-75°C, T_Eva_VARS= 5°C, Approach=, 5°C

Generator temperature of VARS (°C)	% improve-ment in COP_ cascade	% decre-ment in EDR_ cascade	% improvement in Exergetic Efficiency_ cascade
90	3.359	-17.69	17.43
95	3.612	-22.87	23.10
100	3.827	-27.06	28.38
105	3.991	-30.48	33.26
110	4.10	-33.29	37.73
115	4.192	-35.70	41.92
120	4.148	-37.58	45.49

3. Conclusions

Following conclusions were drawn from present investigations

- By reducing the exergy destruction rate by using HFO refrigerants therefore, sustainability improved.
- The best thermodynamic performances were observed by using R1234ze(Z) and R1233zd(E), HFO-1336mzz(Z), and R1225ye(Z) and R1234ze(E) and R1243zf performance
- The lowest thermodynamic performances were fund by using R1234yf in cascaded system in lower circuit up to a VCR evaporator temperature
- In single cascading with VARS, at low temperature applications up to -30°C evaporator temperature, HFO ecofriendly refrigerants (R-1234ze(Z), R-1234ze(E), R1233zd(E), R-1243zf, R1225ye(Z), HFO-1336mzz(Z) and R1234yf will be certainly useful for replacing HFC, HCFC and CFC refrigerants, while R1224yd(Z) will be suitable for -10°C above evaporator temperature for replacing R134a.
- In single cascading with VARS, at low temperature applications up to -50°C evaporator temperature, HFO ecofriendly refrigerants (R1225ye(Z), R1233zd(E), HFO-1336mzz(Z) and R1234yf will be certainly useful for replacing HFC, HCFC and CFC refrigerants.
- In the double cascading with VARS, at low temperature applications up to -75°C evaporator temperature, HFO ecofriendly refrigerants (R1225ye(Z), R1233zd(E), HFO-1336mzz(Z) will be certainly useful for replacing HFC, HCFC and CFC refrigerants and can be better than replacing R134a
- In the triple cascading with VARS, at ultra-low temperature applications up to -150°C evaporator temperature, HFO ecofriendly refrigerants (R1225ye(Z), HFO-1336mzz(Z) will be certainly useful for replacingR32, and, HCFC and CFC refrigerants.

References

[1] Szargut, J.; Morris, D.R.; Steward, F.R. (1987), Exergy Analysis of Thermal, Chemical, and Metallurgical Processes; Hemisphere Publishing: New York, NY, USA.

[2] Rosen, M.A.; Dincer, I. (2005), Efficiency analysis of a cogeneration and district energy system. Applied Thermal Engineering, 25, 147–159

[3] Gebreslassie, B.H.; Medrano, M.; Boer,D. (2010), Exergy analysis of multi-effect water-LiBr absorption systems: From half to triple effect. Renew. Energy, 35, 1773–1782.
[4] Bereche, R.P.; Palomino, R.G.; Nebra, S.A. (2009), Thermoeconomic analysis of a single and double-effect LiBr/H_2O absorption refrigeration system. International journal of thermodynamics, 12, 89–96
[5] Morosuk, T.; Tsatsaronis, G. (2008), A new approach to the exergy analysis of absorption refrigeration machines. Energy ,33, 890–907
[6] Kilic, M.; Kaynakli, O. (2007), Second law-based thermodynamic analysis of water-lithium bromide absorption refrigeration system. Energy, 32, 1505–1512.
[7] Garimella, S., Brown, A.M., Nagavarapu, A.K., (2011), Waste heat driven absorption/vapor-compression cascade refrigeration system for megawatt scale, high-flux, lowtemperature cooling. International journal of refrigeration, 34, 1776–1785.
[8] Radhey Shyam Mishra, (2019), Thermal performance of cascaded vapour compression absorption systems, International Journal of Research in Engineering and Innovation, 3 (1), 61-67**.**
[9] Radhey Shyam Mishra, (2019), Performance evaluation of half effect Li/Br-H2O vapour absorption systems using multi cascading of vapour compression cycles for ultra-low temperature applications International Journal of Research in Engineering and Innovation, 3 (1), 509-526.
[10] Radhey Shyam Mishra, (2020), Single and multiple cascading of VCRS in NH_3H_2O vapour absorption refrigeration systems for improving thermodynamic (energy-exergy) performances using five ecofriendly new HFOs and other low GWP refrigerants for replacing R134a, International Journal of Research in Engineering and Innovation, 4 (2), 96-104.

Cite this article as: R.S. Mishra, Performance evaluation of single effect Li/Br-H_2O vapour absorption refrigeration system with three cascaded vapour compression refrigeration systems using HFO refrigerants for ultra-low temperature applications, International journal of research in engineering and innovation (IJREI), vol 5, issue 1 (2020), 1-9.
https://doi.org/10.36037/IJREI.2021.5101

Computer Networks 201 (2021) 108585

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Performance optimization for UWB communication network under IEEE 802.15.4a channel conditions

Abhishek Sharma *, Amik Garg, Sanjay Kumar Sharma, Vibhav Kumar Sachan, Parvin Kumar

Electronics and Communication Engineering, KIET Group of Institutions, Delhi-NCR, Ghaziabad, Uttar Pradesh, India

ARTICLE INFO

Keywords:
IEEE 802.15.4a
MAN
Time hopping BPM

ABSTRACT

The manuscript presents a performance optimization methodology for UWB signaling under IEEE 802.15.4a channel conditions using pulse shaping. IEEE 802.1.5.4a is widely adapted for wireless personal area network (WPAN), body area network (BAN) as networking standard. The modified pulse shaping is specifically analyzed for the outdoor and industrial IEEE 802.15.4a channel conditions. The pulse shaping is also validated for the effect of interference levels offered to the co-existing Wireless MAN-SCa and Wireless MAN-OFDM (orthogonal frequency-division multiplexing) signaling. The performance has been evaluated for the Gaussian pulse shaping in a multi-user environment. The results for the pulse optimization have been validated for the line of sight and non-line of sight environment to investigate the capability of the UWB pulse for data transmission in multi user scenarios.

1. Introduction

UWB (Ultra Wideband) signals have the potential to provide accurate ranging capabilities over relatively finite distances with indoor and outdoor precision tracking applications for wireless networks due to the characteristic of precise time resolution. UWB communication also provides a viable alternative for high-speed data transfer over short distances while also allowing the user to monitor objects with pinpoint accuracy. Enormous bandwidth greater than 500 MHz and high multiple access capability makes UWB a potential candidate for many wireless applications. UWB is a carrier less technology which uses very short duration (Nano-seconds) pulses that makes it immune to multipath fading. Carrier less transmission gives an edge to UWB communication and makes it a low power, low cost hardware communication that makes it perfectly suitable for distributed sensor network [1]. This multi-path immunity of UWB provides a solution to take an advantage for detection and ranging application. With many advantages, UWB faces limitation due to the huge bandwidth. It causes interference with coexisting narrow-band communication. This limitation put an emission limit for the transmission of UWB. FCC (Federal Communication Commission) accepts the use of UWB for commercial purposes since 2002 for unlicensed use in the frequency range 3.1–10.6 GHz [2]. As UWB devices are intentional radiators, FCC part 15 rules are applicable for such communication which allows UWB devices to operate in a manner not to cause harmful interference to the other co-existing communication systems. Mostly, UWB communication is not affected by any other transmission but it affects the other co-existing narrow-band communication due to the large bandwidth [3]. Currently, ETSI (European Telecommunications Standards Institute), in the ECC (European Communication Committee) do not identify dedicated frequency bands for UWB application. FCC provide a spectrum mask shown in Fig. 1 for indoor transmission of UWB. As per the definition of UWB given by FCC, the bandwidth of UWB must be equal to or greater than 20% of the center frequency or bandwidth is greater than 500 MHz. Due to these major restriction of emission limits imposed on UWB, cannot allow the full utilization of this communication. Some methods has already been discussed to minimize this issue by changing the spectral behavior of the pulse in the frequency band of operation [4–13].

The adjustment of the PSD is compelled by the FCC spectral masking to mitigate signal interference with co-existing spectrum. In literature, authors [15–20] found that pulse shaping is an appropriate method for controlling the PSD as needed. The selection of proper window functions is a must to suppress the ringing that could negatively impact the PSD of the transmitted signal. The selection of the proper impact. Gaussian pulse shaping is an effective window function for reduction in the ringing using windowing [21–23]. In this constraint, the performance of the UWB systems is also affected negatively. To check the capability of any pulse modification, the verification of its capability in standard channel environment is required.

The IEEE 802.15.3a Saleh–Valenzuela and modified IEEE 802.15.4a channel models are the standards for UWB communication systems [21,

* Corresponding author.
E-mail addresses: ymhabhishek@gmail.com (A. Sharma), amikkgarg@gmail.com (A. Garg), drsanjaysharma15@gmail.com (S.K. Sharma), vibhavsachan@gmail.com (V.K. Sachan), parvin.kaushik@gmail.com (P. Kumar).

https://doi.org/10.1016/j.comnet.2021.108585
Received 16 June 2021; Received in revised form 21 September 2021; Accepted 25 October 2021
Available online 9 November 2021

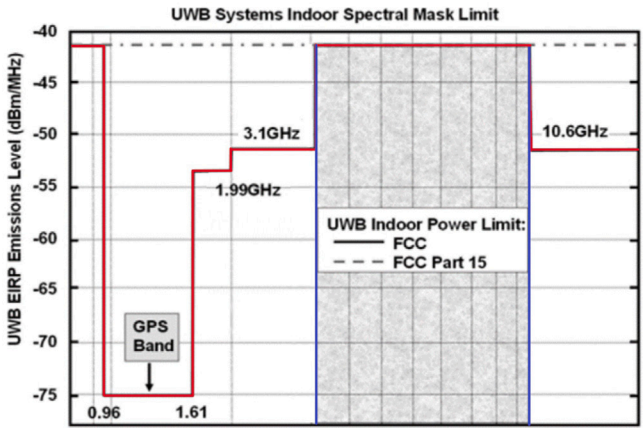


Fig. 1. FCC spectral mask for Indoor UWB communication [14].

24]. Several authors examined their proposals for IEEE 802.15.3a and IEEE 802.15.4a channel models to validate their theories for UWB network. The author of [25] proposed a pulse shaping method with a linear combination of pulse shapes and evaluated its performance for coherent receivers for industrial NLOS IEEE 802.15.4a channel conditions. In indoor personal area networks, the IEEE 802.15.3a channel model conditions are used for high data rate baseband signaling. While the IEEE 802.15.4a was chosen because of its adaptability regarding UWB applications for a broad range of low data rate personal area networks, especially in the fields of sensing, location monitoring, radar imaging, process automation, healthcare, etc. [26–29].

The aim of this work is to obtain the optimum performance of UWB communication for IEEE 802.15.4a outdoor and industrial channel conditions. This work also includes the minimization of interference levels faced by the Wireless MAN SCa and Wireless MAN OFDM transmissions by UWB co-existing transmissions. We have assessed the capability of modified pulse shaping in multi-user environments for aforesaid channel conditions and performance improvement. For each channel model, the performance has been evaluated using both the line of sight (LOS) and non-line of sight (NLOS) conditions. Time Hopping Bi-phase Modulation (TH-BPM) was chosen for UWB signaling in a multi-user environment to test the resilience of Gaussian windowed modified pulse shaping, as TH-BPM is more susceptible to high interferers in near-far scenarios and is most sensitive to time jitters [11]. The properties of various pulse formats for UWB and BPM (Bi-phase Modulation) have already been explored in the literature [13,30,31]. The obtained observations are validated by theoretical analysis and MATLAB simulations in the manuscript for a wide range of frequency operations (2–8 GHz). The comparative outcomes for the same parameters are shown to compare the results of the proposed models in the state-of-the-art literature [25,32,33], base pulse [14,34], and modified pulse shaping.

2. Mathematical modeling for the pulse shaping

In a multiple user environment, accurate localization can be accomplished in wireless communication with a high-performance device with a low bit error rate (BER). Thus Considering, this section consists a complete description of the proposed pulse that can help in multiple ways to achieve the objective of high performance and interference reduction with narrow-band communication. The Gaussian pulse and its derivatives drawn attention of many authors and these signals are widely used in many research papers because of advantage of timing jitter. As antenna has property of differentiation, the output of the transmitter antenna can be provided as derivative of the Gaussian pulse. In [8], the author provide the general expression for nth order derivative which can be obtained recursively from Eq. (1)

$$x^{(n)} = -\frac{n-1}{\varsigma^2}.x^{(n-2)}(t) - \frac{t}{\varsigma^2}x^{(n-1)}(t) \tag{1}$$

From Eq. (1), the 7th order derivative of Gaussian monocycle waveform and the modified pulse with Gaussian windowing is plotted in Fig. 2. Eq. (2) is revealing the maths for the basis pulse waveform with its spectrum presented in Eq. (3).

$$p(t) = K\left(\frac{105t}{\varsigma^8} - \frac{105t}{\varsigma^{10}} + \frac{21t^5}{\varsigma^{12}} - \frac{t^7}{\varsigma^{14}}\right).exp\left(-\frac{t^2}{2\varsigma^2}\right) \tag{2}$$

$$|P(f)| = K(2\pi f)^7.exp\left(-\frac{(2\pi\varsigma f)^2}{2}\right) \tag{3}$$

where K is an scaling coefficient, ς is related with 10 dB RF bandwidth, $\varsigma = \frac{1}{2\pi f_B\sqrt{log_{10}e}}$ [35]. Antenna differentiation characteristics causes undesired damping amplitude oscillations that can impact severely negative on antenna power resolution [6,36–39]. To achieve the desired performance, these oscillations must be suppressed as the base pulse get amplitude modulated by Gaussian window function. The spectrum of the modified pulse is mathematically presented in Eq. (4).

$$G(f) = K.\int_{-\infty}^{\infty}(2\pi\tau)^7 e^{-\frac{(2\pi\varsigma\tau)^2}{2}} e^{-\frac{(2\pi\sigma_1(f-\tau))^2}{2}} d\tau \tag{4}$$

Eq. (4) is presenting the frequency domain convolution of the base pulse with Gaussian pulse. Fig. 2 is presenting both the pulses for comparison. A comparison for the magnitude spectrum between the modified pulse shaping with different pulse shaping factor (PSF) σ_1 values and base pulse is revealing in Fig. 3(a)

The modified pulse is used to modulate with digital bit stream. Corresponding to digital bit stream, the magnitude spectrum for the modified pulse is compared for its different values of PSF and the base pulse, which is revealing in Fig. 3(a). Time hopped Bi-phase (antipodal) UWB (TH-BPM UWB) methodology is used for transmission of N_f pulses per bit. The received signal is expressed as [13,40] in Eq. (5)

$$r(t) = \sum_k \sqrt{\epsilon_k}a_i \sum_{l=0}^{N_f} g'(t - kN_fT_f - lT_f - c_iT_c) \tag{5}$$

ϵ_k is the energy of the kth symbol, a_k is the NRZ digital information symbol in format taking values ±1 with equal probability. $g'(t)$ is the inverse continuous time Fourier transform of $G'(f)$ and the modified pulse shape mentioned in Eq. (5). N_f is the number of frames per symbol, T_f is the frame period, T_c is the chip period, and $\{c\}_{l=0}^{N_f}$ is the time hopping sequence with integer element spacing ranges as $0 < c_j < N_h$. c_l T_c is the time shift in seconds to the jth pulse of the signal, c_j (for different values of j) give the element of PN sequence (pseudo-random sequence). The overlapping of consecutive pulses is avoided by assuming $T_f/T_c = N_c$, where N_c is the number of chips per frame. The minimum criteria that satisfy for T_f is $T_f \geq (c_{N_f} - c_0)T_c + \tau_{L-1,0} + T_p$ [41]. The TH-BPM UWB signal is shown in Fig. 3(b). The block diagram of transmitter and receiver module is presented in Figs. 4 and 5 The power spectral density (PSD) of the M-ary modulation can be the sum of continuous and discrete spectral components. All the aspect corresponding to the behavior of the modified pulse shaping has been discussed previous literature [9,23,24,40,42,43]. Tables 1 and 2 is presenting the PSD levels for the modified pulse shaping [23,43] with different values of pulse shaping factors (PSF) and corresponding center frequencies.

3. Channel model

The most popular standard channel model for UWB communication is IEEE 802.15.4a. For the UWB communication, model is detailed in literature [1,21,40,41,44,45]. Assuming a uniform distribution of

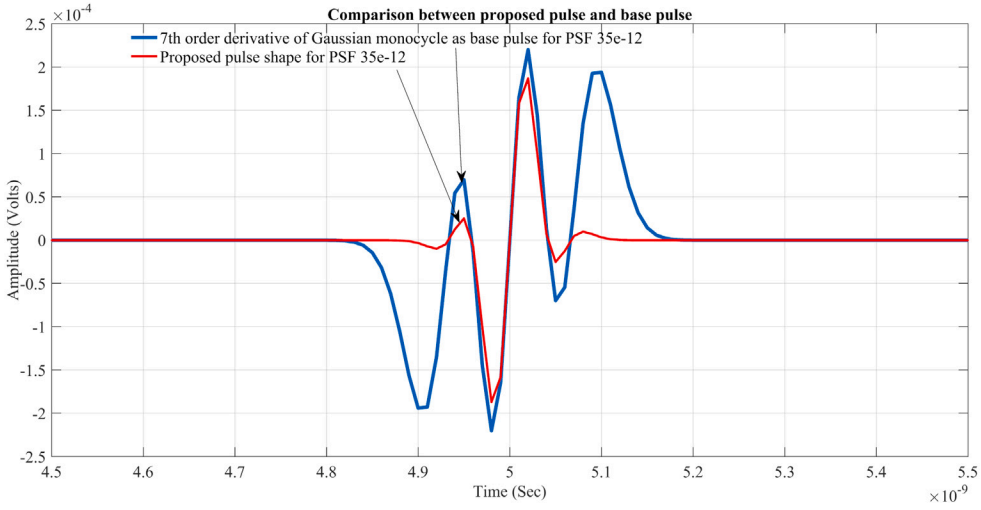


Fig. 2. Comparison between basis pulse and modified pulse shaping [23].

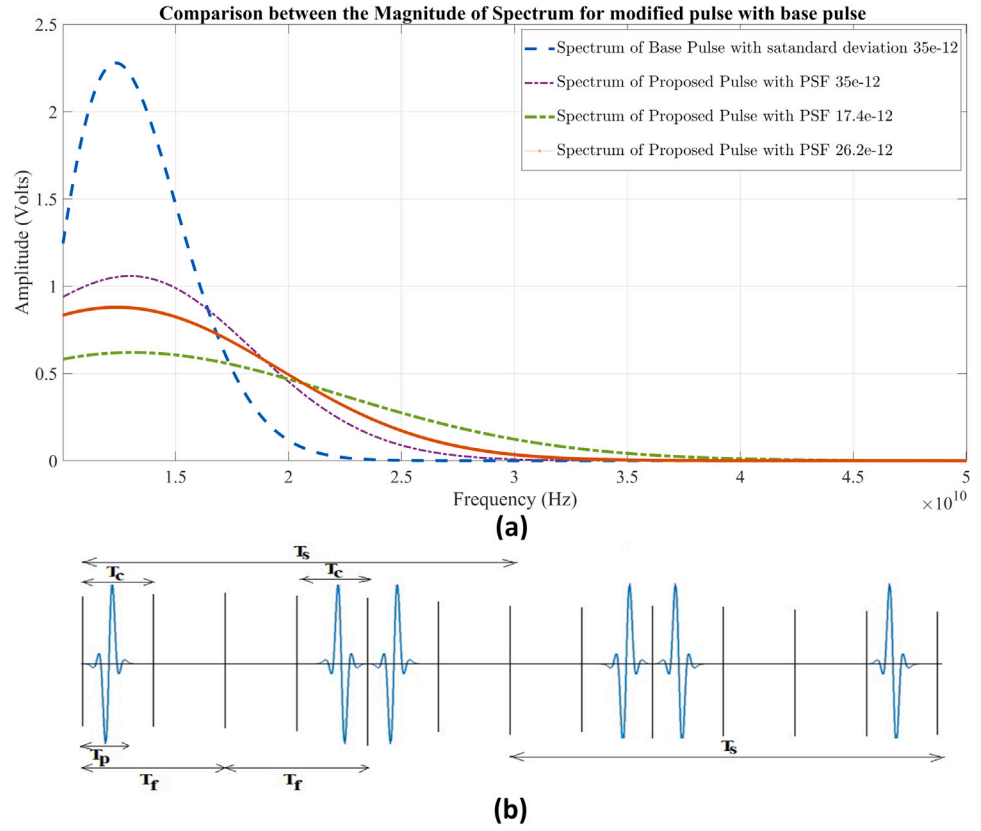


Fig. 3. (a) Comparison between spectrum of basis pulse and modified pulse shaping, (b) TH-UWB Signal with BPM modulation [23].

Table 1
PSD of the transmitted BPM modulated Pulse Train using modified pulse for different PSF values.

PSD for different values of PSF σ_1						
Spectral range	$\sigma_1 = 1.75 \times 10^{-10}$ s	$\sigma_1 = 1.05 \times 10^{-11}$ s	$\sigma_1 = 35 \times 10^{-12}$ s	$\sigma_1 = 26.3 \times 10^{-12}$ s	$\sigma_1 = 17.5 \times 10^{-12}$ s	$\sigma_1 = 8.75 \times 10^{-12}$ s
3 GHz–5 GHz	−62 dBm/MHz	−69 dBm/MHz	−72 dBm/MHz	−94 dBm/MHz	−114 dBm/MHz	−129 dBm/MHz
5 GHz–7 GHz	−64 dBm/MHz	−71 dBm/MHz	−72 dBm/MHz	−95 dBm/MHz	−114 dBm/MHz	−130 dBm/MHz
7 GHz–10.6 GHz	−64 dBm/MHz	−71 dBm/MHz	−73 dBm/MHz	−95 dBm/MHz	−115 dBm/MHz	−130 dBm/MHz

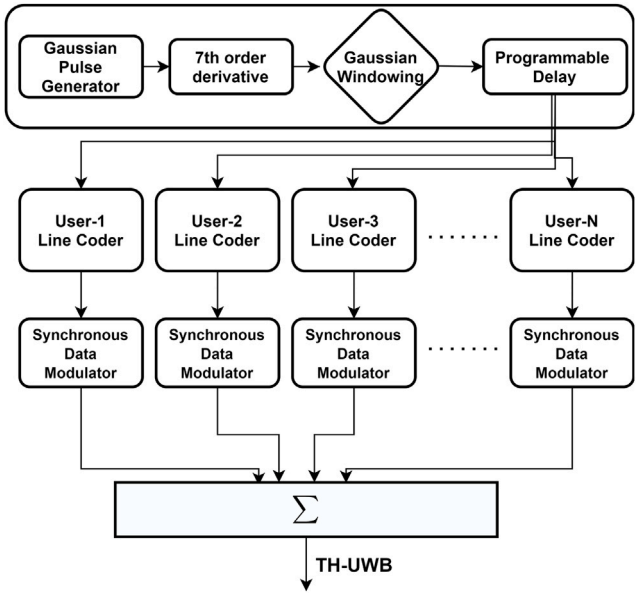


Fig. 4. Transmitter block diagram for Multi-user TH-BPM UWB.

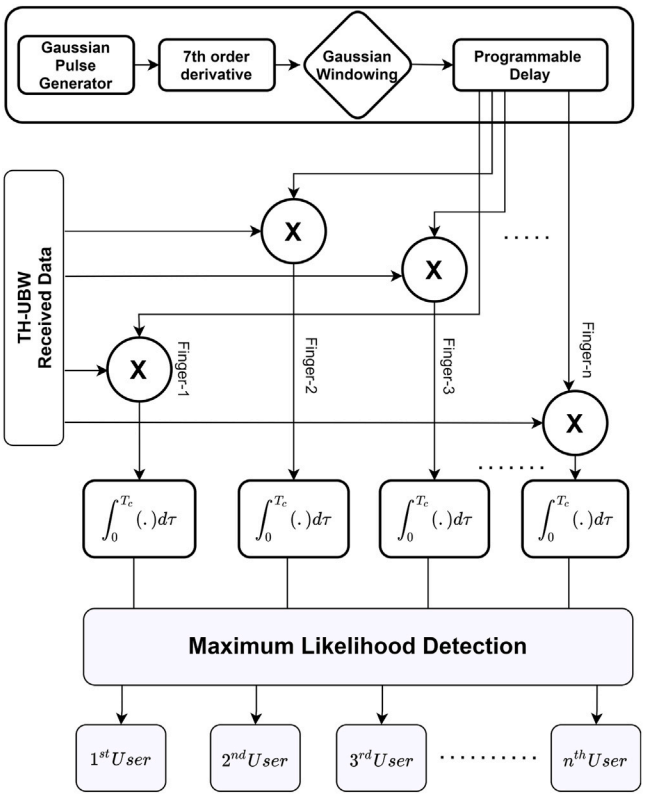


Fig. 5. Receiver block diagram for Multi-user TH-BPM.

symbols with no frame interference for TH-BPM, the performance of the modified pulse shaping has been validated for the IEEE 802.15.4a model. In the aforementioned channel model the tap amplitude statistics are effectively Rayleigh distributed, whereas it was log-normally distributed in IEEE 802.15.3a. The modified Saleh–Valenzuela IEEE 802.15.4a multi-path channel model can be expressed as in (6)

$$h_{discr}(t) = \sum_{k=0}^{L}\sum_{k=0}^{K} \alpha_{k,l} e^{j\phi_{k,l}} \delta(t - T_l - \tau_{k,l}) \tag{6}$$

Table 2
Spectral information for the transmitted sequence using modified pulse.

PSF σ_1 (s)	Approx. Value of RMS for effective signal bandwidth(−10 dB FCC)	Approx. upper frequency	Center frequency
1.75×10^{-10}	19.28 GHz	22 GHz	9.64 GHz
1.05×10^{-11}	21.79 GHz	25 GHz	10.895 GHz
35×10^{-12}	22.18 GHz	30 GHz	11.09 GHz
26.3×10^{-12}	26.2 GHz	42 GHz	13 GHz
17.5×10^{-12}	30.1 GHz	44.8 GHz	15.05 GHz
8.75×10^{-12}	32.54 GHz	50 GHz	16.27 GHz

$\alpha_{k,l}$ is presenting the multi-path gain coefficients or tap weights of the kth and T_l is the delay of the lth cluster respectively. $\tau_{k,l}$ is the delay of the kth multi path component (MPC) relative to the lth cluster arrival time T_l. The phase $\phi_{k,l}$ is a uniformly distributed random variable with a closed range $[0, 2\pi]$. Parameter L follows Poisson Distribution with probability density function(pdf) as revealed in Eq. (7).

$$P_{Poisson}(L) = \frac{(\bar{L})^L}{L!} e^{-\bar{L}} \tag{7}$$

The $\bar{L}$ reveal mean or expectation value of L. Defining $\tau_{(k,l)} = 0$, the distribution of the cluster arrival times is given by a Poisson processes.

$$p(T_l|T_{l-1}) = \Lambda_l e^{-\Lambda(T_l - T_{l-1})}, T_l > T_{l-1}, l > 0 \tag{8}$$

In the Eq. (8), the term Λ_l is the cluster arrival rate (assumed to be independent of l). Poisson process describe ray arrival time for classical Saleh–Valenzuela model. This classical method of Eq. (8) is not perfectly suited for the indoor residential, industrial, and office environments. The authors of [21,24] proposed the modified model ray arrival times with mixtures of two Poisson processes as presented in Eq. (9)

$$\begin{aligned} p(\tau_{k,l}|\tau_{k-1,l}) = {} & \varepsilon\lambda_1 exp[-\lambda_1(\tau_{k,l}|\tau_{k-1,l})] + \\ & (\varepsilon - 1)\lambda_2 exp[-\lambda_2(\tau_{k,l}|\tau_{k-1,l})] \end{aligned} \tag{9}$$

In Eq. (9), ε is the mixture probability, λ_1 and λ_2 are the ray arrival rates. The cluster power and shape are the two important parameters for power delay profile (PDP). Within each cluster, the power delay profile (PDP) is exponential for the different arrival path [21,24] as presented in Eq. (10)

$$E\left\{|\alpha_{k,l}|^2\right\} = \Omega_l \frac{1}{\gamma_l[(1-\varepsilon)+\lambda_1] + \varepsilon\lambda_2 + 1} exp\left(-\frac{\tau_{k,l}}{\gamma_l}\right) \tag{10}$$

The term Ω_l is the integrated energy of the lth cluster, and γ_l is the intra-cluster decay time constant. The normalized approximation works for both typical values of λ and γ. The cluster decay rates are directly proportional to the arrival time of the cluster (as presented in Eq. (11))

$$\gamma_l \propto k_\gamma T_l + \gamma_0 \tag{11}$$

The term k_γ is denoting the increase of the decay constant with delay. The mean (over the cluster shadowing), mean (over the small-scale fading) energy (normalized to) γ_l [21,24], of the lth cluster generally follows exponential decay as revealed in Eq. (12)

$$10log(\omega_l) = 10log\left(exp\left(-\frac{T_l}{\Gamma}\right)\right) + M_{cluster} \tag{12}$$

$M_{cluster}$ is a normal/Gaussian distribution with a variance $\sigma cluster^2$ around it. The PDP follows log normal scale. For the NLOS (Non line of sight), it is equated as

$$\begin{aligned} E\left\{|\alpha_{k,l}|^2\right\} = {} & \frac{\gamma_1+\gamma_{rise}}{\gamma_1} \cdot \frac{\Omega_1}{\gamma_1+\gamma_{rise}(1-\chi)} \times \\ & \left(1 - \chi exp\left(-\frac{\tau_{k,l}}{\gamma_{rise}}\right)\right) . exp\left(-\frac{\tau_{k,l}}{\gamma_{rise}}\right) \end{aligned} \tag{13}$$

The parameter χ describes the attenuation of the first component, the parameter γ_{rise} determines how fast the PDP increases to its local

Table 3
Multipath channel target characteristics and model parameters of standard modified SV model.

Channel model for 2–10 GHz UWB band				
	Outdoor		Industrial	
	LOS	NLOS	LOS	NLOS
Path loss				
n	1.76	2.5	1.2	2.15
σ_s [dB]	0.83	2	6	6
PL_0 [dB]	45.6	73	56.7	56.7
A_{ant}	3 dB	3 dB	3 dB	3 dB
κ	0.12	0.13	−1.103	−1.427
Power Delay Profile (PDP)				
$\bar{L}$	13.6	10.5	4.75	1
Λ [1/ns]	0.0048	0.0243	0.0709	NA
λ_1	0.27	0.15		
	2.41	1.13	NA	NA
β	0.0078	0.062		
Γ [ns]	31.7	104.7	13.47	NA
k_γ	0	0	0.926	NA
γ_0 [ns]	3.7	9.3	0.651	NA
$\sigma_{cluster}$ [dB]	–	–	4.32	NA
Small-scale fading				
m_0 [dB]	0.77	0.56	0.36	0.3
k_m	0	0	0	0
$\hat{m_0}$ [dB]	0.78	0.25	1.13	1.15
$\tilde{m_0}$	–	–	12.99	–
χ	NA	NA	NA	1
γ_{rise}	NA	NA	NA	17.56
γ_1	NA	NA	NA	85.36

maximum, and γ_1 determines the decay at last times. The probability distribution for small scale amplitude fading is Nakagami, as equated,

$$\mathbb{N}(y) = \frac{2}{\Gamma(m)}\left(\frac{m}{\Omega}\right)^m y^{2m-1}.exp(\frac{-my^2}{\Omega}) \tag{14}$$

The term $m \geq \frac{1}{2}$ is the Nakagami m – factor, $\Gamma(m)$ is the gamma function, and Ω is the mean-square value of the amplitude. m is considered to be an independent with delay. The Ω corresponds to the mean power and the delay dependence is given by power delay profile. To obtain optimistic results, a conversion of Nakagami distribution to Rician distribution is approximately possible with the conversion equations are as follows

$$m = \frac{(K_r+1)^2}{[2(K)_r]+1} \tag{15}$$

$$K_r = \frac{\sqrt{m^2-m}}{m-\sqrt{(m^2-m)}} \tag{16}$$

where K_r and m are the Rice factor and Nakagami-m factor respectively. The parameter m is a random variable with log-normally distributed probability. The mean μ_m and standard deviation σ_m are delay dependent, as expressed in Eqs. (17) & (18)

$$\mu_m(\tau) = m_0 - k_m\tau \tag{17}$$

$$\sigma_m(\tau) = \hat{m_0} - k_m\tau \tag{18}$$

As the Nakagami factor m is modeled differently for each cluster and assumed to deterministic and independent of the delay [21], $\hat{m_0} = m$. The model parameters extracted based on measurement for different environment has been summarized below in Table 3 [21,27,46]

At the receiver, the received signal is expressed as [47] in Eq. (19)

$$\mathbb{R}(t) = \sum_{l=0}^{L}\sum_{k=0}^{K}\alpha_{k,l}exp(j\phi_{k,l})\delta(t-T_l-\tau_{k,l})\bigotimes \sum_i \sqrt{\varepsilon_i a_i}\sum_{j=0}^{N_f}k'(t-iN_fT_f-jT_f-c_jT_c)+n_i(t) \tag{19}$$

The term $n_i(t)$ is considered as Additive White Gaussian Noise (AWGN) with two-sided PSD $\frac{N_0}{2}$ and noise samples are presumed to be independent. In practice, these values may not be autonomous and may have an impact on the results. $\bigotimes$ denotes convolution symbol. The receiver structure for an Time Hopped Bipolar antipodal modulation is well explained in [48]. Transmitter and receiver for TH-UWB is shown in Figs. 4 and 5 respectively. Correlative synchronous receiver with Maximum likelihood (ML) estimation detection method as explained in [48], is used for data detection. For this template signal algorithm for data detection, the signal to noise ratio (SNR) is equated as in Eq. (20) [48]

$$SNR_{ML}(N,\alpha) = \frac{\frac{(1-\alpha)\varepsilon^2}{N_d\|h\|^4}}{\left|1+\frac{3(1-\alpha)}{N_d}\right|\varepsilon\sigma^2\|h\|^2+N\sigma^4\phi_2} \tag{20}$$

The term $\phi = 1+\frac{1-\alpha}{N_d(N+1)}$. The relation of $SNR_{ML}(N,\alpha)$ with effective SNR as equated in Eq. (21)

$$\rho(N,\alpha) = \frac{1}{SNR_{ML}^{-1}-\frac{N\sigma^4(2+N)}{\varepsilon^2}\|h\|^4}+\frac{\frac{1-\alpha}{N_d(2L\sigma^2\|h\|^2+N^2\sigma^4)}}{\left(1+\frac{3(1-\alpha)}{N_d}\right)\varepsilon^2\sigma^2\|h\|^2+N\sigma^4(1-\frac{1-\alpha}{N_d})} \tag{21}$$

An approximated value of the BER performance of the receiver in presence of AWGN expressed in Eq. (22)

$$BER_{Avg} = \frac{1}{2\pi}\int_{\sqrt{\rho_{ML}(L,\alpha)}}^{\infty}e^{-\frac{u^2}{2}}du = Q(\sqrt{\rho_{ML}(L,\alpha)}) \tag{22}$$

The Q function is a popular math function related with error function as $Q(x) = \frac{1}{2}erfc\left(\frac{x}{\sqrt{2}}\right)$. α is the energy allocation factor, h is the sampled version of the channel response with $L = WT_{ds}$ number of samples, W is the bandwidth of the received signal, T_{ds} is the constant delay spread of the channel. N is the number of samples taken at a rate of T_s seconds of signal of the received signal $\mathbb{R}$. N_s total number of symbols, N_t are the known number of symbols, the unknown symbols $N_d = N_t - N_s$, ϕ_2 is the moment generating function [48]. As the PSF values are very lower comparably in a nanosecond, so the chance of inter-frame interference is also not considered for simulations.

4. Interference minimization for co-existing networks

In the manuscript we have analyzed the modified pulse shaping for interference scenario. We have evaluated the impact of the transmitted signal on co-existing receivers. The investigation is done for IEEE 802.16 standards such as Wireless-MAN-SCa and Wireless MAN-OFDM for coding rate 1/2 and 3/4. The maximum interference margin I_{Max} is a function of Thermal noise at receiver, N_f also depend upon the signal to average noise ratio [12,49,50].

$$N_{Noise} = -144+10log_{10}(BW_{Rx})+N_f \tag{23}$$

If the medium's propagation loss is avoided, then the ith UWB device's interfering capacity at the target receiver can be expressed as [12]

$$I_{UWB,i} = \int_{f_{low}}^{f_{high}} S_{UWB}(f)df \tag{24}$$

Transmitted signal power for UWB communication will be faced by the victim's receiver, then the PSD at this receiver will be provided by

$$S_{Rx}(f) = \frac{A_f G_{Tx}G_{Rx}S_{UWB}(f)}{P_{th}(r)} \tag{25}$$

In the Eq.(25), G_{Tx} and G_{Rx} are the antenna gain parameters for transmitter and victim's receiver respectively for omni-directional antennas. The activity function, Af, reflects the number of UWB devices that are operational at any given instant of time. $P_{th}(r)$ is a range dependent

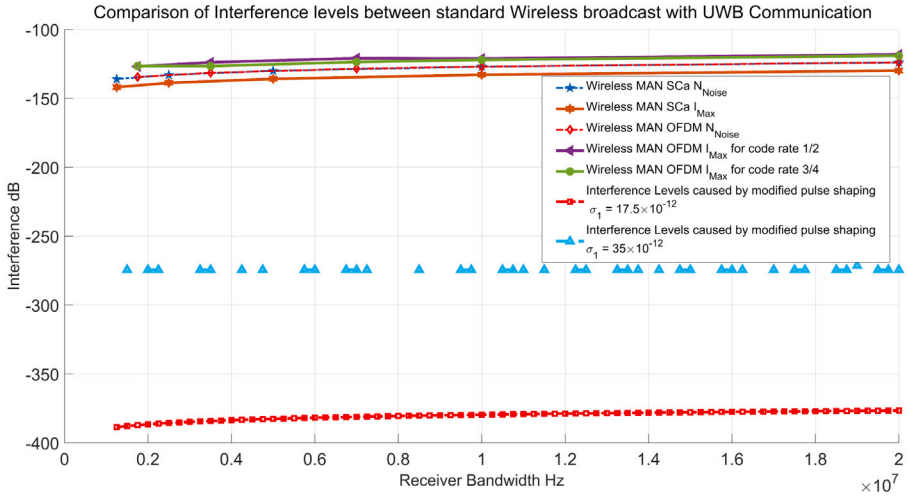


Fig. 6. A comparison for interference levels between IEEE802.16 standards and modified pulse shaping.

path loss parameter. For indoor NLOS communication penetration loss L_p considered as 10 dB. $P_{th}(r) = (\frac{4\pi}{\lambda})^2 L_p r^2$. Let $a_0 = (\frac{4\pi}{\lambda})^2$. If whole area of the target room under study is A and $S_{(Tx,i)}$ is the power transmitted from ith transmitter, then for n number of devices the average transmit power will be expressed as in (26)

$$\varrho_I = \frac{\sum_{i=1}^{n} S_{Tx,i}}{A} \tag{26}$$

In reference to the (24), the interference faced by the target receiver can be equated as in (27)

$$I = \int_{f_{low}}^{f_{high}} \int_{S} S_{Rx}(f) ds df = \int_{f_{low}}^{f_{high}} \int_{S} \frac{A_f G_{Tx} G_{Rx} S_{UWB}(f)}{a_0 L_p r^2} ds df \tag{27}$$

Considering the area R which includes UWB devices. Assuming $ds = rdrd\psi$. If the maximum and minimum distance between the target receiver and UWB transmitter are r_{max} and r_{min} then...

$$I = \int_{f_{low}}^{f_{high}} \int_{\psi=0}^{2\pi} \int_{r=r_{min}}^{r=r_{max}} \frac{A_f G_{Tx} G_{Rx} S_{UWB}(f)}{a_0 L_p r^2} r dr d\psi df \tag{28}$$

$$I = \int_{f_{low}}^{f_{high}} \int_{\psi=0}^{2\pi} \int_{r=r_{min}}^{r=r_{max}} \frac{A_f G_{Tx} G_{Rx} S_{UWB}(f)}{a_0 L_p r^2} r dr d\psi df \tag{29}$$

Assuming $K_1 = \frac{A_f G_{Tx} G_{Rx}}{a_0 L_p}$. The PSD of the modified pulse as presented in (4), will be given by $S_{UWB}(f) = |G(f)|^2$.

$$I = K K_1 ln(\frac{r_{max}}{r_{min}}) \int_{f_{low}}^{f_{high}} \left| \int_{-\infty}^{\infty} (2\pi\tau)^7 e^{-\frac{(2\pi\varsigma\tau)^2}{2}} e^{-\frac{(2\pi\sigma_1(f-\tau))^2}{2}} d\tau \right|^2 df \tag{30}$$

For the Wireless-MAN-SCa NLOS communication the maximum tolerance I_{Max} interference mode for its different licensed bandwidth (BW) is presented [49,51] in Table 4 Similarly for the comparison aspects, the maximum sustainable interference for wireless standard IEEE802.16a (Wireless MAN-OFDM) is also presented in Table 5.

The performance assessment is split into two goals in order to determine the novelty of the modified pulse shaping. The first objective is to assess the performance by examining pulse behavior in a standard outdoor and industrial IEEE 802.15.4a scenario. The second goal is to assess the impact of interference offered by UWB communication with co-existing networks for the modified pulse for data modulation.

Table 4
Maximum allowable interference standards for Wireless-MAN-SCa with various licensed bandwidth [12,49].

Licensed bandwidth					
	1.25 MHz	2.5 MHz	5 MHz	10 MHz	20 MHz
N_{Noise} [dBW]	−136	−133	−130	−127	−124
I_{Max} [dBW]	−141.9	−138.8	−135.9	−132.9	−129.9

Table 5
Maximum allowable interference for Wireless MAN-OFDM for various channelization and code rate values [12,49].

Licensed bandwidth					
	1.75 MHz	3.5 MHz	7 MHz	10 MHz	20 MHz
N_{Noise} [dBW]	−134.5	−131.5	−128.5	−127.0	−123.9
I_{Max} for rate = $\frac{1}{2}$ [dBW]	−126.9	−123.9	−120.9	−121.18	−118.2
I_{Max} for rate = $\frac{3}{4}$ [dBW]	−129.6	−126.6	−123.6	−122.1	−119.1

5. Results

The performance evaluation for the modified pulse has been performed using MATLAB simulations. Fig. 6 is revealing the obtained results for the modified pulse shaping. To achieve the first goal, a comparison has been performed with reference values from Tables 4 and 5 for wireless-MAN-SCa and wireless MAN-OFDM respectively. The investigation is carried out for $\sigma_1 = 17.5\times10^{-12}$ and $\sigma_1 = 35\times10^{-12}$ PSF values of 20 UWB devices with a 50% activity factor and a 10 m by 10 m free space area to acquire the result. The obtained results clearly reveal that the performance of the modified pulse shaping for the UWB signaling is excellent. As the used modified pulse shaping offers a very low interference levels to certain most popular co-existing wireless communication standard such as Wireless MAN-SCa and Wireless MAN-OFDM.

To fulfill the second goal, the performance of modified pulse shaping has analyzed for the outdoor and industrial (LOS & NLOS) channel model IEEE 802.15.4a with multi-user transmission as discussed in Section 3. The environment condition is modeled for $L = 50$ per Channel Response h_{discr} and $\alpha = 0$. For the simulations, the channel modeling is done with the help of parametric values as presented in Table 3. For the multi-user transmission the value PSF σ_1 is 35 pico-sec for the Gaussian window function to obtained SNR vs. BER. The pulse width is 0.25 ns. A random binary sequence a_j with equal probability of 1's and 0's and variable data capacity from 0.25 Gbps to 20 Gbps

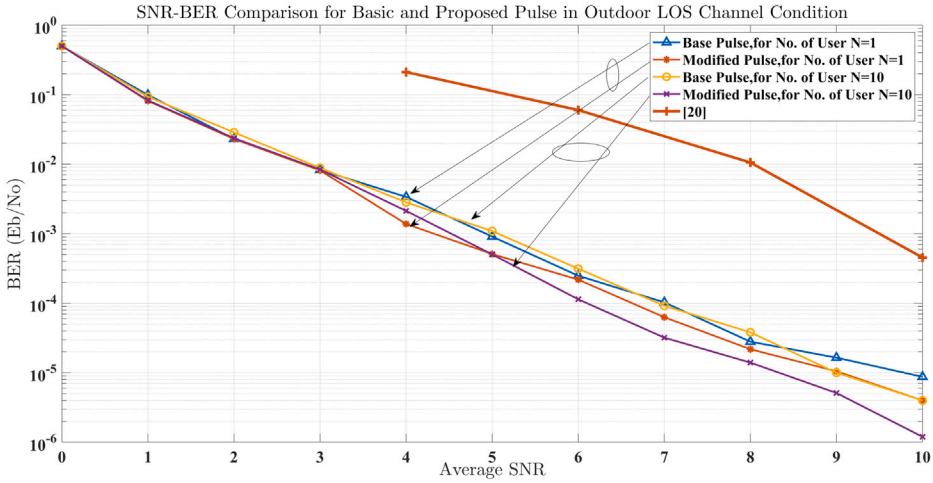


Fig. 7. SNR vs. BER comparison between 7th order derivative base pulse and [33] with proposed pulse in Outdoor LOS for number of users Nu = 1 and Nu = 10.

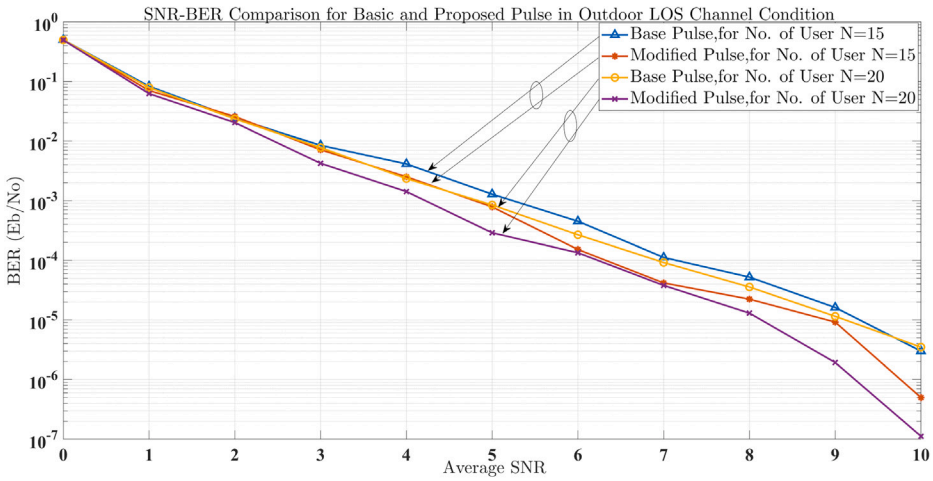


Fig. 8. SNR vs. BER comparison between 7th order derivative base pulse with proposed pulse in Outdoor LOS for number of users Nu = 15 and Nu = 20.

was used for the simulations. The parametric values are same for every simulated comparison. The frame period T_f is 440 ns, $N_f = 1$, T_c is 1 nsec. The time hopping code sequence is randomly selected within the interval $0 \leq c_j \leq 24$ and sampling time interval T_s is .01 ns.

We have presented multiple comparison of $\left(\frac{E_b}{N_0}\right)$ average SNR with BER for number of users(Nu) Nu = 1, Nu = 10, Nu = 15 and Nu = 20. Figs. 7 and 8 are revealing an SNR vs. BER comparison between modified pulse and base pulse for outdoor LOS. Figs. 9 and 10 are revealing an SNR vs. BER comparison between modified pulse and base pulse for outdoor NLOS. From these figures, it is clear that the modified pulse is showing an edge in performance while handling the multi-user signal traffic. In comparison to prove the ability of the modified pulse, we have also compared the obtained results with [33]. In [33], author proposed a solution to improve SNR vs. BER for outdoor LOS and NLOS environment with the least delay spread obtained for perfect synchronize receiver. This presented comparison of Figs. 7 and 9 has been done for the best output achieved by authors [33] with single user transmission and modified pulse shaping with multi-user transmission. The comparison results clearly reveal the performance improvement of the UWB system for the modified pulse shaping in the case of a multi-user outdoor LOS and NLOS scenario.

A similar comparison has been performed for coherent rake receivers in an industrial LOS IEEE 802.15.4a environment with 12 and 96 fingers respectively by [32]. In [32], the authors suggested that, with the increment in the number of fingers in a synchronous receiver can be able to provide the optimum performance for UWB communication. The presented comparison in Fig. 12 has been performed for the optimum results obtained by [32] with system performance for modified pulse shaping. From Fig. 12, the comparison clearly reveals that for similar values of average SNR, the BER values are comparatively much lower for the modified pulse shaping. Fig. 12 is revealing that for the average SNR value “4” and “6”, the improvement in BER is approximately “92%” and “96%” respectively. This clearly indicates the performance improvement of the concerned UWB communication for the modified pulse shaping. The performance of modified pulse shaping has also been evaluated for the industrial channel environment. Figs. 11, 12 are affirming an SNR vs. BER comparison between modified pulse and base pulse for industrial LOS. Similarly, Figs. 13 and 14 are also unfolding the SNR vs. BER comparison for the industrial NLOS. The conceptual simulations are revealing that the systems’ performance is improved for the modified pulse shaping with a reduction in BER values in comparison to its without windowing base pulse. Fro per simulation results, it must be conceded that modified pulse shaping would be the most popular method for obtaining optimal performance of UWB communication systems for outdoor and Industrial multi-user environments.

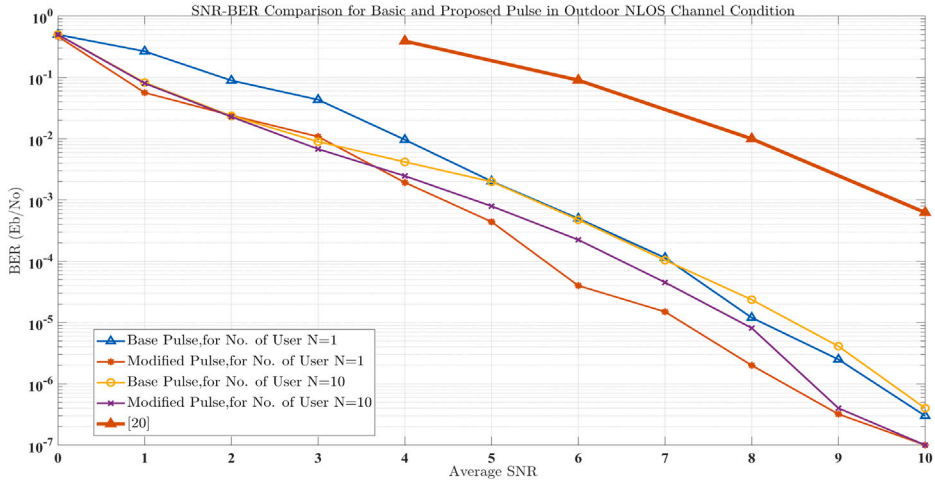


Fig. 9. SNR vs. BER comparison between 7th order derivative base pulse and [33] with proposed pulse in Outdoor NLOS for number of users Nu = 1 and Nu = 10.

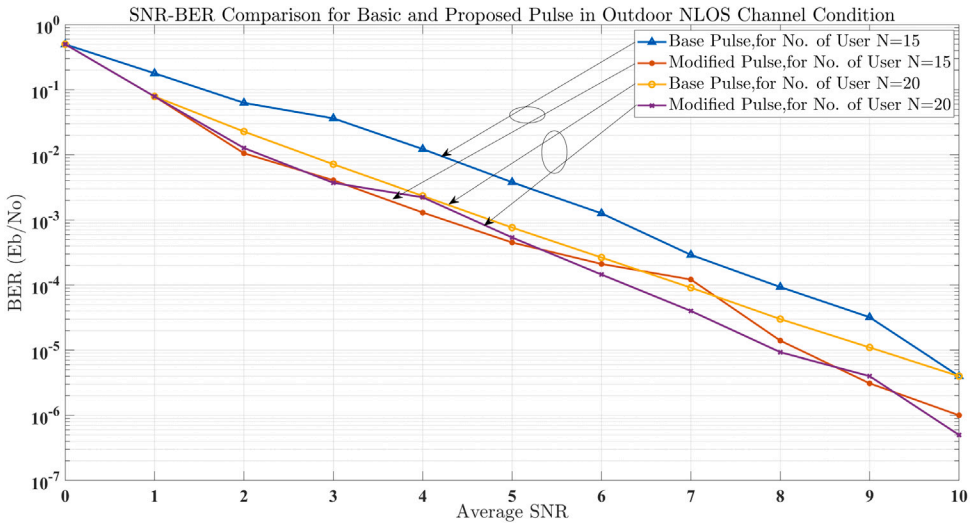


Fig. 10. SNR vs. BER comparison between 7th order derivative base pulse with proposed pulse in Outdoor NLOS for number of users Nu = 15 and Nu = 20.

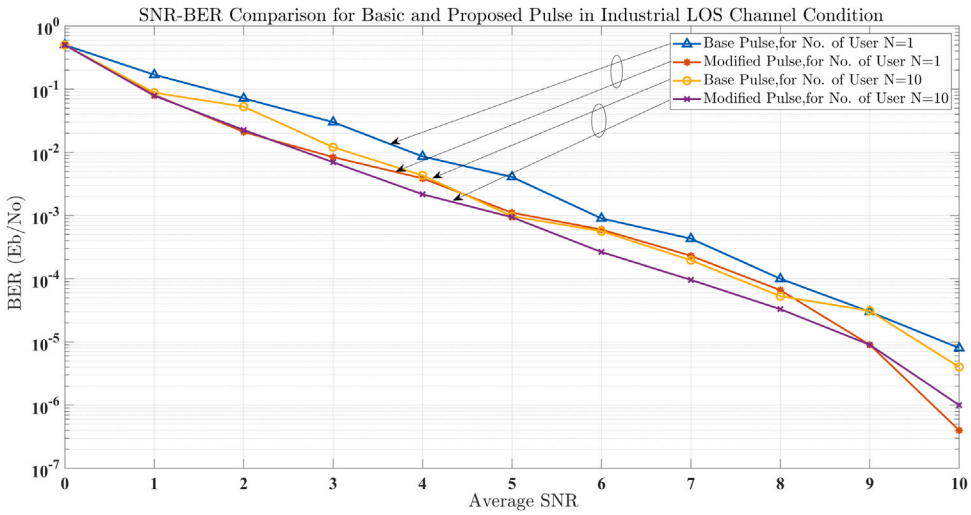


Fig. 11. SNR vs. BER comparison between 7th order derivative base pulse with proposed pulse in Industrial LOS for number of users Nu = 1 and Nu = 10.

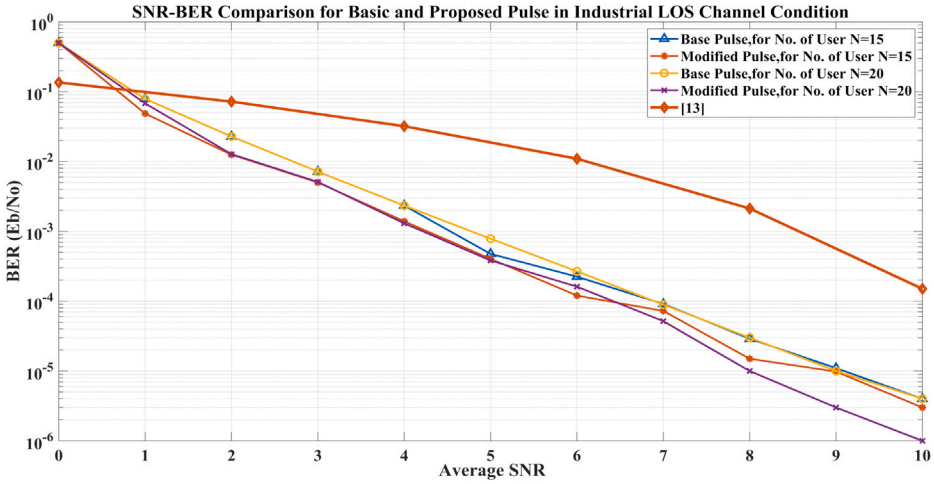


Fig. 12. SNR vs. BER comparison between 7th order derivative base pulse and [32] with proposed pulse in Industrial LOS for number of users Nu = 15 and Nu = 20.

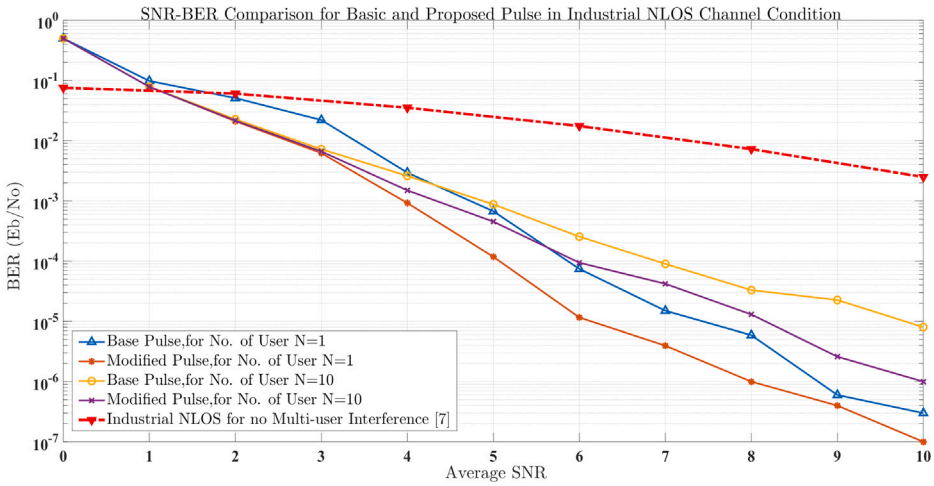


Fig. 13. SNR vs. BER comparison between 7th order derivative base pulse and [25] with proposed pulse in Industrial NLOS for number of users Nu = 1 and Nu = 10.

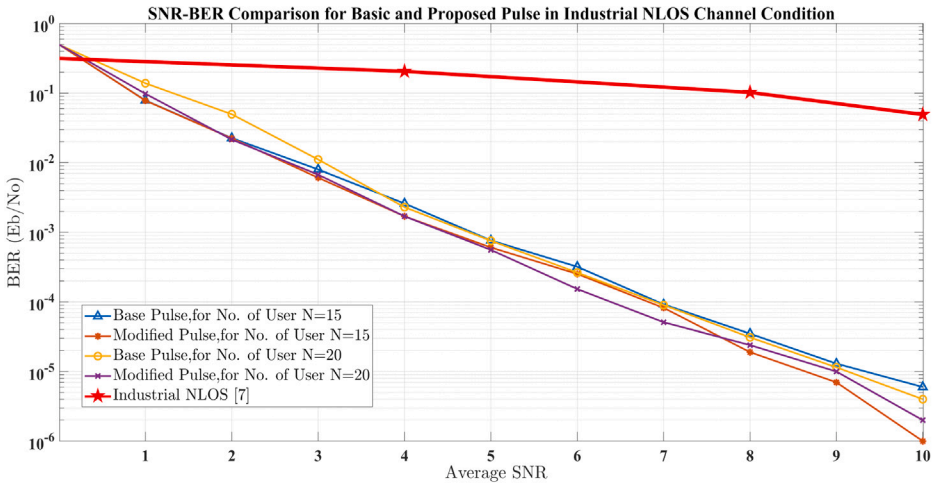


Fig. 14. SNR vs. BER comparison between 7th order derivative base pulse and [25] with proposed pulse in Industrial NLOS for number of users Nu = 15 and Nu = 20.

Another comparison is also revealing in Figs. 13 and 14 for Industrial NLOS IEEE 802.15.4a channel conditions without and with multi-user interference respectively. Figs. 13 and 14 clearly revealing that, in this case also, the behavior of the system performance for the modified pulse shaping is again superior in comparison to [25]. The comparison shows that in presence of multi-user transmission, the

performance of the modified pulse shaping is better in comparison to results with single user interference by [25].

6. Conclusion

In the manuscript, the novelty of the modified pulse shaping has been analyzed in two aspects. First, modified pulse shaping is offering a very low level of interference to its co-existing networks such as Wireless MAN-SCa and Wireless MAN-OFDM. This indicates that the modified pulse would offer a very low level of interference to other co-existing networks which offer a capability to UWB systems to simultaneously exist with the most popular wireless networks.

Second, the obtained results are clearly revealing that by controlling the pulse shaping factor for the pulse, the user might attain particular desired performance as per the required specifications. The BER versus SNR comparison results demonstrate that this optimal modified pulse shaping improves the performance of the UWB system, which improves the efficiency of transmission for multiple users in outdoor and industrial IEEE 802.15.4a channel conditions with AWGN. The comparison results demonstrate that the modified pulse shaping can improve the efficiency of the related UWB applications. The performance of the modified pulse shaping for low rate IEEE 802.15.4z-2020 channel conditions will consider for future scope of work. The effect of modified Gaussian window pulse shaping on antenna behavior and its performance might be a future field of research.

CRediT authorship contribution statement

Abhishek Sharma: Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Drafting the manuscript, Revising the manuscript critically for important intellectual content. **Amik Garg:** Drafting the manuscript. **Sanjay Kumar Sharma:** Drafting the manuscript, Revising the manuscript critically for important intellectual content. **Vibhav Kumar Sachan:** Analysis and/or interpretation of data, Drafting the manuscript. **Parvin Kumar:** Revising the manuscript critically for important intellectual content.

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.

Acknowledgments

All persons who have made substantial contributions to the work reported in the manuscript (e.g., technical help, writing and editing assistance, general support), but who do not meet the criteria for authorship, are named in the Acknowledgments and have given us their written permission to be named. If we have not included an Acknowledgments, then that indicates that we have not received substantial contributions from non-authors. All authors approved the final version of the manuscript.

References

[1] L.E. Miller, Why uwb? A review of ultrawideband technology, Natl. Inst. Stand. Technol. DARPA 78 (2003) URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.76.7051&rep=rep1&type=pdf.

[2] B.T. Ahmed, J.L.M. Campos, J.A. Ruiz-Cruz, Impact of ultra wide band emission on WiMAX systems at 2.5 and 3.5 GHz, Comput. Netw. 54 (10) (2010) 1573–1583, http://dx.doi.org/10.1016/j.comnet.2009.12.010.

[3] K. Ohno, T. Ikegami, Interference mitigation technique for coexistence of pulse-based uwb and OFDM, EURASIP J. Wireless Commun. Networking 2008 (2008) 1–11, http://dx.doi.org/10.1155/2008/285683.

[4] S.S. Mo, A.D. Gelman, On the power spectral density of UWB signals in IEEE 802.15. 3a, in: 2004 IEEE Wireless Communications and Networking Conference (IEEE Cat. No. 04TH8733), Vol. 2, IEEE, 2004, pp. 999–1002, http://dx.doi.org/10.1109/WCNC.2004.1311323.

[5] J. Romme, L. Piazzo, On the power spectral density of time-hopping impulse radio, in: 2002 IEEE Conference on Ultra Wideband Systems and Technologies (IEEE Cat. No. 02EX580), IEEE, 2002, pp. 241–244, http://dx.doi.org/10.1109/UWBST.2002.1006357.

[6] T. Xia, A.S. Venkatachalam, D. Huston, A high-performance low-ringing ultrawideband monocycle pulse generator, IEEE Trans. Instrum. Meas. 61 (1) (2011) 261–266, http://dx.doi.org/10.1002/9780470060490.

[7] A.-H. Mohsenian-Rad, J. Mietzner, R. Schober, V.W. Wong, Pre-equalization for pre-rake DS-UWB systems with spectral mask constraints, IEEE Trans. Commun. 59 (3) (2010) 780–791, http://dx.doi.org/10.1109/TCOMM.2011.122110.090718.

[8] H. Sheng, P. Orlik, A.M. Haimovich, L.J. Cimini, J. Zhang, On the spectral and power requirements for ultra-wideband transmission, in: IEEE International Conference on Communications, 2003. ICC'03., Vol. 1, IEEE, 2003, pp. 738–742, http://dx.doi.org/10.1109/ICC.2003.1204271.

[9] Y.-P. Nakache, A.F. Molisch, Spectral shape of UWB signals-influence of modulation format, multiple access scheme and pulse shape, in: The 57th IEEE Semiannual Vehicular Technology Conference, 2003. VTC 2003-Spring., Vol. 4, IEEE, 2003, pp. 2510–2514, http://dx.doi.org/10.1109/VETECS.2003.1208843.

[10] H. Kim, Y. Joo, Fifth-derivative Gaussian pulse generator for UWB system, in: 2005 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium-Digest of Papers, IEEE, 2005, pp. 671–674, http://dx.doi.org/10.1109/RFIC.2005.1489903.

[11] R. Merz, J. Widmer, J.-Y. Le Boudec, B. Radunović, A joint PHY/MAC architecture for low-radiated power TH-UWB wireless ad hoc networks, Wirel. Commun. Mob. Comput. 5 (5) (2005) 567–580, http://dx.doi.org/10.1002/wcm.313.

[12] M. Ghavami, L. Michael, R. Kohno, Ultra Wideband Signals and Systems in Communication Engineering, John Wiley & Sons, 2007, http://dx.doi.org/10.1002/9780470060490.

[13] M.Z. Win, Spectral density of random UWB signals, IEEE Commun. Lett. 6 (12) (2002) 526–528, http://dx.doi.org/10.1109/LCOMM.2002.806458.

[14] P. Adrian, An optimization of Gaussian UWB pulses, Dev. Appl. Syst. 50 (2010) URL: http://www.dasconference.ro/cd2010/data/papers/B66.pdf.

[15] M. Mirshafiei, M. Abtahi, S. LaRochelle, L.A. Rusch, Wideband antenna EIRP measurements for various UWB waveforms, in: 2008 IEEE International Conference on Ultra-Wideband, Vol. 1, IEEE, 2008, pp. 125–128.

[16] P. Protiva, J. Mrkvica, J. Machac, Universal generator of ultra-wideband pulses, Radioengineering 17 (4) (2008) 74–78.

[17] .

[18] N.C. Beaulieu, B. Hu, On determining a best pulse shape for multiple access ultra-wideband communication systems, IEEE Trans. Wireless Commun. 7 (9) (2008) 3589–3596.

[19] G. Zhang, Y. Dai, X. Zhang, Y. Lv, L. Chen, Design and implementation of UWB pulse with multiple narrow-band interferences mitigation, in: 2011 International Conference on Consumer Electronics, Communications and Networks (CECNet), IEEE, 2011, pp. 1154–1157.

[20] J. Han, C. Nguyen, A new ultra-wideband, ultra-short monocycle pulse generator with reduced ringing, IEEE Microw. Wirel. Compon. Lett. 12 (6) (2002) 206–208, http://dx.doi.org/10.1109/LMWC.2002.1009996.

[21] A.F. Molisch, K. Balakrishnan, C.-C. Chong, S. Emami, A. Fort, J. Karedal, J. Kunisch, H. Schantz, U. Schuster, K. Siwiak, IEEE 802.15. 4a channel model-final report, IEEE P802 15 (04) (2004) 0662, URL: https://www.ieee802.org/15/pub/04/15-04-0662-02-004a-channel-model-final-report-r1.pdf.

[22] Z. Hui, Z. Guiliang, D. Wang, Super-Gaussian window function and its applications, in: China., 1991 International Conference on Circuits and Systems, IEEE, 1991, pp. 595–598.

[23] A. Sharma, S.K. Sharma, Pulse shape optimization for physical layer UWB communication and performance evaluation in saleh–Valenzuela channel IEEE 802.15.3a, AEU - Int. J. Electron. Commun. 95 (2018) 5–15, http://dx.doi.org/10.1016/j.aeue.2018.07.028, URL: https://www.sciencedirect.com/science/article/pii/S1434841117322811.

[24] A.F. Molisch, J.R. Foerster, M. Pendergrass, Channel models for ultrawideband personal area networks, IEEE Wirel. Commun. 10 (6) (2003) 14–21, http://dx.doi.org/10.1109/MWC.2003.1265848.

[25] C. Findikli, S. Erkucuk, M.E. Celebi, Implementation of IEEE 802.15. 4a based UWB systems for coexistence with primary users, Int. J. Adv. Telecommun. Electrotech. Signals Syst. 5 (2) (2016) 86–94, http://dx.doi.org/10.11601/ijates.v5i2.134.

[26] L. Liu, S. Cheung, T. Yuk, Compact MIMO antenna for portable devices in UWB applications, IEEE Trans. Antennas and Propagation 61 (8) (2013) 4257–4264.

[27] R.S. El-Khamy, I. Ghaleb, H.N. Kheirallah, et al., Performance of the IEEE 802.15. 4a UWB system using two pulse shaping techniques in presence of single and double narrowband interferences, Int. J. Wirel. Mob. Netw. 5 (3) (2013) 35, URL: https://www.semanticscholar.org/paper/Performance-of-the-IEEE-802.15.4a-UWB-System-using-El-Khamy-Shaaban/1798d3ae3bcc60d22f0443e53bfd4a5dd545fb88.

[28] A. Ometov, V. Shubina, L. Klus, J. Skibińska, S. Saafi, P. Pascacio, L. Flueratoru, D.Q. Gaibor, N. Chukhno, O. Chukhno, et al., A survey on wearable technology: History, state-of-the-art and current challenges, Comput. Netw. 193 (2021) 108074, http://dx.doi.org/10.1016/j.comnet.2021.108074.

[29] M. Rahman, M. NagshvarianJahromi, S.S. Mirjavadi, A.M. Hamouda, Compact UWB band-notched antenna with integrated bluetooth for personal wireless communication and UWB applications, Electronics 8 (2) (2019) 158.
[30] K. Na, H. Jang, H. Ma, Y. Choi, F. Bien, A 200-mb/s data rate 3.1–4.8-GHz ir-UWB all-digital pulse generator with DB-BPSK modulation, IEEE Trans. Circuits Syst. II 62 (12) (2015) 1184–1188, http://dx.doi.org/10.1109/TCSII.2015.2468951.
[31] S. Franz, U. Mitra, Quantized UWB transmitted reference systems, IEEE Trans. Wireless Commun. 6 (7) (2007) 2540–2550, http://dx.doi.org/10.1109/ACSSC.2005.1599725.
[32] H. Gong, H. Nie, Z. Chen, Performance comparisons of UWB selective rake and transmitted reference receivers under IEEE802. 15.4 a industrial environments, in: 2006 IEEE Annual Wireless and Microwave Technology Conference, IEEE, 2006, pp. 1–5, http://dx.doi.org/10.1109/WAMICON.2006.351956.
[33] P. Medina, J.R. Gallardo, J. Sánchez, F. Ramírez-Mireles, Impact of delay spread on IEEE 802.15. 4a networks with energy detection receivers, J. Appl. Res. Technol. 8 (3) (2010) 352–362, URL: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-64232010000300005.
[34] T.-A. Phan, V. Krizhanovskii, S.-K. Han, S.-G. Lee, H.-s. Oh, N.-S. Kim, 4.7 Pj/pulse 7 th derivative Gaussian pulse generator for impulse radio UWB, in: 2007 IEEE International Symposium on Circuits and Systems, IEEE, 2007, pp. 3043–3046, http://dx.doi.org/10.1109/ISCAS.2007.377988.
[35] T.I. Krébesz, G. Kolumbán, K.T. Chi, F.C. Lau, H. Dong, Use of UWB impulse radio technology in in-car communications: Power limits and optimization, IEEE Trans. Veh. Technol. 66 (7) (2017) 6037–6049, http://dx.doi.org/10.1109/TVT.2017.2647849.
[36] A. Kamal, A. Bhattacharya, M. Tamrakar, C. Roy, Low-ringing and reduced-cost step recovery diode based UWB pulse generators for GPR applications, Microw. Opt. Technol. Lett. 56 (10) (2014) 2289–2294, http://dx.doi.org/10.1002/mop.28581.
[37] M. Rahman, A. Haider, M. Naghshvarianjahromi, A systematic methodology for the time-domain ringing reduction in uwb band-notched antennas, IEEE Antennas Wirel. Propag. Lett. 19 (3) (2020) 482–486, http://dx.doi.org/10.1109/LAWP.2020.2972025.
[38] H. Zhang, X. Liu, N. Li, T.A. Gulliver, A novel UWB pulse generator, in: Proceedings of 2011 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, IEEE, 2011, pp. 691–695, http://dx.doi.org/10.1109/PACRIM.2011.6032977.
[39] Z. Shahid, S. Khan, A.Z. Alam, Sub-nano seconds UWB pulse shaping with reduced ringing phenomenon, in: 2012 International Conference on Computer and Communication Engineering (ICCCE), IEEE, 2012, pp. 617–620, http://dx.doi.org/10.1109/ICCCE.2012.6271261.
[40] J.-Y. Lee, R.A. Scholtz, Ranging in a dense multipath environment using an UWB radio link, IEEE J. Sel. Areas Commun. 20 (9) (2002) 1677–1683, http://dx.doi.org/10.1109/JSAC.2002.805060.
[41] J.C. Adams, W. Gregorwich, L. Capots, D. Liccardo, Ultra-wideband for navigation and communications, in: 2001 IEEE Aerospace Conference Proceedings (Cat. No. 01TH8542), Vol. 2, IEEE, 2001, pp. 2–785, http://dx.doi.org/10.1109/AERO.2001.931259.
[42] K.M. Cohen, C. Attias, B. Farbman, I. Tselniker, Y.C. Eldar, Channel estimation in UWB channels using compressed sensing, in: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2014, pp. 1966–1970, http://dx.doi.org/10.1109/ICASSP.2014.6853942.
[43] A. Sharma, S.K. Sharma, Spectral efficient pulse shape design for UWB communication with reduced ringing effect and performance evaluation for IEEE 802.15. 4a channel, Wirel. Netw. 25 (5) (2019) 2723–2740, http://dx.doi.org/10.1007/s11276-019-01989-6.
[44] Z. Tian, V. Lottici, Low-complexity ML timing acquisition for UWB communications in dense multipath channels, IEEE Trans. Wireless Commun. 4 (6) (2005) 3031–3038, http://dx.doi.org/10.1109/TWC.2005.857996.
[45] L. Yang, G.B. Giannakis, Blind UWB timing with a dirty template, in: 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 4, IEEE, 2004, p. iv, http://dx.doi.org/10.1109/ICASSP.2004.1326875.
[46] J.-Y. Lee, UWB channel modeling in roadway and indoor parking environments, IEEE Trans. Veh. Technol. 59 (7) (2010) 3171–3180, http://dx.doi.org/10.1109/TVT.2010.2044821.
[47] W.-C. Wu, Multiuser detection of time-hopping PPM uwb system in the presence of multipath fading, AEU-Int. J. Electron. Commun. 61 (1) (2007) 22–29, http://dx.doi.org/10.1016/j.aeue.2006.02.001.
[48] S. Franz, U. Mitra, Generalized UWB transmitted reference systems, IEEE J. Sel. Areas Commun. 24 (4) (2006) 780–786, http://dx.doi.org/10.1109/JSAC.2005.863829.
[49] S. Kordnoori, H. Mostafaei, M.H. Behzadi, Evaluation of the error performance of the IEEE802. 16 standard based on a hidden Markov model, Sigma 37 (4) (2019) URL: https://eds.yildiz.edu.tr/AjaxTool/GetArticleByPublishedArticleId?PublishedArticleId=3971.
[50] A. Fernekess, A. Klein, B. Wegmann, K. Dietrich, M. Litzka, Load dependent interference margin for link budget calculations of OFDMA networks, IEEE Commun. Lett. 12 (5) (2008) 398–400, http://dx.doi.org/10.1109/LCOMM.2008.080032.
[51] J. Foerster, Fec performance of concatenated reed-solomon and convolutional coding with interleaving, in: IEEE 802.16. 1pc-00/33, 2000, URL: https://ieee802.org/16/tg1/phy/contrib/802161pc-00_33.pdf.



Dr. Abhishek Sharma received the Master of Science (with honors) degree in Electronics and Computational Physics from Dr. B.R.A. University, Agra in 2005 and Master of Technology (with honors) degree in Telecommunications Technology from R.G.P.V, Bhopal in 2008. He has completed his Ph.D. in Electronics and Wireless Transmission from State Government University, Dr. A.P.J. Abdul Kalam Technical University, Lucknow in 2019. He is currently working as Research Associate Professor in the department of Electronics and Communication Engineering at KIET Group of Institutions, Delhi-NCR, Ghaziabad. He is also in charge of the Centre of Excellence for Space Technologies. He is currently doing his research in Ultra-wideband communication for space technologies. His research interest includes Wireless Communication, Telecommunication Engineering and Deep Learning.



Dr. Amik Garg He is an acting Director of KIET Group of Institutions, Delhi-NCR, Ghaziabad. He has completed engineering professional graduation from Delhi College of Engineering, New Delhi in 1986 and Master of Technology from the Indian Institute of Technology, New Delhi in 1996. He has also completed his Ph.D. from the Indian Institute of Technology, New Delhi in 2010. He is currently leading technical educational organizations to attain greater heights in alignment with their vision. His current research interests include Wireless Communication, Space Sciences, Optical Communication, Computer Networks, and Signal Processing.



Dr. Sanjay Kumar Sharma has completed his Ph.D. (Wireless Communication) from Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia, Delhi, in 2010. He has done M.Tech. in Digital Communication from UPTU, Lucknow (currently — Dr. A.P.J. Abdul Kalam Technical University, Lucknow). He has completed his B.E. in Electronics Engineering from M.M.M. Engineering College, Gorakhpur in 1997. His current research interests include Wireless Communication, Space Technologies, Digital Communication, IOT, Speech and Signal Processing. He is currently Professor and Head in the Department of Electronics and Communication Engineering at KIET Group of Institutions, Delhi-NCR, Ghaziabad.



Dr. Vibhav Kumar Sachan has completed his Ph.D. (Wireless Communication) from Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia, Delhi, in 2012. He has done M.Tech. in Digital Communication from UPTU, Lucknow (currently — Dr. A.P.J. Abdul Kalam Technical University, Lucknow) in 1997. He has completed his B.E. in Electronics and Instrumentation Engineering from Bundelkhand University (B.I.E.T., Jhansi), UP in 1997. Dr. Vibhav Kumar Sachan is Professor in the Department of Electronics and Communication Engineering at Krishna Institute of Engineering and Technology, Ghaziabad. His current research interests include Wireless Communication, Digital Communication, Machine Learning, Computer Networks and Signal Processing.



Dr. Parvin Kaushik is a Research Associate Professor in the Department of Electronics and Communication Engineering at KIET Group of Institutions, Delhi-NCR, Ghaziabad. He is also in In-charge of Space Technologies, Centre of Excellence. He has received his Ph.D. (Electronics Engineering) from AKTU, Lucknow in 2018. He has done his B.Tech. and M.Tech. from Kurukshetra University Kurukshetra, Kurukshetra in 2007 & 2010, respectively. His research interest includes radio over fiber systems, wireless communication systems & Space Communication.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:10/October-2021 Impact Factor- 6.752 www.irjmets.com

REVERSE ENGINEERING

Omkaar Mazumdar*1, Sumeet Padhi*2, Prof. A.K. Madan*3

*1,2,3Department Of Mechanical Engineering, Delhi Technological University (DTU) Rohini, New Delhi, India.

ABSTRACT

Reverse Engineering (RE), or Back Engineering is a course of interpreting design data and information about the construction of an already existing item. This enables you to figure out how a section was planned so you can reproduce it. It is also used to deduce by analysis how a device works, or what it does in a system. The process can be done in two ways: manually, or using computerized methods (CAD/CAM), depending on the usage context.

Keywords: Reverse Engineering, CAD/CAM, 3D Scanning, 3D Model.

I. INTRODUCTION

The purpose of reverse-engineering is to determine how an object or system works. There are various reasons for this. Reverse-engineering can be used to learn how something works and to re-create an object or to build something similar with additional enhancements. Often the goal of reverse-engineering software or hardware is to find a way to create the same product less expensively or because the actual product is no longer available. It may also be used to find the vulnerabilities in a product and to fix them. Obsolete parts can be converted to a 3D CAD model for future references. The only challenge one might face in reverse engineering is about its legality. In the U.S., reverse-engineering is generally considered a legal way to learn about a product as long as the original version is legally available and no other contract agreements are violated. U.S. trade laws aim to allow reverse-engineering if it is in the interest of improving the product or creating compatibility with other products that were previously incompatible. Reverse-engineering a patented product is generally legal under the Defend Trade Secrets Act, but there are cases where its validity can be questioned. Copyright holders have a legal way of dealing with anyone copying their inventions. The various laws relevant to reverse-engineering include the following:

- patent law;
- copyright and fair use law;
- trade secret law;
- any contract law specific to the commodity in question.

The object on which the process of reverse-engineering is being performed governs the steps taken to do so. However, there are three steps common to all reverse-engineering applications. They include:

- **Information extraction-** The object being reverse-engineered is studied, its design is obtained and that information is examined to determine how the pieces fit together.
- **Modeling-** The information collected is summarized into a conceptual model, with each piece of the model defining its function in the whole structure. The purpose of this step is to take specific and factual information and summarize it into a standard model that can be used to guide the design of new objects or systems.
- **Review-** This involves revising the model and testing it in a variety of contexts to ensure that it is a true summary of the original object or system.

Computer-aided design (CAD) is a reverse-engineering technique widely employed to rebuild a manufactured part when the original blueprint is no longer available. It involves producing a 3D model of the part so it can be remanufactured. A coordinate measuring machine measures the part, and as it is measured, a 3D wire frame image is generated using CAD software and displayed on a monitor. After the measuring is complete, the wire frame image is realized. Any part can be reverse-engineered using these methods.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:10/October-2021 Impact Factor- 6.752 www.irjmets.com

II. MOTIVATION

Reverse engineering is the act of dismantling an object to understand its working. It is primarily used to analyze and acquire information about how something works but is often used to replicate or improve an object. Depending on the technology, information acquired during back engineering can be used to remodel obsolete materials, analyze security, gain competitive advantage, or teach someone how something works. Regardless of how the information is used or related, reverse-engineering is the process of obtaining that information from a completed object. In this paper we only discuss about RE from a mechanical engineering standpoint, thus we only consider CAD/CAM applications of 3D manufactured products.

III. LITERATURE REVIEW

S. No.	Author, Year and Journal	Title	Findings
1	Saeid Motavalli 'Journal of Computers and Industrial Engineering', 1998	"Review of Reverse engineering approaches"	This paper reviews various techniques for 'Part digitization', 'segmentation' and 'surface modelling'. Earlier, data classification and component modeling work was done manually by operators which author considered as a disadvantage and discussed the need of research to automate this process.
2	Kevin N. Otto, Kristin L. Wood, 'Journal of Research in Engineering Design', 1998	"Product Evolution: A reverse engineering and redesign methodology"	The paper introduces a new reverse engineering procedure, which considers customer needs for the finished product. A functional model was created by cataloguing client requests, followed by reverse engineering. Next, the models are developed and enhanced using pre-existing designed experiments.
3	William B. Thompson et al, 'IEEE Transactions on Robotics and Automation', 1999	"Feature-Based Reverse Engineering of Mechanical Parts"	The paper describes a feature-based reverse engineering approach that uses manufacturing features as basic building blocks. This method has been found to have great benefits in usability and accuracy. It provides a high definition of the geometric component and allows easy integration into the CAD system.
4	Pal Benko, Ralph R Martin, Tomas Vardy 'Journal of Computer-aided Design',2001	"Algorithms for reverse engineering boundary representation models"	The paper explained the procedure for reconstructing solid models of conventional engineering objects in depth. Emphasis was put on producing accurate and topologically consistent boundary representation models, which can be directly used in CAD/CAM interface. Four key algorithmic components are presented in more detail and application of these is illustrated by means of various examples.
5	K. Schreve et al, 'Journal of Computing and Information Science in Engineering', 2006.	"Interactive Feature modelling for Reverse Engineering"	The paper presented feature based modelling tools for element rendering, edge detection, and adjusted angle measurements with appropriate examples. It was observed that 'edge detection' is a useful tool for reverse engineering lofted and swept features. Also 'draft angles' are critical components when making parts for molding applications.
6	A Durupt, S. Remy, G.Ducelleir 'Journal of Computing and Information	"Knowledge Based Reverse Engineering-An approach for reverse	The paper presented knowledge-based engineering (KBRE) approach as conventional RE approaches based on geometry alone provided frozen and non-reusable geometric models. A piston case study is also used to illustrate the approach. This work also aimed to develop a

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:10/October-2021 Impact Factor- 6.752 www.irjmets.com

	Science in Engineering', 2010	engineering of a mechanical part"	tool for reverse engineering that manages the multitier and multilevel aspects of RE.
7	Jun Wang et al 'Journal of Computer and Industrial Engineering', 2012	"A framework for 3D model reconstruction in reverse engineering"	The paper designed a framework to create 3D models from the boundary surface meshes of parts. The models developed from this framework are known to provide a higher level description of part geometry and allowing easy redesign and reuse for innovation.
8	Nabil Anwer, Luc Mathieu 'CIRP Annals Manufacturing Technology', 2016	"From reverse engineering to shape engineering in mechanical design"	The paper discussed the fundamentals of shape representation, shape processing and mining. It was found that many instincts of reverse engineering in product development may be gained from novel concepts, methods and tools of shape engineering.
9	Besndt Bellay, Harald Gall 'Imagix 4D release', 1997	"A comparison of four Reverse Engineering tools"	The paper investigated and compared the capabilities of four reverse engineering tools that analyze C source code: 'Refine/C', 'Imagix4D', 'Sniff+' and 'Rigi' by applying them to a commercial embedded software system as a case study. Based on evaluation, no single tool is the best. The tools evaluated are all different with varying strengths and weaknesses
10	Rudolf K. Keller et al. 'ICSE Proceedings', 1999	"Pattern-Based Reverse Engineering of Design Components"	The paper presents an environment for reverse engineering of design components based on the structural description of design parts. It also explained three case studies and discussed in-depth about pattern-based reverse engineering.
11	Josh Bongard, Hod Lipson 'PNAS', 2007	"Automated Reverse Engineering of non-linear dynamical systems"	The paper was the first to introduce a method that can automatically generate symbolic equations for a non-linear coupled dynamical system directly from time series data. 'Systems Biology' is one of its important application area for the proposed model.
12	Robin H. Helle, Hirpa G. Lemu. 'Materials Today: Proceedings', 2021	"A case study on use of 3D scanning for reverse engineering and quality control"	The findings of this paper suggest that the use of RE and PC 3D scanning technology is possible but significant challenges exist in the development of accurate plane reconstruction algorithms that address point cloud defects such as noise and holes. Using 3D scanning for RE purposes is found to be very suitable for parts with complex geometry with free areas that are difficult to measure by hand.
13	Zhaohui Geng & Bopaya Bidanda, Virtual and Physical Prototyping 2017	"Review of reverse engineering systems – current state of the art"	This paper discusses the components of RE systems and describes the needs for RE in manufacturing industries. Common contact and non-contact scanning techniques are described, along with possible interfaces with other manufacturing processes, especially 3D printing. Also described are new trends in the field that include more speed, accuracy, and automation in the future. A discussion on two major issues of tolerance regarding the integration of RE and 3DP.

14	Tamás Várady, Ralph R Martin, and Jordan Cox, Computer-Aided Design, 1997	"Reverse engineering of geometric models - an introduction"	Reviews the process of reverse engineering of shapes. The pros and cons of various data acquisition techniques are explained by the problems associated with border representation and model construction. 3D scanning, data acquisition, visualization of data obtained.
15	Francesco Buonamici et al. (2018), 'Journal of Computer-Aided Design and Applications	"Reverse engineering modeling methods and tools: a survey"	The purpose of this paper is to fill in the blanks; from the general definition of the overall RE framework (detection, fragmentation, fragmentation, insertion), both the latest survey and layout of available modeling methods and 3D data-driven tools are provided.
16	Francesco Buonamici et al. 'Journal of Computational Design and Engineering', 2018	"Reverse engineering of mechanical parts: A template-based approach"	Template-based deviant methods represent a lesser tested strategy compared to the field of CAD reconstruction from polygonal models. The reconstruction process is carried out based on the CAD template, its feature tree and the geometric parameters that are defined in terms of advanced material knowledge. The CAD template is embedded in mesh data, improving its size and shape parameters using the particle swarm optimization algorithm. As a result, a CAD parameter model that fully complements the set geometric relationships is produced and a feature tree, which explains the history of the combined model, is available to the reverse engineer.
17	Katsuhiro Saiga, AMM Sharif Ullah, Akihiko Kubo, Tashi, Procedia CIRP, Volume 98, 2021	"A Sustainable Reverse Engineering Process"	This paper concludes that the declining engineering process becomes more sustainable if it does not rely too much on sophisticated equipment, calculations, and human skills. This study presents a five-step sustainable development process. One, describe the given object using some basic geometric shapes. Two, represent the basic geometric shapes using cloud-based point clouds. Three, create cloud points combined by blending cloud points of the second step. Next, create a robust model using an off-the-shelf CAD package. Finally, create a 3D printed template using a robust CAD model.
18	Jin Liu (2020), International Journal of Computer Integrated Manufacturing	"An adaptive process of reverse engineering from point clouds to CAD models"	In this paper, a flexible process for the automatic reconstruction of the CAD model from point to point is proposed. First, this method removes the basic shapes from the point clouds with the RANSAC algorithm, and then analyzes the deviations of points from the fitted primitive shapes by histograms To find the points of the clouds segmented incorrectly, the method updates the separation parameters according to the Gaussian sound and repeats the process of detecting the original shape. After some cycles of repetition, the approach can detect reasonable primitive shapes from the clouds. Then a rule was introduced to acquire CAD models with the guidance of the old form.
19	N. Werghi, R. Fisher, C. Robertson, A.	"Object reconstruction by incorporating	This paper deals with the reconstruction of moderate 3D geometric models of objects from a broad set of data. It describes a new technique of global shape improvement

	Ashbrook, Computer-Aided Design, Volume 31, Issue 6, 1999	geometric constraints in reverse engineering"	based upon feature positions and geometric constraints. It proposes a general framework for growth where constraints can be added and integrated into the model reconstruction process, leading to better trade-offs between reducing the error of balance and tolerance of boundaries.
20	Sebastián Mansilla Mora, Juan Claver Gil, Ana María Camacho López, Procedia Manufacturing, Volume 41, 2019	"Influence of manufacturing parameters in the dimensional characteristics of ABS parts obtained by FDM using reverse engineering techniques"	Different parts were first designed on a CAD software, then realized by an FDM 3D printer. Using a 3D scanner and corresponding data acquisition methods, they were developed into another model- the RE model. This model was compared to the original model by means of a measurement, inspection and quality control software, to calculate the accuracy of an RE model.
21	Research into the engineering application of reverse engineering technology	"Yu Zhang, Journal of Materials Processing Technology, Volume 139, Issues 1–3, 2003"	In this paper, the topic discussed is the engineering application of RE technology in the design and manufacturing of a complex product. Taking a die of the inlet of a diesel engine as an example, they have described the whole processes, from object digitization, CAD model reconstruction to NC machining, and have gained a good result. RE is an effective way to significantly reduce the cycle of product development and ensure product quality through the complex design and production of the product.
22	A comprehensive process of reverse engineering from 3D meshes to CAD models	"Roseline Bénière, Gérard Subsol, Gilles Gesquière, François Le Breton, William Puech, Computer-Aided Design, Volume 45, Issue 11, 2013"	Discussed in this paper is a comprehensive automation process for the continuous B-Rep model from the selected 3D CAD mesh. This method involves three steps: the extraction of geometric primitives, which works successfully, unlike many existing methods, in the case of sparse meshes, by using an adapted pre-segmentation step; wire construction, which incorporates intersecting curves in a consistent manner using a new layout based on a graph that is close to the creation of an all-encompassing model to form B Representation. This method gives good results with CAD meshes, independently of the mesh structure, which can be dense or sparse, regular or not and even if there are many geometric primitives.
23	Jorg Seewig, Matthias Eifler, Frank Schneider, Jian C. Aurich, 'Procedia CIRP', 2016	"Design and verification of geometric roughness standards by reverse engineering"	The paper investigated examples of different design and manufacturing of accordant measurement standards. Based on the research, the author proposed a standard engineering method for the design of measurement standards. The proposed method was confirmed by the production of calculated data from the experimental setup. Only a slight deviation was observed.
24	Seokbae Son, Hyunpung Park, Kwan H. Lee, 'International Journal of Machine	"Automated laser scanning system for reverse engineering and inspection"	The paper proposed an automatic laser scanning system in response to already existing laser scanning system operated manually by skilled operators. The manual approach is expensive and prone to errors. To automate this process, proper hardware and software modules are required.

	tools and Manufacture', 2002		Hardware system consists of laser scanning device and fixture setups whereas software modules generate optimal scan plan. The limitation of proposed method is that it is better suited for inspection than for reverse engineering as CAD model is assumed to be given for comparison purposes.
25	Mudit Bansal, Sukhdeep S. Dhami, 'International Journal of Emerging Technologies in Engineering Research (IJETER)', 2017	"3D Scanning for Reverse Engineering-Technological Advancements, Process Overview, Accuracy Inspection, Challenges and Remedies"	The author performed reverse engineering of an industrial component using 3d scanners to get data points and then use it to generate CAD model and prototype it using 3D printing. The paper's objective was to show how to carry out the whole process, how data processing is the major challenge and how with the help of quality check, inspection and taking appropriate precautions while scanning, can reduce the standard deviation and improve accuracy.
26	V.D. Yadav, V.M. Jamdar, G.S. Jadhav, P.S. Gunavant, P.S. Mohite, 'International Advanced Research Journal in Science, Engineering and Technology', 2017	"Reverse Engineering of Crank Shaft"	The author created 3D model of crankshaft by using old unused crankshaft and applying reverse engineering techniques. Vernier caliper was mainly used for determining the dimensions whereas Coordinate measuring machine (CMM) along with ProE software was used to trace out critical and complicated profiles.
27	Tiziana Segreto et al. 'Procedia CIRP', 2019	"Integration of reverse engineering and ultrasonic non-contact testing procedure for quality assessment of CFRP aeronautical components"	The paper integrated the use of laser-based reverse engineering method and ultrasonic testing technique to inspect carbon fiber reinforced polymer (CFRP) components for any external and internal defects respectively. The author also developed a custom-made software to analyze and visualize the information provided by laser scan and ultrasonic testing simultaneously
28	Bernd Engel, Sara Salman Hassan Al-Maeeni, 'Procedia CIRP', 2019	"An integrated reverse engineering and failure analysis approach for recovery of mechanical shafts"	The paper proposed an integral approach combining reverse engineering and failure inspection approach using analytical, FEA and experimental methods for recovering a shaft of a rotary draw bending machine. Dimensions are retrieved using non-contact laser scanning and manual measurement methods. It was concluded that the integral approach proposed can be used for any type of machine tools/parts.
29	Lingling Li, Congbo Li, Ying Tang, Yanbin Du, 'Robotics and Computer	"An integrated approach of reverse engineering aided remanufacturing	The author applied reverse engineering techniques to remanufacture a used forging die and a worn gear bracket. A prominent cross-section algorithm embedded with curvature constraint was proposed in order to automatically eliminate the defective point clouds from reconstruction

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:10/October-2021 Impact Factor- 6.752 www.irjmets.com

	Integrated Manufacturing', 2017	process for worn components"	process. Also, a modified ICP algorithm was used to integrate curvature and distance constraints. Overall, it was shown that the method used is effective and reliable and has the potential to be improved further.
30	Robin H Helle, Hirpa G. Lemu, 'Materials Today: Proceedings',	"A case study on use of 3D scanning for reverse engineering and quality control"	The paper showed a case-study of reverse engineering a 3D printed components. Complex internal feature were 3D scanned and point clouds were optimized to generate error-free CAD design. The paper presented the challenges in generating accurate surface reconstruction algorithms and discussed about point cloud imperfections.

IV. RESEARCH GAP

Key research areas which still need further work before general-purpose reverse engineering becomes widely available include: improving data capture and measurement, coping with noise, visual integration, coping with missing data, reliable classification, fair surface fitting, recognizing natural or human-intended structure of the geometry of the part, and eventually the development of consistent models.

- The specific application of reverse engineering 3D CAD meshes has been studied very little.
- The differences in measurements and structure of a reverse engineered 3D CAD model compared to its original model can vary depending on multiple factors not researched yet.
- A standard operating procedure for reverse engineering complex 3D structures can be developed.
- A systematic dissertation of model long methods for a mechanical engineering point of view is still missing.

V. CONCLUSION

RE technology is not about copying an existing product, but getting a design idea from an existing visual model and creating a complete digital product model and continuing to improve product design. This paper reviews and summarizes a bunch of other research papers and compiles them in a legible format for reference purposes. All discussion done in this paper is on reverse engineering from a mechanical engineer standpoint, with CAD, 3D models and 3D printing being in context.

VI. REFERENCES

[1] Geng, Z., & Bidanda, B. (2017). Review of reverse engineering systems – current state of the art. Virtual and Physical Prototyping, 12(2), 161–172

[2] Geng, Z., & Bidanda, B. (2017). Review of reverse engineering systems – current state of the art. Virtual and Physical Prototyping, 12(2), 161–172.

[3] Várady, T., Martin, R. R., & Cox, J. (1997). Reverse engineering of geometric models—an introduction. Computer-Aided Design, 29(4), 255–268.

[4] Buonamici, F., Carfagni, M., Furferi, R., Governi, L., Lapini, A., & Volpe, Y. (2017). Reverse engineering of mechanical parts: A template-based approach. Journal of Computational Design and Engineering, 5(2), 145–159.

[5] Saiga, K., Ullah, A. S., Kubo, A., & Tashi. (2021). A Sustainable Reverse Engineering Process. Procedia CIRP, 98, 517–522.

[6] Liu, J. (2020). An adaptive process of reverse engineering from point clouds to CAD models. International Journal of Computer Integrated Manufacturing, 33(9), 840–858.

[7] Buonamici, F., Carfagni, M., Furferi, R., Volpe, Y., & Governi, L. (2020). Reverse engineering by CAD template fitting: study of a fast and robust template-fitting strategy. Engineering with Computers, 37(4), 2803–2821.

[8] Werghi, N., Fisher, R., Robertson, C., & Ashbrook, A. (1999). Object reconstruction by incorporating geometric constraints in reverse engineering. Computer-Aided Design, 31(6), 363–399.

[9] Mora, S. M., Gil, J. C., & Camacho López, A. M. (2019). Influence of manufacturing parameters in the dimensional characteristics of ABS parts obtained by FDM using reverse engineering techniques. Procedia Manufacturing, 41, 968–975.

[10] Buonamici, F., Carfagni, M., Furferi, R., Governi, L., Lapini, A., & Volpe, Y. (2017). Reverse engineering modeling methods and tools: a survey. Computer-Aided Design and Applications, 15(3), 443–464.

[11] Keller, R. K., Schauer, R., Robitaille, S., & Pagé, P. (1999). Pattern-based reverse-engineering of design components. Proceedings of the 21st International Conference on Software Engineering - ICSE '99.

[12] Motavalli, S. (1998). Review of reverse engineering approaches. Computers & Industrial Engineering, 35(1-2), 25–28.

[13] Otto, K. N., & Wood, K. L. (1996). A Reverse Engineering and Redesign Methodology for Product Evolution. Volume 4: 8th International Conference on Design Theory and Methodology.

[14] Benkő, P., Martin, R. R., & Várady, T. (2001). Algorithms for reverse engineering boundary representation models. Computer-Aided Design, 33(11), 839–851.

[15] Bellay, B., & Gall, H. (n.d.). A comparison of four reverse engineering tools. Proceedings of the Fourth Working Conference on Reverse Engineering.

[16] Thompson, W. B., Owen, J. C., & De St. Germain, H. J. (1995). Feature-Based Reverse Engineering of Mechanical Parts. Revision.

[17] Wang, J., Gu, D., Yu, Z., Tan, C., & Zhou, L. (2012). A framework for 3D model reconstruction in reverse engineering. Computers & Industrial Engineering, 63(4), 1189–1200.

[18] Durupt, A., Remy, S., & Ducellier, G. (2010). Knowledge Based Reverse Engineering—An Approach for Reverse Engineering of a Mechanical Part. Journal of Computing and Information Science in Engineering, 10(4).

[19] Anwer, N., & Mathieu, L. (2016). From reverse engineering to shape engineering in mechanical design. CIRP Annals, 65(1), 165–168.

[20] Schreve, K., Goussard, C. L., Basson, A. H., & Dimitrov, D. (2006). Interactive Feature Modeling for Reverse Engineering. Journal of Computing and Information Science in Engineering, 6(4), 422–424.

[21] Bénière, R., Subsol, G., Gesquière, G., Le Breton, F., & Puech, W. (2013). A comprehensive process of reverse engineering from 3D meshes to CAD models. Computer-Aided Design, 45(11), 1382–1393.

[22] Zhang, Y. (2003). Research into the engineering application of reverse engineering technology. Journal of Materials Processing Technology, 139(1-3), 472–475.

[23] Seewig, J., Eifler, M., Schneider, F., & Aurich, J. C. (2016). Design and Verification of Geometric Roughness Standards by Reverse Engineering. Procedia CIRP, 45, 259–262.

[24] Son, S., Park, H., & Lee, K. H. (2002). Automated laser scanning system for reverse engineering and inspection. International Journal of Machine Tools and Manufacture, 42(8), 889–897.

[25] V D, Y., V M, J., G S, J., P S, G., & P S, M. (2017). Reverse Engineering of Crank Shaft. IARJSET, 4(1), 20–22.

[26] Segreto, T., Bottillo, A., Caggiano, A., & Martorelli, M. (2019). Integration of reverse engineering and ultrasonic non-contact testing procedures for quality assessment of CFRP aeronautical components. Procedia CIRP, 79, 343–348.

[27] Engel, B., & Al-Maeeni, S. S. H. (2019). An Integrated Reverse Engineering and Failure Analysis Approach for Recovery of Mechanical Shafts. Procedia CIRP, 81, 1083–1088.

[28] Li, L., Li, C., Tang, Y., & Du, Y. (2017). An integrated approach of reverse engineering aided remanufacturing process for worn components. Robotics and Computer-Integrated Manufacturing, 48, 39–50.

[29] Helle, R. H., & Lemu, H. G. (2021). A case study on use of 3D scanning for reverse engineering and quality control. Materials Today: Proceedings, 45, 5255–5262.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/355807477

Reviewing the Analeptic Activity of Calcium Citrate Malate

Article *in* International Journal of Pharmaceutical Sciences Review and Research · October 2021

DOI: 10.47583/ijpsrr.2021.v70i02.036

CITATIONS
0

READS
211

6 authors, including:



Bharat Kwatra
Sabarmati University
52 PUBLICATIONS **428** CITATIONS
SEE PROFILE



Hiya Abrol
Jain University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Ishika Gulati
Delhi Technological University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Pankaj Pradeep
VIT University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Antimicrobial Discovery (Project ADMIS) View project

All content following this page was uploaded by Hiya Abrol on 01 November 2021.

The user has requested enhancement of the downloaded file.

Review Article

Reviewing the Analeptic Activity of Calcium Citrate Malate

Bharat Kwatra*, Chelsea Rumao, Hiya Abrol, Ishika Gulati, Pankaj Pradeep, Srashti Bajpai
Sabarmati University, Sophia College - Mumbai, Jain University - Bangalore, Delhi Technological University- Delhi, Vellore Institute of Technology- Vellore, MIT University- Pune, India.
***Corresponding author's E-mail:** info.researchbulb@gmail.com

Received: 11-07-2021; **Revised:** 18-09-2021; **Accepted:** 26-09-2021; **Published on:** 15-10-2021.

ABSTRACT

Calcium Citrate Malate's (CCM) health benefits were patented over three decades ago. It is involved in calcium retention, in children and adolescents. In adulthood, it is seen to promote the maintenance of bone health. In conjunction with vitamin D, CCM also decreases the risk of bone fracture in aged people and ensures that postmenopausal women's health is made better. CCM, unlike other supplements, does not need to be taken along with a meal and is seen to deliver benefits to individuals of any age group. CCM is usually used when the person has achlorhydria where the gastric acid secretion is reduced and there is an absence of an acidic environment (required by calcium carbonate). Calcium citrate Malate possesses outstanding bioavailability properties and is a useful extension to the calcium available presently, either for direct supplementation or for food fortification.

Keywords: Calcium Citrate Malate, Dental care, Gynaecology, Calcium deficiency, Osteoporosis.

QUICK RESPONSE CODE →

DOI: 10.47583/ijpsrr.2021.v70i02.036

DOI link: http://dx.doi.org/10.47583/ijpsrr.2021.v70i02.036

INTRODUCTION

The most abundant mineral in the human body is calcium. Calcium deficiency can lead to osteoporosis. It can be prevented by various calcium preparations, organic salts like tricalcium citrate, calcium lactate, calcium lactate gluconate, calcium gluconate, and inorganic salts like calcium chloride, calcium carbonate, and calcium phosphate.[1]

Public awareness of the importance of adequate calcium in the diet has led to a greater demand for supplements and calcium supplementation. Calcium solubility is believed to be an important factor in its absorption.[2] The main function of consuming calcium is to strengthen the skeleton. Insufficient consumption of calcium during the early stages of life hinders the development of bone and further fractures in adulthood.[3] Calcium is present in blood, muscle, and cell fluids. It is present both in plant and animal foods. Amongst the animals' food, the major source is milk and its products and amongst the plants, it is green leafy vegetables. There are 6 calcium salts[4]-

1) lactate

2) carbonate

3) citrate

4) gluconate

5) phosphate

6) citrate malate.

Rapidly increasing awareness of bone health has increased the demand for different types of Ca supplements and Ca fortified food products.[5] Calcium carbonate and calcium citrate are the most common calcium supplements that are taken by a large variety of people. This is because calcium carbonate is affordable.[5] Calcium supplements are consumed by people more than 50 years of age. Studies suggest that high calcium intake may prevent vascular diseases. However, calcium supplements increase vascular calcification and increase the mortality rate in patients with renal failure.[3] CCM is a metastable compound consisting of calcium salts of citric acid and malic acid. The structure of CCM is unknown. It is present as a mixture of various crystalline and non-crystalline forms.[5,8] It exists in several stages of hydration. It has a molecular ratio of 6:2:3 of calcium: citrate: malate.[2,5] The structure of CCM makes it possible and easily soluble in the stomach than calcium citrate. Calcium citrate malate is well-absorbed taken with or without food. CCM is a composite salt with 10 times more solubility than Calcium citrate or calcium malate independently. Ca absorption from CCM fortified orange/ apple juice is more than from milk and Calcium carbonate.[7] The chemistry of CCM makes it a particularly beneficial calcium source for individuals with hypochlorhydria or achlorhydria, which generally includes the elderly and those on medications that decrease gastric acid secretion.[1] Calcium Supplementation is available commercially in different forms. These forms are said to have different levels of efficacy amongst different people. This efficacy can also differ through the method that it is administered and based on the individual. Calcium is essential for the metabolic functioning of the body. There is a certain level

©Copyright protected. Unauthorised republication, reproduction, distribution, dissemination and copying of this document in whole or in part is strictly prohibited.

of Calcium required by the body to function properly. When that level drops below the threshold, there are bone defects and can lead to metabolic disorders in the long run.

The calcium balance in the body is maintained by vitamin D and PTH. The required calcium intake for individuals of different ages and gender should be between 400 and 1200 mg.[8] But most diet routines of people don't satisfy this. The most common defect due to low intake of calcium is osteoporosis which compromises bone health, and leads to reduced bone mass and fragility, making the bones weaker.[9] Calcium being a mineral contributes to bone mineral density (BMD), and when calcium is lost from bone, the BMD reduces. The key focus is to identify osteopenia early so that calcium supplementation can be given to solve osteopenia before it leads to osteoporosis.[9] Calcium levels are monitored by calciotropic hormones like vitamin D and parathyroid, which control the transport of calcium in the intestine, kidney and bone. Any imbalance in the serum calcium concentrations are detected by the Calcium Sensing Receptor (CaSR) and the parathyroid, therefore inactivating the PTH receptor, and this leads to jumping in the PTH secretion.[10] CCM is one of the most important minerals that a pregnant woman needs as it gives the baby the aid to develop its vital structures like the skeleton. Since calcium is so important to both the mother and growing baby, pregnant women should be given supplements in case their body is not able to cope with the calcium levels and requirements in that vulnerable state.[11] Therefore, monitoring patients after surgery is one of the key steps towards postoperative care, which mainly involves measuring serum calcium levels and taking necessary action in terms of giving vitamin D and/or calcium supplements.[12]

METHODS

The study was conducted using four databases Google Scholars SAGE, DOAJ and PubMed. The selection of papers was done based on keywords and themes relevant to this review. Further the published papers from these databases were arranged in systematic order with respect to year of publication.

RESULTS

Role of CCM in ENT

Many Controlled intervention trials with CCM were performed. A double-blind, placebo-controlled calcium supplementation study has been carried out in children and adolescents.[3] The primary endpoints were time to first myocardial infarction and time to first stroke and the secondary endpoint was time to death.[13] Eleven adults were taken as subjects and upper removable intra- oral slabs were made for each subject. The erosion was measured by the loss of enamel of the casts of the enamel surface. The casts were obtained and compared before and after the exposure to oral fluids. Oral isotope was used to measure Ca absorption in healthy women.[7] The procedure was approved by the Human Subjects Committee at Indiana University. A blind and non-randomized clinical study was taken into consideration on a group of 100 healthy Indian children.[14] A questionnaire was given including the medical history and juice consumption of the person.

The results obtained were similar in both the cases i:e rats and humans (women). In the first experiment, CCM was compared against CaCO3, and in the second, it was compared against milk.[1] It was observed that CCM had higher 6-day retention values. Calcium absorption was significantly better than milk in the case of orange juice beverages.[1] There is a significant effect of pH treatment on the amount of soluble calcium present in ionic form and the total soluble calcium.[4] When the pH is 2.0, all the soluble calcium is present in the ionic state. When the pH increases to 7.0, there is a different solubility pattern for each calcium salt.[4] Both calcium citrate and CCM exhibited higher amounts of soluble calcium complex. At pH 2.0, there was a difference in the soluble calcium levels of all the calcium salt sources. When the pH was increased to 7.0, all the calcium phosphate that was present, precipitated.[4] This unique change in the solubility patterns at different pH describes the importance of pH adjustments during in vitro testing of bioavailability.

Results from double-blind, placebo-controlled calcium studies showed that the absorption rate of milk or calcium carbonate is much lesser than the calcium supplementation or CCM fortified juices.[3] This result also suggested that CCM has also proven to be a bigger help than calcium carbonate to prevent bone loss in older postmenopausal women.[3] It could be observed that total calcium intake in the prepubertal CCM supplemented group is greater than that of the placebo group. On the other hand, increased calcium intake was not associated with the CCM supplemented group, rather it was greater in the placebo group in pubertal subjects.

A calcium citrate malate addictive is formulated to curb erosion of soft drinks.[15] Calcium and phosphate supplementation techniques have been a source for a new method to curb erosion. The effectiveness of this additive is measured gravimetrically as the percentage of weight loss. Experiments show that the percentage has been reduced from 8.7% to figures in the range of 0.05% to 0.75%, in the presence of varying concentrations of CCM.[15] Experiments done on a group of rats show that erosive property has been reduced from 4.29±0.85 to 0.95±0.38 and 0.71 ±0.55 by the CCM supplement (recorded on a 0-6 grading scale of increasing severity).[2] Magnesium intake did not differ between the groups and a mean was observed as 176mg/day.[16] Calcium intake was different in low and high calcium groups and the mean was calculated as 667mg/day and 1667mg/day respectively.[16] However, magnesium balance was not affected due to low or high calcium intake and average as 21mg/day.[16] In another experiment conducted, the mean depth of enamel loss was 12.8 µm for the 9 enamel slabs exposed to diet cola compared with 3.8 µm of enamel lost in the 9 enamel slabs exposed to distilled water.[17] This difference of 9.0 µm was

©Copyright protected. Unauthorised republication, reproduction, distribution, dissemination and copying of this document in whole or in part is strictly prohibited.

statistically highly significant. There was no difference in depth of loss of enamel between distilled water (5.0 µm), citric acid-based orange drink CCM (5.2 µm), and citric acid-based orange drink- CCM (6.1 µm, p 0.05).[17]

Calcium absorption from calcium carbonate was noted as 27% and that from CCM was noted as 36%.[5] It was concluded that calcium absorption from calcium citrate malate was significantly higher in these children compared to calcium absorption from calcium carbonate.[5] There were 5 patient-level data and 11 with trial-level data eligible for inclusion. In the studies inclined towards patient-level data, 143 people had myocardial infarctions which were allocated to calcium as compared to 111 allocated to placebo.[13] The meta-analysis of trial-level data showed similar results wherein 166 allocated to calcium has myocardial infarction and 130 to placebo.[13]

The adequate intake of calcium is important for the systematic and full development of bones. An experiment in rodents suggested that the adequate intake of Calcium affects the oral bone surrounding and supporting the teeth, especially alveolar bone.[6] Calcium-fortified food products help in preventing tooth loss and tooth erosion. Consumption of acidic food and beverages harms dental enamel erosion. There have been various studies to prove that CCM reduces dental erosion and reduces the erosive effect of much acidic food products.[17]

Role of CCM Post Surgery

After parathyroidectomy surgery, it is insisted that patients are observed primarily for calcium levels. In many cases, it was seen that a drop in calcium levels in serum occurred on the first day after surgery and lasted for some time after that. Over the years, parathyroidectomy has evolved into a minimally invasive procedure that can now be completed within 20 minutes for most patients, as it requires smaller dissections and lesser anaesthesia. A study shows that in 1997, 80% of patients were sent home immediately after surgery and by 2003, 100% were sent home except for a few rare and complex cases.[18] But, here there was a problem, the drop in serum calcium levels post-surgery was not accounted for and these patients dealt with complications like hypocalcaemia and definitely required supplements.[18]

Thyroidectomy surgery has been proven to be a safe technique. However, it does have one complication post-surgery, which is that of hypocalcaemia and this is the only reason that patients are kept under observation and not discharged early. Usually, patients are all given supplements after thyroidectomy as there are calcium losses. These losses can be temporary and can revert back faster in some cases, and giving supplements might cause more harm than good in these scenarios. Therefore, it is a new finding that it is better to check PTH levels on the day of thyroidectomy and along with the calcium levels, during discharge and before giving calcium supplements.[19]

People suffering from obesity usually go in for surgeries that involve a bypass that can lead to a huge loss of nutrients and this, in turn, results in complications. A decrease in calcium and vitamin D levels compromises bone health as it increases parathyroid hormone in the body, this condition is called secondary hyperparathyroidism. Secondary refers to a negative calcium balance that increases PTH (parathyroid hormone) levels, along with bone defects like osteopenia or osteoporosis.[20] Most protocols take into account the calcium (and now PTH) levels after surgery, but there would be a marked benefit seen in case pre-and post-operative supplementation was done.[21]

The comparison of the effects of Calcium carbonate and calcium citrate are directly done on calcium regulating hormones and marker of bone resorption in postmenopausal women that usually have a low intake of calcium and bone density. 40 postmenopausal women, for 12 weeks, were randomly administered 1000mg/day of calcium carbonate or calcium citrate with their meals.[22] After this time period, these women went through a 2 week washout time period without calcium supplements and after which they underwent a 12 week time period with alternate calcium supplements. Postmenopausal women with any disease or injuries related to bone health were not picked for this study as it would hamper the process.[22] Out of the 40 chosen women, 34 women completed the study as 6 were either not eligible or were withdrawn. Over the due course of the study, these women were evaluated for markers of bone resorptions, ionized calcium, PTH and urinary calcium excretion. In this time, the effect of Calcium Citrate was found to result in the decrease of all four markers of bone resorption the markers being NTX, CTX, Dpyr and sNTX.[22] This decrease did not occur with Calcium Carbonate. Calcium Citrate and Carbonate: Calcium Citrate had better bioavailability in 25 postmenopausal women, depicted by an increase in serum and urinary calcium levels, and PTH levels.[22] The reason for calcium citrate being more bioavailable when had with a meal is that it has a low gastric acid output since carbonate requires an acidic environment for absorption. Calcium carbonate does not work in patients that have achlorhydria, and here calcium citrate was significantly better. In the case of proton pump inhibitors and H2 blockers decrease acidity thereby making it difficult for carbonate to work. In terms of cost-benefit, calcium carbonate had the advantage.

A total of 6000 patients were taken for the study over a period of 4 years. All patients were ones suffering from primary hyperparathyroidism, and had parathyroidectomy surgery and were also discharged after about 2.5 hours. 3/4th of the population were women, and the remaining 1/4th were men.[18] The average age was 59.8 years and the ages ranged from 14-95 years. Patients were divided into groups based on their average calcium levels before surgery. Every patient of a group was given the same number of calcium supplement tablets (each having 315 mg of calcium citrate and 200IU vitamin D-3.[18] It is a controlled study with high doses given immediately post-surgery and then the amount tapers towards the next two

©Copyright protected. Unauthorised republication, reproduction, distribution, dissemination and copying of this document in whole or in part is strictly prohibited.

weeks. The first dose should be administered within 3 hours post-surgery.[18] Out of 6000 patients, 7 patients had severe complications and needed intravenous calcium, between days 3 and 5 post-surgery. Also, 5 out of these 7 had a case of 4-gland hyperplasia and underwent parathyroidectomy needed IV calcium, while only the other 2 with adenomas required IV calcium.[18] The most common symptoms of hypocalcaemia in these patients included hand tingling and paraesthesia, mental fog and hand cramping. Every patient was given 10 pills maximum per day, and if the need arose, then they were given 2 more every 2 hours, but no one would be given more than 10 pills in the 24-hour period without asking the surgeon.[18]

A study conducted in Christian Medical College Hospital in Vellore took into consideration 125 patients who had undergone thyroidectomy. The difference in hypocalcaemia between low and high PTH was that of 25%.[12] After the surgery, patients were observed for any complications such as haemorrhage, voice changes and any symptoms of hypocalcaemia. Samples taken to estimate total calcium in serum, albumin and PTH levels were done by 6 AM on the day that followed the surgery (no exceptions). Low serum calcium concentration gave an indication of hypocalcemia, but they also measured serum magnesium, to check if the hypocalcaemia was severe or resistant to calcium and vitamin D therapy.[12]

A study was conducted on 20 patients with prior consent to be included in the study. Patients who had other co-morbidities and medications that affect bone health were excluded. Choosing patients for the study was by conducting biochemical tests that showed any abnormalities in the alkaline phosphatase, vitamin D and/or PTH levels.[20] However, during statistical analysis and data tabulation, their age, body mass index, serum calcium levels were also taken into account. Patients were put in 2 separate groups- one treated with calcium citrate (600mg) with 400IU of vitamin D, given twice every day for 2 months; The other group was given calcium carbonate (600mg), with the same amounts of vitamin D, for the same duration.[20]

CCM in Gynaecology

Earlier, there were doubts about the effectiveness of calcium supplementation in improving bone density reduction in postmenopausal women and many contrasting evidences were found. A study had previously indicated that the rate of bone loss was higher in women who had a calcium intake under 400 mg per day.[24] Thus the subjects of this study were divided into two groups: the first group with daily calcium intake under 400 mg, and the second group with daily calcium intake above 400 mg. Further considering the fact that women closer to menopause (5 years or less) had a higher rate of bone loss compared to women who had menopause more than 5 years ago, the trial was divided into two phases, with early postmenopausal and late postmenopausal women.[24] In this two-year-long double-blind, placebo-controlled randomized trial, each category was subdivided into three groups: one received placebo tablets (microcrystalline cellulose), or 500 mg doses of either calcium carbonate or calcium citrate malate.[24] The early-postmenopausal women had a major decrease in bone density at the spine after 2 years of the study, regardless of their calcium intake. The femoral neck and the radius bone density did not change significantly. In the late- postmenopausal women with low dietary calcium intake, the ones who received calcium citrate malate showed no significant size of bone loss at any site.[24] The bone density of the spine decreased in both the calcium carbonate and placebo group, and the bone density of the femoral neck significantly decreased only in the placebo group. In the women with higher dietary calcium intake, the treatment didn't lead to any sort of change at any site. It was ultimately concluded that late postmenopausal women with low dietary calcium intake benefit the most from calcium supplementation.[24]

It has been established that bone density increases rapidly during puberty and the highest bone density at hip and spine of females is achieved by 20 years of age, the study was carried out to study the effect of calcium supplementation on the overall bone mineral content in adolescent girls. A total of 94 premenarchal girls with a mean age of 11.9±0.5 years participating in the study were randomly divided into two groups during the 18-month course of the trial: one group received placebo tablets and the other received 250 mg calcium citrate malate (6:2:3).[25] The dietary calcium intake of the groups did not differ. After 18 months, the mean height, weight, BMI and body fat percentage of both groups did not differ. Trends were the same for urinary gonadotropin, estradiol, testosterone and Integrated Estrogen Exposure score. The bone mineral density and bone mineral content, both overall and site-specific, showed a great increase in the supplemented group as compared to the placebo group.[25] Urinary calcium excretion was also higher in the supplemented group.[25] Thus, in conclusion, a modest increase in calcium intake via supplementation can help increase peak bone mass and avoid the possibility of osteoporotic fractures in future.

A double-blind, placebo-controlled 2 year-long trial was carried out to study the role of essential trace elements copper, zinc and manganese, in combination with or without supplemental calcium on spinal bone loss in older women.[26] 59 out of the total 120 participants completed the study. Among other reasons, 21 subjects dropped out of the study due to adverse side effects including gastrointestinal issues such as nausea, gas and indigestion, which were believed to be caused by supplements. The subjects were divided into four groups of study: one receiving only placebo tablets, the second receiving only active trace minerals, the third receiving only active calcium and the fourth receiving both active trace minerals and active calcium. Active calcium was given in the form of calcium citrate malate (ratio 6:2:3, 1000mg/ day). Active mineral tablets consisted of 15mg Zinc, 5mg Manganese and 2.5mg Copper.[26] At the end of the 2-year course of the trial, the greatest decline in bone mineral density was seen

©Copyright protected. Unauthorised republication, reproduction, distribution, dissemination and copying of this document in whole or in part is strictly prohibited.

in the group receiving placebo tablets (-3.53 ± 1.24), the decline was relatively less in the group receiving only trace minerals (-1.89 ±1.40) and even better in the group receiving only calcium supplements (-1.25 ±1.46).[26] The best results were obtained in the group receiving both active calcium and active trace minerals; there was no loss in bone mineral density, rather there was a net gain of 1.48 ±1.40.[26] This established that trace minerals, alongside calcium supplementation, are also important for improving and maintaining a healthy bone mineral density.

Other than bone mass and density, bone area and geometry could be equally important variables due to their role in bone strength. Thus, the primary aim of this calcium intervention trial was to analyse the effect on the bone area and bone density. For this study, 112 Caucasian premenarchal girls with a mean age of 11.9±0.5 years were shortlisted, 21 of which eventually dropped out at a regular rate throughout 24 months of this study.[27] These participants were divided into two groups: one group received placebo tablets (microcrystalline cellulose) and the other received 250mg CCM tablets (6:2:3 molar ratio).[27] To control the role of genetic variation in average BMI and initial bone density, the groups were divided by stratified randomization into two groups with balanced baseline values of BMI and LSBMD. At the end of the 2-year span, both groups showed normal growth and pubertal progression. Bone acquisition levels were higher in the calcium supplemented group than in the placebo group.[27] There was a 12-24% difference in the increases in the lumbar spine and pelvis bone measurements in supplemented groups as compared to the placebo group.[27] To further understand the role of estrogen in bone gain, Tanner score and IEES were both compared with TB BMD.[27] Even though IEES did not yield clear results when participants were subdivided into groups according to below- and above-median values for Tanner score and below- and above-median values for dietary calcium intake, the association between the two was significant. It was found that only the girls with above-median values of Tanner score benefited from calcium supplementation and had greater gains in TB BMD, the girls with below-median scores on the other hand showed no major effect of supplementation.[27]

Epidemiological studies have been both in favour of and against the effectiveness of calcium supplementation. Positive results have been efficiently obtained with much difficulty as some factors like multifactorial nature of osteoporosis, variable absorption of calcium, dietary calcium deficiency, unique biochemistry of postmenopausal women and readjustment of bone mass due to varying hormones (in both sexes) hinder the outcomes.[28] The majority of studies indicate that calcium supplementation cannot stop postmenopausal bone loss.[28]

Adolescents and the effect of calcium supplementation on their bone mass gain and accretion have been widely analysed. A study of 11-12-year-old adolescent females to analyse the effect of calcium supplementation showed an average increase of 24 grams of bone per year in the supplemented group.[28] If this trend was kept up for 30 years, the peak bone mass would have significant protection from postmenopausal bone loss. In another study in 45 pairs of pre-pubertal identical twins, one twin was given 700mg of CCM supplement.[28] Thus one twin had a daily intake of 1370 +/- 303 mg while the other had 888+/- 173 mg. There was a significant gain in bone mass in the supplemented twins.[28] Similar trials have been performed on adolescent girls who were divided into three groups receiving placebo or 500mg/day or 1000mg/day elemental calcium as CCM. The first two groups did not have a significant bone mass gain but the group 1000mg /day had an average skeletal mass gain of 29 +/- 7 gm. These results were, however, obtained at a daily calcium intake of 1618 +/- 288 mg which sheds light on the need to re-evaluate the recommended intake.[28]

Many studies have established that pre-and post-menopausal women benefit from calcium supplementation more than menopausal women. In Hugh-Dawson study with postmenopausal women, it was clear that none of the calcium supplements could prevent bone loss in menopausal women but in late postmenopausal women, CCM showed a 60-per cent reduction in spinal bone loss, which was a lot better than placebo and even calcium carbonate supplements (only 15 per cent reduction).[28] Better absorption of calcium citrate malate was demonstrated by an increase in urinary calcium in women with lower dietary intake. Later studies that used higher levels of calcium supplementation were unable to halt spinal bone loss at doses as high as 2500mg/day.[28] This points towards the importance of trace minerals (TMIN) such as copper, zinc and manganese. Thus, postmenopausal women were categorised as estrogen using and non-estrogen using groups and were given an only placebo, only TMIN, only CCM or CCM plus TMIN.[28] No differences in bone mass were noticed in estrogen using women, but in non-estrogen using groups, the greatest decline in bone mineral density was seen in the placebo group, then in the TMIN group and the least decline was in the CCM group. Additionally, in the CCM plus TMIN group, there was no decline, rather there was a slight gain in bone mineral density.[28] The effect of the combination of CCM and vitamin D in postmenopausal women living at 42-degree latitudes was also studied. They are unable to manufacture vitamin D during the winter months. The group receiving vitamin D and CCM showed an increase in bone mineral density, while the other groups had a net decrease.[28]

Insufficient dietary intake of calcium and vitamin D is a primary concern because when calcium or vitamin D is lower in the body, calcium homeostasis is maintained by parathyroid hormone at the expense of bone.[29] Thus lack of calcium accumulation negatively impacts peak bone mass. Many studies have shown that increasing dietary calcium intake up to 1200-1600 mg/ day has positive effects on bone acquisition.[29] In a study with subjects aged

©Copyright protected. Unauthorised republication, reproduction, distribution, dissemination and copying of this document in whole or in part is strictly prohibited.

7-17 years, the group of subjects with higher calcium intake had 1.6-5.1% higher bone density than the control group.[29] Similar results were obtained in a co-twin placebo-controlled trial where there was a 3.7% difference in TBBMC when one consumed a more calcium-rich diet than the other. Vitamin D deficiency leads to decreased calcium absorption which is a prime cause of bone abnormalities and rickets due to defective mineralization and altered growth plate formation.[29] Studies have also related vitamin D deficiency with a rise in parathyroid hormone levels. Vitamin D receptor is important in all calcium-regulated tissues as it is in ovarian function. Vitamin D deficient mice were less fertile compared to replete littermates. It has even led to uterine hypoplasia and impaired folliculogenesis, alongside low levels of estradiol and higher LH and FSH.[29] VDR-null mutant mice showed decreased gene expression as well as lower activity of estrogen synthase or aromatase; estrogen supplementation increased uterine weight in these mice and normalised histological abnormalities. Calcium supplementation in VDR- null mice corrected low serum levels, and more importantly, improved CYP19 gene (encodes P450 of aromatase) expression to 20% and aromatase cytochrome 450 activity up to 60% of the normal.[29] Thus calcium and vitamin D have a key role in estrogen biosynthesis and normal aromatase gene expression.

Postmenopausal women are at a greater risk of obesity due to reduced circulating sex hormones and metabolism disbalance due to the absence of the menstrual cycle. Calcium and Vitamin D supplementation have proven benefits in improving body composition: high calcium intake is associated with high-fat oxidation and increased satiety; improved vitamin D signaling enhances expression of lipolytic and thermogenic genes in adipose tissue.[30] A preclinical trial also reported that higher dairy calcium intake and higher serum vitamin D levels are associated with greater diet-induced weight loss, and as reported in another study they may even facilitate weight loss. Contrasting evidence has given rise to another possibility that dairy calcium has this effect due to other bioactive compounds present in dairy working in unison with calcium. This was a randomized, parallel armed study carried out in 128 subjects to identify the impact of calcium supplementation (with or without vitamin D) on different groups of women having different diet regimes but having exercise training regimes for 14 weeks.[30] There were three groups divided based on the assigned diet regime:

(1)No dietary modification or CTL group

(2) Low-calorie high- protein or LCHP diet group

(3) low-calorie high-carbohydrate or LCHC diet group.

Each of the diet based groups was subdivided into three categories which were given either:

(1) 800 mg of elemental calcium as calcium carbonate (Ca),

(2) 800mg calcium (calcium tetrahydrate + calcium malate) with 400 IU/day of vitamin D, 300 mg of magnesium, 7.5 mg of zinc, 2 mg of copper and 2 mg of manganese (Ca+D)

(3) maltodextrin placebo.

Analyses at the end of the 14-week intervention showed a major decline in resting energy expenditure by 80 kcal/day. This decrease was greater in the Ca+D group during the weight-loss intervention but not in comparison to total body weight. LCHC and LCHP, both groups experienced great fat loss, this loss was greater in the Ca+D subcategory compared to Ca supplemented subcategory in the LCHC group.[30] Supplementation showed no effect on weight loss in the LCHF diet group. Thus calcium supplementation benefits people having a hypo energetic diet during weight loss. Finally, supplementing calcium supplementation with vitamin D does not have any major benefit with weight loss compared to only calcium supplementation.[30]

Role of CCM in Orthopedics

Calcium is under close homeostatic control with cycles like assimilation, discharge, and emission and capacity in bone being engaged with keeping up with the grouping of ionized Ca in the plasma inside a firmly managed range.[31] This tight guideline of plasma Ca focus is accomplished through a complex physiological framework containing the association of the calcitropic chemicals, like parathyroid chemical (PTH), 1,25 dihydroxycholecalciferol (1,25 (OH)2D3), and calcitonin, with explicit objective tissues (kidney, bone, and digestive tract) which serve to increment or to diminish the section of Ca into the extracellular space.[31]

Investigations of Ca supplementation in postmenopausal ladies, commonly of one to two years, have shown that Ca forestall bone misfortune can decrease the pace of bone misfortune partly. These investigations uncover that the viability of Ca fluctuates by skeletal site, by menopausal age, and with common Ca admissions of the examination subjects.[31] In an examination researching, caffeine consumption in 489 ladies matured 66 – 77 years, there was a positive relationship between caffeine admission and bone loss. Bone misfortune at the spine was more prominent in ladies devouring 300 mg of caffeine each day than in ladies burning through 300 mg/d.[32] Ladies with the tt hereditary variation of VDR had fundamentally more noteworthy deficiency of bone at the spine than ladies with the TT hereditary variation when their admission was 300 mg/d.[32] The most widely recognized types of calcium accessible to the buyer are calcium carbonate and calcium citrate. Different types of calcium incorporate lactate, gluconate, bone feast, and hydroxyapatite. Calcium supplements are accessible as containers, tablets, bites, wafers, powders, and fluids.

Calcium citrate malate is portrayed as a metastable complex of calcium, citrate, and malate or as a combination of calcium salts involving the calcium salt of citrus extract and malic corrosive.[5] The metastable materials have more than one glasslike state, reflected by

©Copyright protected. Unauthorised republication, reproduction, distribution, dissemination and copying of this document in whole or in part is strictly prohibited.

the presence of various hydration states. Calcium citrate malate is depicted as containing between 20 – 24 % of calcium on a dry premise having a pH of 4.0 – 8.0.[5] Hefty metal degrees of under 20 mg/kg are proposed, for lead two solicitors propose a most extreme degree of 2 mg/kg and two others propose a limit of 10 mg/kg.[33] The impact of calcium supplementation on skeletal advancement in four gatherings of 14 C/D female rodents controlled for 4 or 12 weeks as calcium citrate malate.[5]

Customary ingestion of satisfactory calcium (Ca), especially via utilization of food varieties normally high in Ca, requires a conscious way to deal with adjusted nourishment for an enormous number of individuals. Ordinary openness to an inexorably wide assortment of engaging food and refreshment decisions can redirect the accentuation and inclination from innately Ca-rich food sources that very frequently are to a great extent under-consumed by those most needing them.[33] Under particular conditions, like lactose narrow-mindedness, hypersensitivities, severe vegetarianism, and individual abhorrences, dairy food varieties are not considered as a feasible choice or are burned-through in deficient amounts to meet Ca prerequisites by numerous subgroups.[33]

The Food and Drug Administration (FDA) is altering its marking guideline approving a wellbeing guarantee on the connection among calcium and a decreased danger of osteoporosis to incorporate nutrient D so that, notwithstanding the case for calcium and osteoporosis, an extra case can be made for calcium and nutrient D and osteoporosis; dispense with the prerequisite that the case list sex, race, and age as explicit danger factors for the advancement of osteoporosis; dispose of the necessity that the case doesn't state or infer that the danger of osteoporosis is similarly material to the overall U.S. populace, and that the case distinguish the populaces at specific danger for the improvement of osteoporosis; wipe out the necessity that the case recognize the component by which calcium diminishes the danger of osteoporosis and rather make it discretionary; dispose of the prerequisite that the case incorporate an articulation that a complete dietary admission more noteworthy than 200% of the suggested every day consumption (2,000 milligrams (mg) of calcium) has no further advantage to bone wellbeing when the food contains 400 mg or a greater amount of calcium per reference sum usually devoured or per all out day by day suggested supplement admission; and permit reference for the need of active work in both of the wellbeing professes to be discretionary rather then required.[34]

Ca assimilation happen prevalently in the jejunum and the ileum and colon. Take-up happens by dynamic vehicle and basic aloof dispersion.[35] At low Ca admissions dynamic vehicle prevails, however as admissions increment more is consumed by vague pathways.[35] The metabolite of nutrient D (1,25-dihydroxycholecalciferol) animates Ca transport across the intestinal cells by inciting the creation of a Ca-restricting protein.[35] This cycle happens inside the villus cells through the ordinary interaction of receptor restricting, DNA communication, and courier RNA creation. Thus, nutrient D is basic for powerful Ca ingestion.[35]

Around 97% of the sifted Ca load is re-consumed by the renal tubules and the leftover 3%, which is discharged in pee, addresses the mandatory misfortune.[35] The significant determinant of urinary Ca discharge is dietary Ca admission. Urinary Ca discharge is higher when protein and Na admissions are high than when admissions of these two supplements are low.[35]

Intense oral calcium load showed, through the concurrent expansion in serum calcium and urinary calcium discharge, that the bioavailability of the marine calcium sources is as great (Phoscalim) or better (Glycollagen) as a similar admission of dairy calcium from milk.[36] Besides, the marine calcium load actuated a more prominent diminishing of iPTH in the Glycollagen bunch than milk and a more noteworthy lessening in bone resorption reflected by s-CTX, since Phoscalim has an almost comparable impact on serum PTH than milk and a lesser impact on s-CTX than milk.[36]

This examination was intended to decide if a was diminished in kids with CF and to build up impacts of calcium and nutrient D3 supplementation on a. Likewise, it was intended to decide if expanded dietary enhancements of nutrient D3 and calcium would further develop serum minerals, markers of bone turnover, and estimations of bone mineral substance (BMC) in these kids. Subjects were enlisted from the Cystic Fibrosis Clinic at the University of Missouri. The Institutional Review Board of the School of Medicine supported the investigation and all guardians gave composed assent with all kids more than seven giving consent. Kids somewhere in the range of 3 and 15 years were qualified, the two guys and females. Kids stayed on their standard drugs, which included pancreatic compounds and generally ADEK nutrients containing another 400 IUD. Kids on oral or intravenous glucocorticoids were avoided. Leather treater Staging utilizing a self-distinguished picture framework was done toward the start of each stage. The significance of Ca supplementation for bone wellbeing is grounded and, along with nutrient D, is considered as a vital part of any preventive or restorative routine for osteoporosis.[37] The gainful impact of long haul Ca and nutrient D supplementation on bone and break hazard in old ladies has additionally been recorded in a few clinical preliminaries.[37]

Bioavailability in the gut and for bone is the initial standard to consider while choosing a Ca salt.[38] Great intestinal absorbability is required, yet it is just the first step in quite a while of bioavailability for both bone mineral growth and bone misfortune avoidance, especially in the old. Some dietary variables impact bone Ca accumulation and additionally resorption, just as urinary, Ca discharge. Accordingly, this load of components will decide the general bioavailability of different Ca salts. The worth of

©Copyright protected. Unauthorised republication, reproduction, distribution, dissemination and copying of this document in whole or in part is strictly prohibited.

advantageous Ca and nutrient D in bone wellbeing is grounded, there is an unmistakable absence of agreement in regards to different nutrients and minerals embroiled in bone wellbeing.[38] Ca and nutrient D supplementation is suggested in postmenopausal ladies and the old, and instances of osteoporosis. Calcium lysine doesn't ionize in the gut, in this manner wiping out the possibility to meddle with iron ingestion. L-lysine can both upgrade intestinal Ca assimilation and work on the renal protection of the retained Ca.[38] The consolidated impacts might add to a positive Ca balance, subsequently recommending the likely convenience of L-lysine supplements for both preventive and remedial intercessions in osteoporosis.[38]

DISCUSSION

Calcium is the fifth most abundant mineral in our body. Calcium supplementation plays a vital role in maintaining bone health throughout life. There is a difference in the retention values in both A and B experiments in rats. Younger animals absorb better and have higher turnover. Both factors elevate retention. The results from the CCM supplemental group provide evidence that there has to be an increase in the calcium intake above RDA (Recommended Dietary Allowance) to optimize bone mass acquisition during childhood and adolescence.

With the increasing use of soft drinks, especially in adolescent groups, there is a growing concern in the dental sector. The problem is that product reformulation is not a simple matter and many suggested additives can have adverse effects. It is necessary to keep in mind that acids also have a preservative effect so if there is any reduction in acidity levels, it should be observed.[17] This study reflects chronic magnesium utilization and not acute responses to the increase in calcium diet. There were several limitations in the study conclusions. It can be possible that the effect of calcium on magnesium utilization takes place gradually and not immediately in a period of 14 days. Calcium intake does not influence magnesium utilization.[16]

Although the enamel slab is a well-known system, its usage as an experimental erosion system was new. The palatal area of the maxillary anterior teeth is the surface that is most exposed and affected by erosion. There was a loss of enamel, to some extent, when exposed to distilled water. It may be caused by erosion during the 6 days and nights of intra- oral use. There are two ways to design these experiments.

1) leaving the slabs in the mouth all of the time and rinsing with the test solution. This method measures the real-life situation.

2) insert the slabs into test solutions out of the mouth before putting the appliance back into the mouth. This method allows the comparison between two solutions which helps in controlling and reducing the experimental variations.[16]

Calcium from CCM is more bioavailable, around 8-15% than from other sources of calcium.[5] This does not affect the safety of use in the long term, supplementation studies show that CCM is well tolerated by humans. For many people around the globe, the adequate intake of calcium is quite low. This is a major health problem and is getting worse day by day. People are attracted to food products that have acidic effects without even knowing their downfalls. Calcium-fortified food products must be palatable to the extent that their taste and texture are not affected. The cost of habitually purchasing good quality Ca-fortified foods and supplements is negligible compared to the exorbitant costs associated with substandard nutrition and the resulting pathologies.[6] From the experiments conducted, it can be concluded very clearly that calcium supplements are linked with an increased risk of myocardial infarction and cardiovascular disease. As calcium supplements are in great demand nowadays, these risks can turn into a big burden on the world population.

Calcium absorption has to vary differently in healthy adult humans as they have their normal self-selected diets having different Ca sources. The conclusions were limited due to the use of two different groups of women. Although they were from the same age group and in the same geography, there was not enough information regarding their vitamin D status or self-selected dietary intake. Calcium fortification is an important aspect for increasing the calcium intake in people especially those who are lactose intolerant. The most effective way to prevent the loss of tooth minerals is the modification of beverages with the fortification of calcium. The effect of calcium fortification on oral pH changes and taste perception should be evaluated in further studies.

Doctors are still unsure of the safety of the parathyroidectomy since hypocalcaemia occurs anytime past the second day after surgery and patients go into complications if precautions are not taken. Every patient must comply with the calcium supplementation therapy immediately post-surgery, irrespective of how successful the surgery was or what their serum levels were. This gives a broader range of control over the post-operative measures and can keep any unwanted things coming the patient's way in check.

Usual therapy for patients with end-stage renal disease due to other morbidities is to put them on calcium supplements with activated vitamin D (calcitriol or its analogues). Many go with calcium carbonate since it is more cost-effective, but in cases where there are antacids in the medication that the patients take, it is difficult to see the real effectiveness of calcium carbonate. In fact, in general, calcium carbonate has about 40% calcium content, while calcium citrate has only a meagre 21.1%. But since calcium citrate does not require an acid environment and due to much higher bioavailability, it tends to have a significantly greater effect on lowering PTH levels. In this case study, calcium citrate was used for a very short time to accelerate the calcium deposition into the bone post-surgery.

©Copyright protected. Unauthorised republication, reproduction, distribution, dissemination and copying of this document in whole or in part is strictly prohibited.

This research should be extended to patients with other morbidities and also the interactions of these calcium supplements must be studied so that there are no side effects to this on top of the existing complications. Doctors should take note and observe what calcium levels are, before, during and after the surgery. But sometimes, the calcium levels after surgery might drop later and thus there is a need for preventive measures so that the hypocalcaemia even if it comes, will be mild and well within handleable limits.

The upper limit for a single dose is 500 mg, and it was seen that only doses below this limit gave the highest rates of absorption of calcium. The dose of calcium had a direct effect on the PTH levels, the higher the dose, the lower the PTH, thus it proved vital in eradicating bone defects and also secondary hyperparathyroidism. The greater bioavailability of CCM is due to the fact that malate and citrate are easily ionized, hence their increased absorption rates. Citrate and malate ions are absorbed usually in the upper digestive tract. There have also been studies done on single and double stable isotopes, and it was found that the stable isotopes are 44Ca and 42Ca, which were found in serum and urine. It was also understood through tracer as well as varied molar ratios, that CCM absorption exceeds that of the calcium absorbed from various dietary intakes and other calcium supplements. Although it is fair to note that calcium lysinate might change this due to its chelated nature.

The latest development in the field of calcium supplementation after surgeries is because PTH levels are a stronger indicator when measured alongside calcium, and this adds to the efficiency of the process post-operation. Further research can be done based on this as this can be a very good foundation in terms of the accuracy of the result as well as having an added bonus of being a randomized trial. Key experiments done could be extensions of these into a larger group of people including some patients with certain pre-existing defects or medications.

Coming to vitamin D along with calcium supplements, it was also seen that the use of Vitamin D along with the calcium administered resulted in higher levels of calcium in the patients' bodiesy and reduced risk for them. Hence, postoperative early administration of Vitamin D or Calcium supplements or both together showed higher chances for the patient to not be hypocalcaemic. Unfortunately, the study of pre-operative administration of Calcium and Vitamin D supplements has not resulted in any solid evidence and has been seen to have many limitations.

A new form of calcium supplementation that might change the scenario is that of calcium lysinate, which has the advantage of the fact that L-lysine has a role in enhancing calcium absorption from the intestine as well as in improving renal conservation of absorbed calcium. Thus, proving useful in both prevention and also curing of osteoporosis. In today's world, there's always a concern on how long someone is at the hospital and how much it would cost to do a treatment. But there is never a perspective that revolves around how meagre the cost of taking up treatment is compared to paying to fix the complications. By giving patients the necessary supplements, based on their postoperative calcium concentrations in their serum, their stay at the hospital is minimized. Yes, there is no doubt that their operative costs will go up, but definitely no complications, and that would mean no return to the hospital after this. A combination of calcitriol and calcium seems to be less expensive as they work with most patients, and patient's bodies too comply with this combined supplementation in most cases. This is because many calcium supplements are selective and they have to dive deep into details of a patient's proton pump characteristics, gastric acid secretions and other drugs that might interact with the supplement. There is a downside to giving all patients supplementation, as that might increase calcium levels, in case a patient actually copes with the calcium loss during surgery. But then again, if monitored properly then it would not cross the maximum limit.

There are upcoming strategies now to supplement based on intact PTH levels and that can be a boon in handling calcium supplementation better.

CONCLUSION

This research review's purpose is to help the reader understand different aspects posed by the research on the analeptic properties of Calcium Citrate Malate. This is significant because it gives insights into how CCM can be used for the treatment and management of various diseases in different aspects of medicine like gynaecology, orthopaedics and even ENT. It is noteworthy to understand the analeptic properties of CCM post-surgery. There has been much research and discussion conducted on these opinions of the pharmacological properties of CCM**.** Most of the research found was on the CCM supplementation post-surgery, in menstruating women to combat osteoporosis. More research and testing are required to gain a better understanding of the effective ways in which CCM can be used to treat and manage various other diseases and disorders.

Acknowledgement: We would like to thank our supervisor/guide Bharat Kwatra, from Invenzion Labs Inc and sub mentor Chelsea Rumao, whose expertise was invaluable in formulating the research questions, methodology and drawing conclusions. His insightful feedback and guidance pushed us to sharpen our thinking and brought our work to a higher level.

Ethics Approval and Consent to Participate

Not applicable.

Human and Animal Rights

No Animals/Humans were used for studies that are the basis of this research.

©Copyright protected. Unauthorised republication, reproduction, distribution, dissemination and copying of this document in whole or in part is strictly prohibited.

Consent for Publication

Not applicable.

Availability of Data and Materials

The author confirms that the data supporting the findings of this research are available within the article.

Funding Acknowledgement and Conflict of Interest

The authors whose names are listed immediately above certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

REFERENCES

1. Trailokya A, Srivastava A, Bhole M, Zalte N. Calcium and calcium salts. J Assoc Physicians India. 2017;65(2): 100-103.

2. Smith K, Heaney R, Flora L, Hinders S. Calcium absorption from a new calcium delivery system (CCM). Calcified Tissue International. 1987;41(6):351-352.

3. Andon MB, Lloyd T, Matkovic V. Supplementation trials with calcium citrate malate: Evidence in favor of increasing the calcium RDA during childhood and adolescence. In: Journal of Nutrition. 1994; 124(8): 1412S- 1417S.

4. Roth-Bassell HA, Clydesdale FM. In vitro solubility characteristics of six calcium salts. Vol. 55, Journal of Food Protection. 1992; 55(12):1003-1005.

5. Calcium citrate malate as source for calcium for use in foods for Particular Nutritional Uses and in foods for the general population (including food supplements) - Scientific Opinion of the Panel on Food Additives, Flavourings, Processing aids and Materials in Contact with food (AFC). EFSA J. 2007;5(12):1-24.

6. Reinwald S, Weaver CM, Kester JJ. The Health Benefits of Calcium Citrate Malate: A Review of the Supporting Science. In: Advances in Food and Nutrition Research. 2008;54: 219-346.

7. Andon MB, Peacock M, Kanerva RL, De Castro JAS. Calcium absorption from apple and orange juice fortified with calcium citrate malate (CCM). J Am Coll Nutr. 1996;15(3): 313-316.

8. Rzymski P, Pischel I, Conrad F, Zwingers T, Rzymski P, Opala T. The bioavailability of calcium in the form of pyruvate, carbonate, citrate–malate in healthy postmenopausal women. European Food Research and Technology. 2015;242(1):45-50.

9. Shankar K, M S, Raizada P, Jain R. A Randomized Open-Label Clinical Study Comparing the Efficacy, Safety, and Bioavailability of Calcium Lysinate with Calcium Carbonate and Calcium Citrate Malate in Osteopenia Patients. J Orthop case reports. 2018;8(4)15-19.

10. Palermo A, Naciu A, Tabacco G, Manfrini S, Trimboli P, Vescini F et al. Calcium citrate: from biochemistry and physiology to clinical applications. Reviews in Endocrine and Metabolic Disorders. 2019;20(3):353-364.

11. Sampoorna DM, Bhavani S V., Jayasurya K, kumar BK. Calsiara Tablets: A Clinical Study on Calcium Citrate malate for Bone Health. Asian J Appl Sci Technol. 2020;04(04): 130-138.

12. Alhefdhi A, Mazeh H, Chen H. Role of Postoperative Vitamin D and/or Calcium Routine Supplementation in Preventing Hypocalcemia After Thyroidectomy: A Systematic Review and Meta-Analysis. Oncologist. 2013;18(5):533-542.

13. Bolland M, Avenell A, Baron J, Grey A, MacLennan G, Gamble G et al. Effect of calcium supplements on risk of myocardial infarction and cardiovascular events: meta-analysis. BMJ. 2010;341(jul29 1):c3691-c3691.

14. Franklin S, Masih S, Thomas AM. Effect on oral pH changes and taste perception in 10–14-year-old children, after calcium fortification of a fruit juice. Eur Arch Paediatr Dent. 2015;16(6): 483-489.

15. Grenby TH. Lessening dental erosive potential by product modification. Vol. 104, European journal of oral sciences. 1996;104(2):221-228.

16. Andon MB, Ilich JZ, Tzagournis MA, Matkovic V. Magnesium balance in adolescent females consuming a low- or high-calcium diet. Am J Clin Nutr. 1996;63(6):950-953

17. Rugg-Gunn AJ, Maguire A, Gordon PH, McCabe JF, Stephenson G. Comparison of Erosion of Dental Enamel by Four Drinks Using an Intra-Oral Applicance. Vol. 32, Caries Research. 1998; 32(5): 337-343

18. Vasher M, Goodman A, Politz D, Norman J. Postoperative Calcium Requirements in 6,000 Patients Undergoing Outpatient Parathyroidectomy: Easily Avoiding Symptomatic Hypocalcemia. J Am Coll Surg. 2010;211(1):49-50.

19. Sekar S, Belavendra A, Jacob P. Early discharge and selective calcium supplementation after thyroidectomy based on post-operative day 1 parathormone and calcium level: A prospective study. Indian Journal of Endocrinology and Metabolism. 2020;24(4):319.

20. Baretta G, Cambi M, Rodrigues A, Mendes S. Secondary Hyperparathyroidism after Bariatric Surgery: treatment is with Calcium Carbonate or Calcium Citrate?. ABCD Arquivos Brasileiros de Cirurgia Digestiva (São Paulo). 2015;28(suppl 1):43-45.

21. Afshan S, Farah Musa AR, Echols V, Lerant AA, Fülöp T. Persisting Hypocalcemia After Surgical Parathyroidectomy: The Differential Effectiveness of Calcium Citrate Versus Calcium Carbonate With Acid Suppression. Am J Med Sci. 2017;353(1): 82-86.

22. Kenny AM, Prestwood KM, Biskup B, Robbins B, Zayas E, Kleppinger A, et al. Comparison of the effects of calcium loading with calcium citrate or calcium carbonate on bone turnover in postmenopausal women. Osteoporos Int. 2004;15(4): 290-294.

23. Straub DA. Calcium supplementation in clinical practice: A review of forms, doses, and indications. Nutrition in Clinical Practice. 2007; 22(3):286-296.

24. Dawson-Hughes B, Dallal GE, Krall EA, Sadowski L, Sahyoun N, Tannenbaum S. A Controlled Trial of the Effect of Calcium Supplementation on Bone Density in Postmenopausal Women. N Engl J Med. 1990;323(13): 878-883.

25. Lloyd T, Andon MB, Rollings N, Martel JK, Landis JR, Demers LM, et al. Calcium Supplementation and Bone Mineral Density in Adolescent Girls. JAMA J Am Med Assoc. 1993;270(7):841-844.

©Copyright protected. Unauthorised republication, reproduction, distribution, dissemination and copying of this document in whole or in part is strictly prohibited.

26. Strause L, Saltman P, Smith KT, Bracker M, Andon MB. Spinal bone loss in postmenopausal women supplemented with calcium and trace minerals. J Nutr. 1994;124(7): 1060-1064.

27. Lloyd T, Martel JK, Rollings N, Andon MB, Kulin H, Demers LM, et al. The effect of calcium supplementation and Tanner Stage on bone density, content and area in teenage women. Osteoporos Int. 1996;6(4): 276-283.

28. Patrick L. Comparative absorption of calcium sources and calcium citrate malate for the prevention of osteoporosis. Vol. 4, Alternative Medicine Review. 1999; 4(2):74-85.

29. Harkness LS, Bonny AE. Calcium and vitamin D status in the adolescent: Key roles for bone, body weight, glucose tolerance, and estrogen biosynthesis. Vol. 18, Journal of Pediatric and Adolescent Gynecology. 2005; 18(5):305-311.

30. Kerksick CM, Roberts MD, Campbell BI, Galbreath MM, Taylor LW, Wilborn CD, et al. Differential impact of calcium and vitamin d on body composition changes in post-menopausal women following a restricted energy diet and exercise program. Nutrients. 2020;12(3):713.

31. Cashman K.D. Calcium intake, calcium bioavailability and bone health. British Journal of Nutrition. 2002;87(6):169-177.

32. Straub DA. Calcium supplementation in clinical practice: A review of forms, doses, and indications. Nutrition in Clinical Practice. 2007;22(3):286–96.

33. Reinwald S, Weaver CM, Kester JJ. The health benefits of CALCIUM Citrate Malate: A review of the supporting science. Advances in Food and Nutrition Research. 2008;54:219–346.

34. Food labeling: health claims; calcium and osteoporosis, and calcium, vitamin D, and osteoporosis. Final rule. Fed Regist. 2008;73(189):56477-56487.

35. Lanham-New SA. Importance of calcium, vitamin D and vitamin K for osteoporosis prevention and treatment. In: Proceedings of the Nutrition Society. 2008; 67(2):163-176.

36. Lecerf JM, Lamotte C, Boukandoura B, Cayzeele A, Libersa C, Delannoy C, et al. Effects of two marine dietary supplements with high calcium content on calcium metabolism and biochemical marker of bone resorption. Eur J Clin Nutr. 2008;62(7):879-884.

37. Hillman LS, Cassidy JT, Popescu MF, Hewett JE, Kyger J, Robertson JD. Percent true calcium absorption, mineral metabolism, and bone mineralization in children with cystic fibrosis: Effect of supplementation with vitamin D and calcium. Pediatr Pulmonol. 2008;43(8):772-780.

38. Bonjour JP, Guéguen L, Palacios C, Shearer MJ, Weaver CM. Minerals and vitamins in bone health: The potential value of dietary enhancement. Vol. 101, British Journal of Nutrition. 2009;101(11):1581-1596.

Source of Support: The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

For any question relates to this article, please reach us at: editor@globalresearchonline.net

New manuscripts for publication can be submitted at: submit@globalresearchonline.net and submit_ijpsrr@rediffmail.com

©Copyright protected. Unauthorised republication, reproduction, distribution, dissemination and copying of this document in whole or in part is strictly prohibited.

View publication stats

SmartSplit: Latency-Energy-Memory Optimisation for CNN Splitting on Smartphone Environment

Ishan Prakash
Delhi Technological University
New Delhi, India
ishan.0114@gmail.com

Aniruddh Bansal
Delhi Technological University
New Delhi, India
aniruddhbansal2000@gmail.com

Rohit Verma
University of Cambridge
Cambridge, UK
rv355@cam.ac.uk

Rajeev Shorey
UQIDAR, IIT Delhi
New Delhi, India
rshorey@iitd.ac.in

arXiv:2111.01077v1 [cs.LG] 1 Nov 2021

***Abstract*—Artificial Intelligence has now taken centre stage in the smartphone industry owing to the need of bringing all processing close to the user and addressing privacy concerns. Convolution Neural Networks (CNNs), which are used by several AI applications, are highly resource and computation intensive. Although new generation smartphones come with AI-enabled chips, minimal memory and energy utilisation is essential as many applications are run concurrently on a smartphone. In light of this, optimising the workload on the smartphone by offloading a part of the processing to a cloud server is an important direction of research. In this paper, we analyse the feasibility of splitting CNNs between smartphones and cloud server by formulating a multi-objective optimisation problem that optimises the end-to-end latency, memory utilisation, and energy consumption. We design *SmartSplit*, a Genetic Algorithm with decision analysis based approach to solve the optimisation problem. Our experiments run with multiple CNN models show that splitting a CNN between a smartphone and a cloud server is feasible. The proposed approach, *SmartSplit* fares better when compared to other state-of-the-art approaches.**

***Index Terms*—smartphones, CNNs, multi-objective optimisation, edge computing**

I. Introduction

With the first set of AI applications launched on smartphones in 2017 [1], the AI on smartphone industry has seen several improvements and use-cases. These use-cases, many of which utilise Convolution Neural Networks (CNNs) [2], include facial recognition, natural language translation, behaviour prediction, sensor data analysis for activity recognition, and many more [3], [4]. Unlike other deep learning approaches that utilise matrix multiplication, CNNs utilise convolution in at least one of the network layers. Convolution is a resource and computation-intensive task. Seeing the importance of AI on smartphones the smartphone industry has been building specialised chips. The number of smartphones with specialised AI hardware has increased from 3% in 2017 to 33% in 2020 and is expected to grow further [5]. Huawei's Kirin processor [6] and Apple's Bionic AI Chip [7] are some of the popular examples of such specialised hardware.

A few crucial points need to be considered while running CNNs on the smartphone environment. First, although the new AI-enabled chips are suitable for running CNNs on smartphones, reducing resource and computation requirements is essential as the remaining resource or computation power could be utilised by other processes running simultaneously on the smartphone. Second, the AI-enabled chips are only available for the newer smartphones and this restricts the reach of several AI applications. Furthermore, a strategy that is likely work for earlier generation of smartphones would be useful for the newer ones to optimise already efficient AI-enabled chips.

Several studies have been done to expand the reach of AI on smartphones. One such direction of study develops compressed architecture of existing CNN models [8]–[10]. However, these approaches are limited to the type of model they are compressing. Moreover, compressed architectures almost always compromise on accuracy. Another direction of research is to develop lightweight libraries that could be used to implement CNNs [11]. Some works prune the redundant and non-informative weights from the network architecture to develop models similar to the compressed models. However, both lightweight libraries and network pruning compromise on accuracy [12]. Another popular approach is splitting the CNN between the smartphone and a cloud server where a part of the processing is done on the cloud server. Such splitting could be done using N-step algorithms [13] or using latency-based optimisation [14]. These approaches have three major shortcomings when applied to smartphones. First, splitting-based approaches have been specifically designed for edge devices designed to run AI applications. The constraints that work in the context of edge devices would be different compared to smartphones. Smartphones are used for performing concurrent tasks that need to be taken into consideration. Second, these approaches optimise only a single system parameter, be it memory utilisation [13] or latency [14]. However, considering the multi-tasking capability of smartphones, we need to optimise both the parameters. Optimising latency is important; otherwise, the non-splitting approaches would be the better alternative. Furthermore, optimising memory utilisation is crucial as smartphone memory is used for additional parallel processes. Finally, none of the state-of-the-art works considers optimising the energy consumed when running computation-extensive applications.

In light of the limitations of the existing works, we formulate a multi-objective optimisation problem that optimises latency, memory utilisation, and energy consumption when a CNN is split between a smartphone and a cloud server. The key contributions in this paper are as follows;

- Design a latency and energy model for a CNN being split

between a smartphone and the cloud server (§ III).
- Formulate a multi-objective optimisation problem that defines three objective functions to optimise latency, memory utilisation, and energy consumption along with the necessary constraints (§ IV).
- Design a genetic algorithm and decision analysis approach *SmartSplit* to solve the formulated multi-objective optimisation problem. (§ V).
- Design an Android Application that uses PyTorch for Android [15] for running the CNN models post splitting between the smartphone and the cloud server.

We evaluate our optimisation algorithm under varying conditions and compare it with different state-of-the-art approaches. Our results show that *SmartSplit* fares better than the existing approaches (§ VI). In the next section, we briefly describe the related literature in the area (§ II).

II. Related Work

Recent strides in Deep learning techniques, namely Convolution Neural Networks (CNN) [2] have surpassed conventional computer vision techniques and have given superior accuracy for tasks ranging from image classification [2], [8], [16]–[18] to action recognition. Convolution Neural Networks are a class of Neural Networks using convolution operation in at least one of the layers. However, CNN-based techniques are computationally expensive as they have a large number of model parameters. The performance of CNN on mobile devices has been suboptimal compared to running CNN on superior computing hardware such as GPUs and Neuromorphic chips.

Over the last few years, the computational power of mobile devices has increased drastically [19]. This advancement prompts the development of techniques that will enable us to utilise the computing power by running CNNs on resource-constrained mobile devices. Several bottlenecks are encountered when running CNNs on mobile devices such as network delays, hardware limitations, and privacy concerns [20]. There exist different techniques that can overcome the bottlenecks. Some techniques focus on CNN memory management; they address this problem by using feature map pruning and using certain encoding strategies to compress the neural network [12], [21]. Memory reduction can also be achieved by designing efficient neural networks [22]. Another class of techniques focus on decreasing the complexity of the network [23]. Tan et al. [24] design CNN models using automated search (AutoML). There are lightweight libraries that are optimised to run on mobile devices [11], [25]. Yanai et. al. [26] implement multi-threaded mobile applications by utilising either NEON SIMD [27] instructions or the BLAS Library [28] for effective computation of convolution layers.

The other class of techniques focuses on neural network splitting [29]. In this technique, the neural network is divided between the mobile device and a cloud server. In tasks such as image recognition, there is no straightforward implementation. However, several strategies have been studied for optimal splitting of CNNs on edge devices [13], [30]–[33]. Other techniques have studied the performance of the above approaches on cloud and mobile devices separately [34]. Eshratifar et al. [35] introduce a splitting algorithm for reducing the size of the feature map before sending it to the server. The primary limitation of these techniques is that they depend on the architecture of the neural network and cannot be generalised for smartphones.

III. Latency and Energy Model

In order to incorporate latency and energy consumption as objective functions in the optimisation problem, we first need to design a latency and an energy model. The latency model defines how much time is consumed performing different tasks in the splitting process. The energy model defines how much energy the smartphone consumes when splitting a CNN model between the smartphone and the cloud server. We first perform a pilot study to observe how latency and energy consumption vary when different CNN models are split between smartphones and the cloud server.

A. Pilot Study

We set up the experiment between two smartphones and a Windows 10 cloud server. The first smartphone is a Samsung Galaxy J6 and has 4 GB RAM, 64 GB internal storage, 3000 mAh battery. It runs on an Exynos 7 Octa 7870 processor. The second smartphone is a Redmi Note 8, has 4 GB RAM, 32 GB internal storage, 4000 mAh battery. It runs on a Qualcomm SDM665 Snapdragon 665 Octacore processor. Both devices run Android version 10 as the operating system. The cloud server runs a Windows 10 operating system with 8 GB RAM and a 1.6 GHz quad-core Intel i5 processor. All three devices are connected to one Wi-Fi network with bandwidth of 10 Mbps. We run our experiments with four pre-trained CNN models, Alexnet (21 layers) [16], VGG11 (29 layers), VGG13 (33 layers), and VGG16 (39 layers) [17].

We design an Android application using PyTorch for Android [15] to run the CNN models on the smartphone. PyTorch for Android has the advantage that it does not prune or automatically quantise the CNN model, unlike other optimised deep learning libraries. PyTorch for Android ensures that there is no compromise on accuracy. The application stores an instance of a pre-trained model split at a certain layer (called the *split index*) on the smartphone and loads it on execution. The application takes a set of images and the CNN model as input and performs an image classification task. As the experiment is run for all possible split indices for all models, we uploaded a different model instance for each run.

1) Latency Analysis: There are four contributing tasks involved in the end-to-end process of running a split CNN between the smartphone and the cloud server. These tasks are local computation of the CNN model on the smartphone (*Client Latency*), sending the intermediate results to the cloud server (*Upload Latency*), local computation of the remaining CNN layers at the cloud server (*Cloud Server Latency*), and retrieving the final results from the cloud server (*Download*

Latency). The Android Application logs the timestamp when each task starts and the corresponding latency could be computed. However, we observe that the Download Latency is negligible and hence is not included in our results. Therefore, we plot the other three latency values and the corresponding total latency for every possible split index.

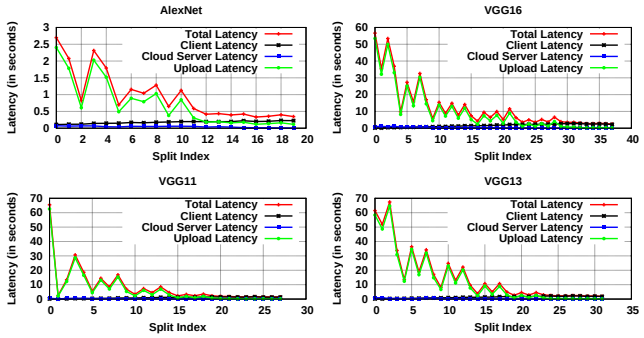


Fig. 1. Latency vs CNN split index for Samsung J6.

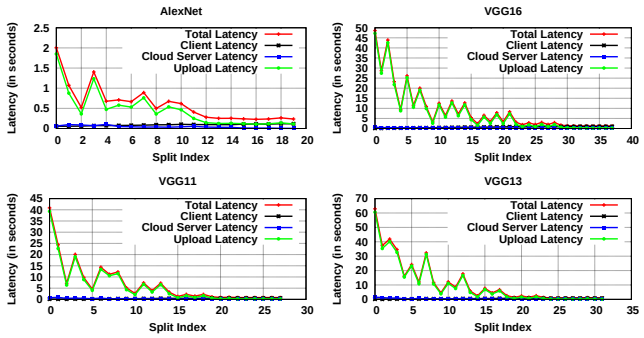


Fig. 2. Latency vs CNN split index for Redmi Note 8.

Figure 1 and Figure 2 show the variation of the different latency values for all four CNN models when run on the two smartphones and split at all possible split indices. Both devices show similar trends for latency, with the Upload Latency being the primary contributing factor to the total latency. The Cloud Server Latency shows low variations because of the high resource availability at the cloud server. As expected, the Client Latency increases as more and more layers are processed on the smartphone.

2) Energy Analysis: The energy consumption calculations are done using the Android Battery Stats tool. For each run of the experiment, we retrieve the bug report file using the Android Debugger. The bug report file contains the voltage (V) and charge (Q) information at regular timestamps. With the voltage and charge values, we compute the energy consumed (E) as follows [36];

$$\mathcal{E} = V * Q \tag{1}$$

We consider the four contributing tasks. The cloud server computation would not contribute to energy consumption on the smartphone. We consider the other three tasks and calculate *Client Energy*, *Upload Energy*, and *Download Energy*. Once the application run is complete, we compute the energy for the three tasks and add all three to get the *Total Energy*. We plot the four energy values for all models when split at all possible split indices.

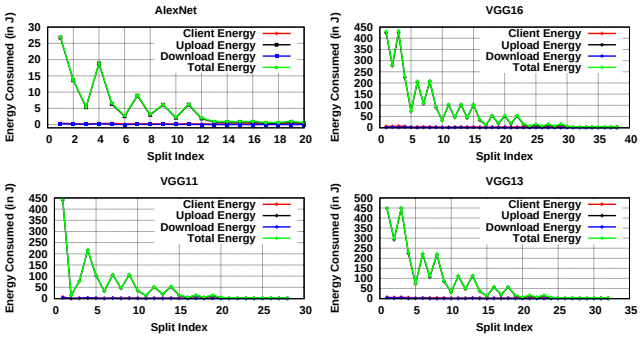


Fig. 3. Energy consumption vs CNN split index for Samsung J6.

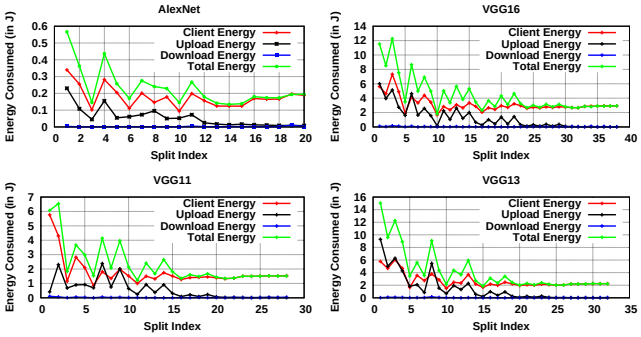


Fig. 4. Energy consumption vs CNN split index for Redmi Note 8.

Figure 3 and Figure 4 show the variation of energy consumption at all split indices for all the four CNN models. We observe that the download energy is very low for all scenarios; however, contrasting results are obtained for the two smartphones for upload and client energy. The upload energy is the primary contributing factor for Samsung J6, while client energy is the primary contributing factor for Redmi Note 8. The contrasting result is due to the different WiFi standards followed by the two devices. Samsung J6 uses the WiFi 802.11 b/g/n standard, which consumes higher energy to perform the upload task. Redmi Note 8 uses the WiFi 802.11 ac standard, which is much more energy-optimised and hence consumes considerably less energy [37], [38]. As can be seen in Figure 5, the client energy consumption remains almost similar for both the devices.

B. Latency Model

As discussed in the pilot study, three latency components contribute to the overall latency of running a CNN between the smartphone and the cloud server. We develop a model to compute the three components used to define the latency objective function.

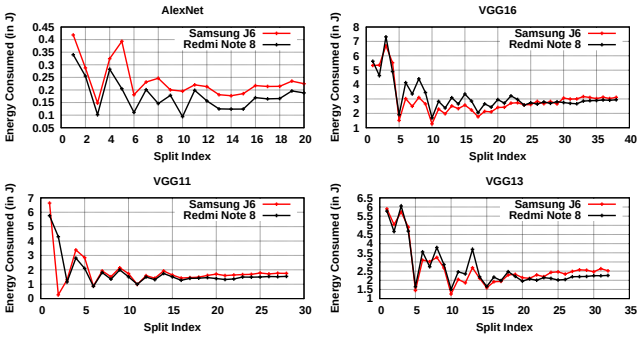


Fig. 5. Client energy consumption for Samsung J6 and Redmi Note 8.

1) Client Latency ($\mathcal{T}_{client}$): Given that l_1 layers are computed on the smartphone, the client latency is given as;

$$\mathcal{T}_{client} = \frac{\mathcal{M}_{client}|l_1}{\mathcal{C}_{client} * \mathcal{S}_{client}} \tag{2}$$

In Equation 2, $\mathcal{M}_{client}|l_1$ is the total memory utilised when l_1 layers of the CNN model are run on the smartphone. The memory utilisation is computed using the tensor size, kernel size, and the type of CNN layer [39]. $\mathcal{C}_{client}$ is the number of cores, and $\mathcal{S}_{client}$ is the processor speed of the smartphone.

2) Cloud Server Latency ($\mathcal{T}_{server}$): Given that l_2 layers are computed at the server we can calculate $\mathcal{T}_{server}$ as;

$$\mathcal{T}_{server} = \frac{\mathcal{M}_{server}|l_2}{\mathcal{C}_{server} * \mathcal{S}_{server}} \tag{3}$$

In Equation 3, $\mathcal{M}_{server}|l_2$ is the total memory utilised when l_2 layers of the CNN model are run on the server, $\mathcal{C}_{server}$ is the number of cores, and $\mathcal{S}_{server}$ is the processor speed of the cloud server.

3) Upload Latency ($\mathcal{T}_{upload}$): The upload latency $\mathcal{T}_{upload}$ is dependent on the intermediate model size and the network bandwidth. It can be computed as;

$$\mathcal{T}_{upload} = \frac{\mathcal{I}|l_1}{\mathcal{B}} \tag{4}$$

In Equation 4, the intermediate model size given that l_1 layers are computed at the smartphone ($\mathcal{I}$) is dependent on the kernel size, stride, padding, and the type of layer [39]. The network bandwidth ($\mathcal{B}$) determines the transmission speed of the intermediate model to the cloud server.

The overall latency is represented as follows:

$$\mathcal{T} = \frac{\mathcal{M}_{client}|l_1}{\mathcal{C}_{client} * \mathcal{S}_{client}} + \frac{\mathcal{I}|l_1}{\mathcal{B}} + \frac{\mathcal{M}_{server}|l_2}{\mathcal{C}_{server} * \mathcal{S}_{server}} \tag{5}$$

C. Energy Model

The total energy consumed depends on three energy components, *Client Energy*, *Upload Energy*, and *Download Energy*, as we have shown earlier in the pilot study. We compute the energy as a product of Power and Time. The time component is available using our latency model. The power consumption at the client and network transmission follow different models that we describe next.

1) Client Energy ($\mathcal{E}_{client}$): The dynamic power consumption when executing an application on a smartphone is proportional to the number of cores ($\mathcal{C}_{client}$) and the operating frequency (ν). The client power is expressed as follows [40];

$$P_{client} = k * \mathcal{C}_{client} * \nu^3 \tag{6}$$

where k is an empirical constant.

Using Equation 6 and 2, we express the Client Energy $\mathcal{E}_{client}$ when computing l_1 layers at the smartphone as follows;

$$\mathcal{E}_{client} = \left(k * \mathcal{C}_{client} * \nu^3\right) * \left(\frac{\mathcal{M}_{client}|l_1}{\mathcal{C}_{client} * \mathcal{S}_{client}}\right) \tag{7}$$

Using the results we obtain in the pilot study, we calculate the value of the constant k as 1.172.

2) Upload Energy ($\mathcal{E}_{upload}$): We use the data transfer energy model given by Huang et al. [41] to compute the power consumption during model upload. Given that the upload throughput of the network is τ_u, the upload power consumption is given as;

$$P_{upload} = \alpha_u * \tau_u + \beta_u \tag{8}$$

where α_u and β_u are empirical constants. Using Equation 8 and 4, we express the Upload Energy $\mathcal{E}_{upload}$ when sending the intermediate model after computing l_1 layers at the smartphone as follows;

$$\mathcal{E}_{upload} = (\alpha_u * \tau_u + \beta_u) * \left(\frac{\mathcal{I}|l_1}{\mathcal{B}}\right) \tag{9}$$

3) Download Energy ($\mathcal{E}_{download}$): Similar to $\mathcal{E}_{upload}$, we use the data transfer energy model given by Huang et al. [41] to compute the power consumption during result download. Given that the download throughput of the network used is τ_d, the download power consumption is given as;

$$P_{download} = \alpha_d * \tau_d + \beta_d \tag{10}$$

where α_d and β_d are empirical constants. We need to define the time consumed when retrieving the result from the cloud server. We can express the download time as follows;

$$\mathcal{T}_{download} = d/\mathcal{B} \tag{11}$$

where d is the download size. Using Equation 10 and 11, we express the Download Energy $\mathcal{E}_{download}$ when retrieving the result from the server as follows;

$$\mathcal{E}_{download} = (\alpha_d * \tau_d + \beta_d) * \left(\frac{d}{\mathcal{B}}\right) \tag{12}$$

With the three energy components defined, we model the overall energy consumption as follows;

$$\begin{aligned}\mathcal{E} = &\left(k * \mathcal{C}_{client} * \nu^3\right) * \left(\frac{\mathcal{M}_{client}|l_1}{\mathcal{C}_{client} * \mathcal{S}_{client}}\right) \\ &+ (\alpha_u * \tau_u + \beta_u) * \left(\frac{\mathcal{I}|l_1}{\mathcal{B}}\right) \\ &+ (\alpha_d * \tau_d + \beta_d) * \left(\frac{d}{\mathcal{B}}\right)\end{aligned} \tag{13}$$

We use the values of the constants as defined by Huang et al. [41]. These being;

$$\alpha_u = 283.17 \text{ mW/Mbps}$$
$$\alpha_d = 137.01 \text{ mW/Mbps}$$
$$\beta_u = \beta_d = 132.86 \text{ mW}$$

With the latency and energy model in place, we now define the optimisation problem.

IV. Problem Definition

Recall that, pilot study shows that the latency and energy consumption varies with the layer at which the CNN model is split. Furthermore, we also observe that the variation in both latency and energy consumption is not monotonously increasing with split index. Moreover, since the smartphone memory is used for other applications running simultaneously, it is crucial that we also minimise the memory utilised by the application.

We define the three objective functions as follows;

$$f_1(l_1, l_2) = \frac{\mathcal{M}_{client}|l_1}{\mathcal{C}_{client} * \mathcal{S}_{client}} + \frac{\mathcal{I}|l_1}{\mathcal{B}} + \frac{\mathcal{M}_{server}|l_2}{\mathcal{C}_{server} * \mathcal{S}_{server}} \quad (14)$$

$$f_2(l_1) = \left(k * \mathcal{C}_{client} * \nu^3\right) * \left(\frac{\mathcal{M}_{client}|l_1}{\mathcal{C}_{client} * \mathcal{S}_{client}}\right) + (\alpha_u * \tau_u + \beta_u) * \left(\frac{\mathcal{I}|l_1}{\mathcal{B}}\right) + (\alpha_d * \tau_d + \beta_d) * \left(\frac{d}{\mathcal{B}}\right) \quad (15)$$

$$f_3(l_1) = \mathcal{M}_{client}|l_1 \quad (16)$$

Equation 14 and Equation 15 are directly derived from the Latency and Energy model respectively (Section III), for variables l_1 and l_2, which are the number of CNN layers at the smartphone and on the cloud server. Equation 16 represents the memory requirement when l_1 layers are computed at the smartphone. Based on these three equations, we formulate the optimisation problem as follows,

$$\begin{aligned} \min \quad & \mathbb{F} = (f_1, f_2, f_3) \\ \text{s.t.} \quad & \mathcal{M}_{edge}|l_1 \leq \mathbb{M}, \\ & l_1 + l_2 = \mathbb{L}, \\ & 1 \leq l_1 \leq \mathbb{L}, \\ & 1 \leq l_2 \leq \mathbb{L}, \\ & \tau_u \leq \mathcal{B}, \\ & \tau_d \leq \mathcal{B} \end{aligned} \quad (17)$$

Equation 17 gives the formulated optimisation problem that minimises f_1, f_2, and f_3 under the six constraints. The first constraint restricts the memory requirement from exceeding the total available memory at the smartphone ($\mathbb{M}$). The second constraint ensures that the number of layers on the smartphone and the cloud server add up to the total number of layers in the model ($\mathbb{L}$). The third and fourth constraints ensure at least one layer is computed at the smartphone and the cloud server. Finally, the last two constraints guarantee that the upload and download throughput is never above the available network bandwidth.

V. The *SmartSplit* Optimisation Algorithm

In this section, we describe the *SmartSplit* algorithm that utilises a genetic algorithm along with decision analysis to solve the multi-objective optimisation problem formulated in the previous section. The results obtained from the pilot study (Section III) show that an increasing trend in energy does not necessarily imply a strictly increasing trend in latency. Hence, a truly optimal solution cannot be obtained for the optimisation problem $\mathbb{F}$. Instead, Pareto-optimal [42] results need to be calculated to obtain the Pareto set.

Algorithm 1: The *SmartSplit* Optimisation Algorithm

Result: The Pareto Optimal Solution for $\mathbb{F}$

```
// NSGA-II to get the Pareto set of
   all Pareto optimal solutions
```

$\mathbb{O} \leftarrow NSGA2(\mathbb{F})$ `// The Pareto Set`

```
// TOPSIS to get Pareto optimal
   solution from 𝕆
```

$\mathcal{F} \leftarrow \left[(f_1)_i \quad (f_2)_i \quad (f_3)_i\right]_{3*n}$

$\mathcal{F}' \leftarrow Column - Normalised(\mathcal{F})$

$\mathcal{F}'' \leftarrow \left[(f'_1)_i \quad (f'_2)_i \quad (f'_3)_i\right]_{3*m}$ `// The reduced matrix of size 3*m where the solutions match the constraints`

$(f'_j)_{ideal} \leftarrow min_{i=1}^{m}(f'_j)_i \; where \; j \in [1, 3]$

$\mathcal{P} \leftarrow [\sqrt{\sum_{j=1}^{3}((f'_j)_i - (f'_j)_{ideal})^2}]_{1*m}$

$\mathbb{O}_{optimal} = \mathcal{F}_i$ corresponding to $min(\mathcal{P})$

Algorithm 1 shows the *SmartSplit* optimisation algorithm that has two parts. NSGA-II [43] to get the Pareto set ($\mathbb{O}$), and TOPSIS [44] to get the best Pareto optimal solution ($\mathbb{O}_{optimal}$). The two parts of the algorithm are described next in the subsequent subsections.

A. Pareto Set Computation using NSGA-II

The Non-dominated Sorting Genetic Algorithm II [43] has been designed to overcome the limitations of NSGA [45] and other state-of-the-art evolutionary strategies such as PAES [46] and SPEA [47]. Employing elitism [48], a faster sorting algorithm, and ensuring diversity of Pareto-optimal solutions using crowding distance improves the convergence property of the algorithm. Furthermore, the NSGA-II provides solutions much closer to the Pareto front than other multi-objective optimisation approaches such as ϵ-constrained optimisation [49], weighted sum [50], or weighted metric methods [51]. We opt for NSGA-II to obtain the Pareto set in *SmartSplit*.

NSGA-II follows the standard genetic algorithm approach during execution; however, it modifies the mating and survival selection strategies. The individuals are first selected per the Pareto front, which leads to the fronts being split as not all individuals are able to survive [52]. In these fronts that are

created due to splitting, the solutions are selected based on *crowding distance*. The crowding distance is the Manhattan distance in the objective space and is computed based on the density of solutions around each solution. NSGA-II also follows a binary tournament mating selection. Hence, every individual is compared based on their rank, followed by the crowding distance. Once all the Pareto optimal solutions are obtained to create a Pareto set ($\mathbb{O}$), we employ TOPSIS to get the best out of these solutions.

B. Optimal Solution Estimation using TOPSIS

In order to obtain the best solution from the Pareto set, we employ decision analysis based on TOPSIS [44]. Say the Pareto set ($\mathbb{O}$), of size n, contains the Pareto optimal solutions $\{\mathbb{F}_1, \mathbb{F}_2, ..., \mathbb{F}_n\}$. Then for every i^{th} solution, the objective functions could be represented as; $\mathbb{F}_i = [(f_1)_i, (f_2)_i, (f_3)_i]$. For TOPSIS, we obtain the decision matrix for all solution in the pareto set as;

$$\mathcal{F} = \begin{bmatrix} (f_1)_1 & (f_2)_1 & (f_3)_1 \\ (f_1)_2 & (f_2)_2 & (f_3)_2 \\ . & . & . \\ . & . & . \\ (f_1)_n & (f_2)_n & (f_3)_n \end{bmatrix}$$

The matrix elements in $\mathcal{F}$ are then column-wise normalised to obtain $\mathcal{F}'$. Next, we remove all the solutions that do not meet the constraints in $\mathbb{F}$ to get $\mathcal{F}''$ which has a reduced dimension of $3 * m$. Now, for each objective function, the ideal value is calculated as;

$$(f'_j)_{ideal} \leftarrow min_{i=1}^{m}(f'_j)_i$$

where $j \in (1, 3)$

The Euclidean distance between all elements of $\mathcal{F}$ and the ideal points is calculated to get the distance matrix as;

$$\mathcal{P} = \begin{bmatrix} \sqrt{\sum_{j=1}^{3}((f'_j)_1 - (f'_j)_{ideal})^2} \\ \sqrt{\sum_{j=1}^{3}((f'_j)_2 - (f'_j)_{ideal})^2} \\ . \\ . \\ . \\ \sqrt{\sum_{j=1}^{3}((f'_j)_m - (f'_j)_{ideal})^2} \end{bmatrix}$$

The solution corresponding to the minimum element in $\mathcal{F}$ is then selected as the Pareto optimal solution ($\mathbb{O}_{optimal}$) of the multi-objective optimisation problem $\mathbb{F}$.

VI. Performance Evaluation

In this section, we first describe the experimental setup followed by an evaluation of the *SmartSplit* algorithm. We then provide the comparison of *SmartSplit* with other state-of-the-art approaches.

A. Experimental Setup

We run our experiments on Samsung Galaxy J6 with 4 GB RAM, 64 GB internal storage, 3000 mAh battery, and an Exynos 7 Octa 7870 processor with an Android 10 operating system. The cloud server is a Windows 10 system with 8 GB RAM and a 1.6 GHz quad-core Intel i5 processor. We utilise a Wi-Fi connection with 10 Mbps bandwidth to connect both devices. For our experiments, we use five pre-trained CNN models: Alexnet (21 layers) [16], VGG11 (29 layers), VGG13 (33 layers), VGG16 (39 layers) [17], and MobileNetV2 [8] (21 layers).

B. Performance Evaluation of SmartSplit

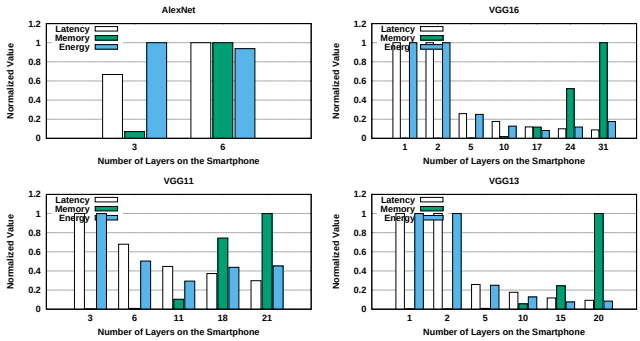


Fig. 6. Values of Latency, Energy consumed, and Memory utilised for all the solutions in the Pareto set obtained using NSGA-II.

	AlexNet	VGG11	VGG13	VGG16
Smartphone Layers	3	11	10	10

TABLE I
THE OPTIMAL NUMBER OF LAYERS FOR THE FOUR MODELS AFTER APPLYING TOPSIS ON THE PARETO SET.

We evaluate the *SmartSplit* algorithm to observe if the objective function values obtained based on the algorithm output are optimal or not. We test *SmartSplit* over AlexNet, VGG11, VGG13, and VGG16. In Figure 6 we plot the normalised latency, energy consumed, and memory utilised values for all solutions in the Pareto set obtained from NSGA-II. The corresponding best optimal solution obtained after performing TOPSIS is given in Table I. We observe that *SmartSplit* gives the optimal solution in all four cases. For example, when split at layer 3 in AlexNet, latency and memory utilisation are lowest, and the energy consumption is comparable. Similarly, for VGG11, when split at layer 11, memory utilisation and energy consumption are lower than other solutions in the Pareto set. Moreover, the latency is also comparable to the lowest value. In the case of VGG13 and VGG16, splitting at layer ten is close to the next best solutions, i.e., layer 15 for VGG13 and layer 17 for VGG16. However, in both CNN models, the memory utilisation is very low when splitting at layer 10, while the other two parameters are comparable to the next best solutions.

C. Comparison with Competing Algorithms

We compare *SmartSplit* with five competing algorithms that split the CNN between the smartphone and the cloud server based on other strategies.

1) Latency-based Optimisation (LBO): This class of optimisation algorithms only consider the best layer to split the CNN, such that the latency is minimum.

2) Energy-based Optimisation (EBO): Since energy-based optimisation algorithms are not available in the literature for smartphones; we design our own algorithm. In EBO, the optimal layer to split the CNN is the one that minimises the total energy consumed by running the CNN application on the smartphone. Effectively, EBO only has to solve f_2 of $\mathbb{F}$.

3) CNN on Smartphone (COS): This approach runs the complete CNN model on the smartphone.

4) CNN on Cloud Server (COC): This approach runs the complete CNN model on the server.

5) Random Splitting (RS): A random number is selected for each run, and the CNN is split at that random layer.

Algorithm	AlexNet	VGG11	VGG13	VGG16
SmartSplit	3	11	10	10
LBO	3	21	20	25
EBO	6	11	15	17
COS	21	29	33	39

TABLE II
NUMBER OF LAYERS AT SMARTPHONE FOR COMPETING ALGORITHMS. COC HAS ZERO LAYERS AT THE SMARTPHONE AND RS GENERATES RANDOM LAYERS.

We evaluate all six competing algorithms over the four CNN models and compare the corresponding latency, energy consumption and memory utilisation values. in Table II, we provide the number of layers computed on the smartphone when using all the algorithms except RS and COC. COC has all the layers computed at the cloud server, and RS always gives a random value. We perform 100 runs on Samsung Galaxy J6, and the average values are reported.

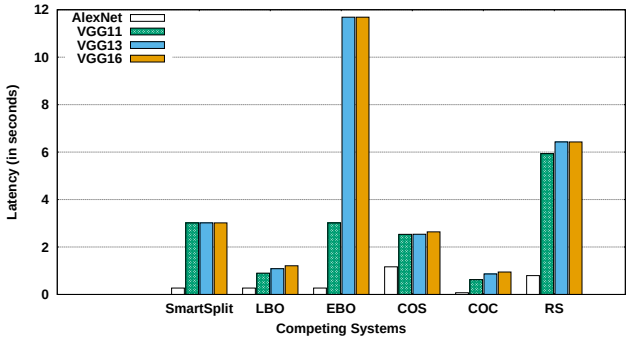


Fig. 7. **Comparing latency values achieved when using different competing algorithms**

In Figure 7, Figure 8, and Figure 9, we plot the latency, energy consumption, and memory utilisation values for all competing algorithm respectively. Intuitively, the results obtained with RS is non-consistent owing to the random nature

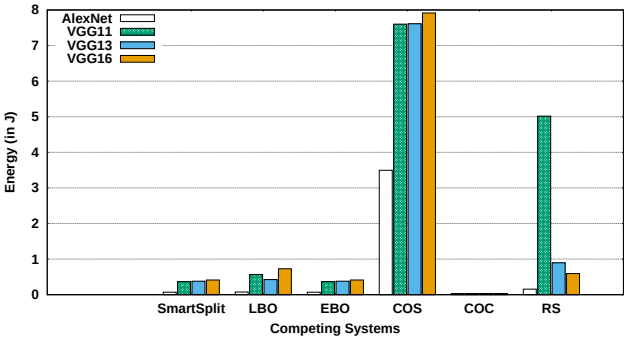


Fig. 8. Comparing energy consumption values achieved when using different competing algorithms

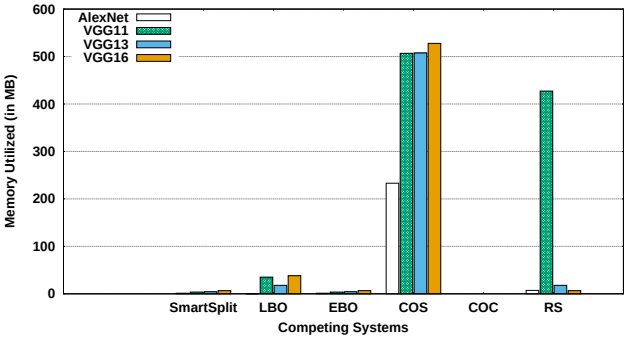


Fig. 9. Comparing memory utilisation values achieved when using different competing algorithms

of the algorithm. COC, as expected, has the minimum values for latency and energy consumption and negligible memory requirement. However, it defeats the purpose of running AI on a smartphone environment. On the other hand, COS has the highest energy and memory consumption than the other competing algorithms because the smartphone has to perform all computation by itself. The COS algorithm's latency is comparatively low due to the absence of any interaction with the cloud server. Although the EBO algorithm has low energy consumption as expected and low memory consumption, it has high latency requirements. Moreover, the energy consumption values are also comparable to *SmartSplit* and LBO. LBO shows the closest results when compared to *SmartSplit*. However, *SmartSplit* has lower energy consumption and memory requirement than LBO and comparable latency values. The comparison results show that *SmartSplit* gives the best trade-off between latency, energy consumption and memory utilisation as compared to all the competing algorithms, thus making it a better alternative.

D. Comparison with Smartphone Optimised Approach

We compare *SmartSplit* with MobileNetv2 [8] which was explicitly designed to run on smartphones. MobileNetv2 has 21 layers and decreases the total number of model parameters using depth-wise separable convolutions [53], thus reducing the computational complexity and memory requirements. We

compare MobileNetv2 with the four CNN models and the COS algorithm with VGG16. We run our experiments for 100 input images, and the reported results are averaged over 100 runs.

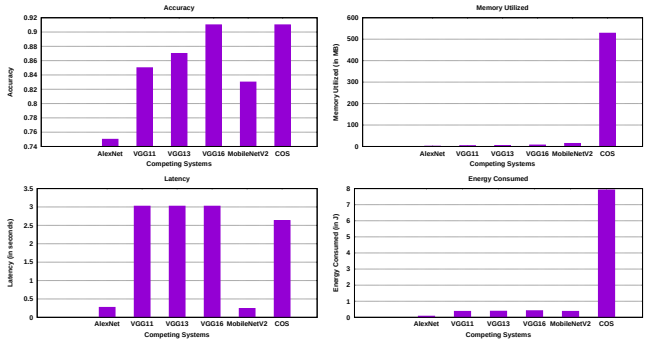


Fig. 10. Comparison of *SmartSplit* with MobileNetV2 and COS

In Figure 10, we plot accuracy, latency, energy consumed and memory utilised values for the four CNN models that are split using *SmartSplit*, MobileNetv2, and COS. MobileNetV2 is a compressed model for smartphones and has almost 10% less accuracy than VGG16 with *SmartSplit*. Furthermore, unlike COS and MobileNetV2, since not all layers of the CNN model is run on the smartphone, the memory utilised by the CNN models after splitting is lower than COS and MobileNetV2. Moreover, the energy consumed is similar to MobileNetV2 for the VGG models with split and lower for AlexNet with *SmartSplit*. MobileNetV2 does have lower latency. However, a latency difference of ≈ 2.7 seconds is a small trade-off for improving accuracy by 10% and a lower memory utilisation. We conclude from Figure 10 that splitting is a superior alternative for running CNNs on a smartphone environment.

VII. Conclusion

Smartphones are personal devices, and hence bringing AI-based applications to the smartphone has become crucial in recent times. Although the smartphone industry has been performing extensive research towards developing better hardware to run such applications, the earlier generation of smartphones without such hardware would miss out on the developments in the field. In this paper, we develop *SmartSplit* that makes resource-intensive AI-based applications accessible for most smartphones and ensures that resource utilisation is minimised, thus also favouring new smartphones with better hardware. We assert that running a CNN on a smartphone could be done more efficiently by splitting CNNs between the smartphone and a cloud server. We formulate this problem as a multi-objective optimisation problem that minimises the smartphone's end-to-end latency, memory utilisation, and energy consumed. We design a latency and energy model utilised to define the objective functions of the optimisation problem. In order to solve the multi-objective optimisation problem, we design *SmartSplit*, that uses NSGA-II to obtain a set of Pareto optimal solutions and then utilise TOPSIS to estimate the best Pareto optimal solution. Experiments on non-AI optimised smartphones show that *SmartSplit* gives optimal solution and performs better than other state-of-the-art approaches. The key takeaways from the paper are: (i) the upload energy is the primary factor affecting energy consumption in the earlier generation of smartphones, and hence network bandwidth is a crucial parameter to consider when splitting CNNs, (ii) *SmartSplit* is a better alternative compared to other splitting based approaches since it optimises three different objective functions, (iii) as compared to compressed CNN models, that compromise on accuracy, splitting the CNN between the smartphone and cloud server ensures higher accuracy and comparable latency along with low energy and memory consumption.

Some key aspects require further analysis in order to improve *SmartSplit*. First, generalising *SmartSplit* for other types of neural networks such as RNNs, LSTMs, or GANs would make more use-cases accessible on smartphones. Second, performing a splitting of CNNs in a centralised network topology. This direction of work would eliminate the single point of failure in *SmartSplit*. The third direction of work is to investigating how *SmartSplit* could be generalised such that smartphones are combined with other edge devices to create a heterogeneous edge ecosystem performing shared AI tasks such as localisation, smart home management, intruder detection, and several others. This work would involve investigating privacy concerns and fault mitigation.

References

[1] Micron. Artificial intelligence in your smartphone? you'll need faster storage. [Online]. Available: https://in.micron.com/insight/artificial-intelligence-in-your-smartphone

[2] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in *Shape, contour and grouping in computer vision*. Springer, 1999, pp. 319–345.

[3] Q. Kong, R. M. Allen, L. Schreier, and Y.-W. Kwon, "Myshake: A smartphone seismic network for earthquake early warning and beyond," *Science advances*, vol. 2, no. 2, p. e1501055, 2016.

[4] Apple. An on-device deep neural network for face detection. [Online]. Available: https://machinelearning.apple.com/research/face-detection

[5] Counterpoint. The rise of ai: One in three smartphones will be ai capable in 2020. [Online]. Available: https://www.counterpointresearch.com/the-rise-of-ai-one-in-three-smartphones-will-be-

[6] Huawei. Huawei reveals the future of mobile ai at ifa 2017. [Online]. Available: https://consumer.huawei.com/sg/press/news/2017/ifa2017-kirin970/

[7] Apple. iphone 8 and iphone 8 plus: A new generation of iphone. [Online]. Available: https://www.apple.com/newsroom/2017/09/iphone-8-and-iphone-8-plus-a-new-generati

[8] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[9] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[10] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv*, 2018.

[11] TensorFlow. Tensorflow lite. [Online]. Available: https://www.tensorflow.org/lite

[12] R. A. Khalil, N. Saeed, M. Masood, Y. M. Fard, M.-S. Alouini, and T. Y. Al-Naffouri, "Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications," *IEEE Internet of Things Journal*, 2021.

[13] R. Mehta and R. Shorey, "Deepsplit: Dynamic splitting of collaborative edge-cloud convolutional neural networks," in *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 2020, pp. 720–725.

[14] X. Tang, X. Chen, L. Zeng, S. Yu, and L. Chen, "Joint multi-user dnn partitioning and computational resource allocation for collaborative edge intelligence," *IEEE Internet of Things Journal*, 2020.

[15] PyTorch. Pytorch for android. [Online]. Available: https://pytorch.org/mobile/android/

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[18] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[19] Y. Huang, X. Ma, X. Fan, J. Liu, and W. Gong, "When deep learning meets edge computing," in *2017 IEEE 25th international conference on network protocols (ICNP)*. IEEE, 2017, pp. 1–2.

[20] A. Ignatov, R. Timofte, W. Chou, K. Wang, M. Wu, T. Hartley, and L. Van Gool, "Ai benchmark: Running deep neural networks on android smartphones," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 0–0.

[21] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding," *arXiv preprint arXiv:1510.00149*, 2015.

[22] S. Teerapittayanon, B. McDanel, and H.-T. Kung, "Branchynet: Fast inference via early exiting from deep neural networks," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 2464–2469.

[23] C. Gamanayake, L. Jayasinghe, B. K. K. Ng, and C. Yuen, "Cluster pruning: An efficient filter pruning method for edge ai vision applications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 4, pp. 802–816, 2020.

[24] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and Q. V. Le, "Mnasnet: Platform-aware neural architecture search for mobile," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2820–2828.

[25] Y. Matsubara, S. Baidya, D. Callegaro, M. Levorato, and S. Singh, "Distilled split deep neural networks for edge-assisted real-time systems," in *Proceedings of the 2019 Workshop on Hot Topics in Video Analytics and Intelligent Edges*, 2019, pp. 21–26.

[26] K. Yanai, R. Tanno, and K. Okamoto, "Efficient mobile implementation of a cnn-based object recognition system," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 362–366.

[27] ARM. Simd neon. [Online]. Available: https://developer.arm.com/architectures/instruction-sets/simd-isas/neon

[28] C. Nugteren, "Clblast: A tuned opencl blas library," in *Proceedings of the International Workshop on OpenCL*, 2018, pp. 1–10.

[29] S. S. Latifi Oskouei, H. Golestani, M. Hashemi, and S. Ghiasi, "Cndroid: Gpu-accelerated execution of trained deep convolutional neural networks on android," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 1201–1205.

[30] Y. Zhang, D. Zhao, J. Sun, G. Zou, and W. Li, "Adaptive convolutional neural network and its application in face recognition," *Neural Processing Letters*, vol. 43, no. 2, pp. 389–399, 2016.

[31] E. MalekHosseini, M. Hajabdollahi, N. Karimi, S. Samavi, and S. Shirani, "Splitting convolutional neural network structures for efficient inference," *arXiv preprint arXiv:2002.03302*, 2020.

[32] T. Jin and S. Hong, "Split-cnn: Splitting window-based operations in convolutional neural networks for memory system optimization," in *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, 2019, pp. 835–847.

[33] P. A. Leroy and T. Goedemé, "Optimal distribution of cnn computations on edge and cloud." in *ICAART (2)*, 2021, pp. 604–613.

[34] T. Guo, "Cloud-based or on-device: An empirical study of mobile deep inference," in *2018 IEEE International Conference on Cloud Engineering (IC2E)*. IEEE, 2018, pp. 184–190.

[35] A. E. Eshratifar, A. Esmaili, and M. Pedram, "Bottlenet: A deep learning architecture for intelligent mobile cloud computing services," in *2019 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*. IEEE, 2019, pp. 1–6.

[36] W. Oliveira, R. Oliveira, and F. Castor, "A study on the energy consumption of android app development approaches," in *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*. IEEE, 2017, pp. 42–52.

[37] L. Sun, H. Deng, R. K. Sheshadri, W. Zheng, and D. Koutsonikolas, "Experimental evaluation of wifi active power/energy consumption models for smartphones," *IEEE Transactions on Mobile Computing*, vol. 16, no. 1, pp. 115–129, 2016.

[38] G. P. Noordbruis, B. van Gastel, and M. van Eekelen, "Energy efficient wlan using wifi standards b/a/g/n/ac on an archer c7 ac1750 access point," 2020.

[39] S. N. Satya Mallick. Number of parameters and tensor sizes in a convolutional neural network (cnn). [Online]. Available: https://learnopencv.com/number-of-parameters-and-tensor-sizes-in-convolutional-neura

[40] V. Sundriyal and M. Sosonkina, "Modeling of the cpu frequency to minimize energy consumption in parallel applications," *Sustainable Computing: Informatics and Systems*, vol. 17, pp. 1–8, 2018.

[41] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4g lte networks," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*, 2012, pp. 225–238.

[42] G. Debreu, "Valuation equilibrium and pareto optimum," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 40, no. 7, p. 588, 1954.

[43] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[44] M. Behzadian, S. K. Otaghsara, M. Yazdani, and J. Ignatius, "A state-of the-art survey of topsis applications," *Expert Systems with applications*, vol. 39, no. 17, pp. 13 051–13 069, 2012.

[45] N. Srinivasan and K. Deb, "Multi-objective function optimisation using non-dominated sorting genetic algorithm," *Evolutionary Comp*, vol. 2, no. 3, pp. 221–248, 1994.

[46] J. Knowles and D. Corne, "The pareto archived evolution strategy: A new baseline algorithm for pareto multiobjective optimisation," in *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, vol. 1. IEEE, 1999, pp. 98–105.

[47] E. Zitzler and L. Thiele, "An evolutionary algorithm for multiobjective optimization: The strength pareto approach," *TIK-report*, vol. 43, 1998.

[48] H. Du, Z. Wang, W. Zhan, and J. Guo, "Elitism and distance strategy for selection of evolutionary algorithms," *IEEE Access*, vol. 6, pp. 44 531–44 541, 2018.

[49] V. Chankong and Y. Y. Haimes, *Multiobjective decision making: theory and methodology*. Courier Dover Publications, 2008.

[50] R. T. Marler and J. S. Arora, "The weighted sum method for multi-objective optimization: new insights," *Structural and multidisciplinary optimization*, vol. 41, no. 6, pp. 853–862, 2010.

[51] A. R. NOUR and A. M. KAMALI, "A weighted metric method to optimize multi-response robust problems," 2009.

[52] M. A. Albadr, S. Tiun, M. Ayob, and F. AL-Dhief, "Genetic algorithm based on natural selection theory for optimization problems," *Symmetry*, vol. 12, no. 11, p. 1758, 2020.

[53] L. Kaiser, A. N. Gomez, and F. Chollet, "Depthwise separable convolutions for neural machine translation," *arXiv preprint arXiv:1706.03059*, 2017.

Silicon
https://doi.org/10.1007/s12633-021-01472-y

ORIGINAL PAPER

TCAD investigation of ferroelectric based substrate MOSFET for digital application

Rashi Mann [1] · Rishu Chaujar [1]

Received: 26 July 2021 / Accepted: 11 October 2021
© Springer Nature B.V. 2021

Abstract
The present investigation is focused on the analog/RF performance of ferroelectric (FE) based substrate metal oxide semiconductor field effect transistor (MOSFET) for digital application as an inverter. Short channel effect (SCEs) reduced significantly with improvement in switching ratio (I_{on}/I_{off}). The simulation shows the immunity towards the parameters like transconductance (g_m), transconductance generation factor (TGF), early voltage (V_{EA}) and the intrinsic gain (A_V) which allows the device to look into digital application. Also, the RF performance parameters come out to be superior for the proposed device structure, as the gain frequency product (GFP) and gain bandwidth product (GBP) enhanced by 100 and 10 folds, respectively. Further, RF investigation shows the enhanced results as attributed with peak values for transconductance frequency product (TFP) and gain transconductance frequency product (GTFP). The transient analysis shows an improvement in the noise margin (NM) for FE-MOSFET structure. Thus, the proposed architecture can be an attractive alternative for the digital application.

Keywords Ferroelectric material · Subthreshold swing · analog/RF parameters · Transient analysis · Drain induced barrier lowering

1 Introduction

As the situation of current increasing demand to follow the high performance and ultra-low power circuits has motivated the semiconductor industry to downscale the transistors which also states by Moore's law [1]. But with the following of this trend, CMOS technology is getting towards its fundamental limits of scaling down. As the size of the transistor scales down that means the length of the channel in the device decreases, the SCEs like channel tunneling, drain induced barrier lowering, subthreshold swing value start dominating and affect the performance of the device [2–4]. With scaling below 100 nm, the control of the gate over the device channel becomes poor, due to which the SCEs start to dominate. Other issues from which the device performance suffers is pin-point ultra-sharp S/D junction's formation. Some of these effects are reduced by junctionless field effect transistors in which doping species in all regions are the same [5]. Also gives better protection against SCEs in comparison to conventional CMOS devices.

Another challenge with the decreasing size of the transistor comes from power dissipation. Now, with reduction of threshold voltage for same device speed and increasing current exponentially, the static power dissipation is new objective sets for transistors. Whether it looks very difficult to achieve both things with balance but there is a possible way to maintain the stability between device performance and power dissipation which is the subthreshold swing value of a device [6]. It represents the amount of applied gate voltage that is required to change the current by one decade and given by

$$SS = \frac{k_B T}{q} \frac{d\log_{10} I_D}{dV_{gs}} \quad (1)$$

For a conventional MOSFET, the value of SS is ≥ 60mV/decade. From the formula (1), we can say that as the value of SS lowers, this means the needed applied gate voltage becomes less to switch on the transistor [7, 8].

Furthermore, the next challenge is to increase the gate direct tunneling by reducing gate oxide thickness which gives the decrement in leakage current [9, 10]. The technical

✉ Rishu Chaujar
chaujar.rishu@dtu.ac.in

Rashi Mann
rashimann08@gmail.com

[1] Microelectronics Research Lab, Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110042, India

Published online: 28 October 2021

solution for this situation is proved by the use of high-k dielectric materials like hafnium oxide (HfO_2) (k=25). This can be used in the gate stack of the transistor with a combination of silicon oxide (SiO_2) as the effective oxide thickness (EOT) stays constant [11–14]. As HfO_2 cannot be directly deposited on the Si substrate because it forms the oxide composites at the interface of the gate stack and substrate region which offers the resistance at the interface. So, with maintaining the interface quality, HfO_2 always deposited on the top of SiO_2 in the gate stack of the transistor.

In the near past, the use of ferroelectric material in the gate stack of MOSFET gives better immunity to the device performance because FE material shows the behavior of negative capacitance which can be observed from their charge energy curve [15, 16]. So we can use the ferroelectric materials in many possible ways to redefine the architecture of existing MOSFET structures. For that, a suitable example of FE materials is HfO_2 based FE (HfO_2FE) [17]. Basically, we know that the HfO_2 is high-k dielectric material but when it deposited with chemical vapor deposition, its properties show the ferroelectric behavior [18].

Thus, by considering all the conceptions, we designed a new architecture of MOSFET in which we use a ferroelectric material (HfO_2FE) sandwiched in between the high-k dielectric material HfO_2 in the substrate region with the addition of HfO_2 in the gate stack of the device. We named it as ferroelectric based substrate metal oxide semiconductor field effect transistor (FE-MOSFET) as shown in Fig. 1. In the last several years, there is a sufficient increase in demands for higher battery backup and portable electronic devices. And to the matchup, this demand with the addition of speed performance, the concept of Moore's law has been followed for decades. And with help of this system-on-chip (SoC) concept, we can integrate several independent systems on a single chip. The main challenge in the SoC technology is device optimization. Accordingly, the persistence of this work is to consider the structure and performance of the device for the digital application. Analog/RF performance parameters for FE-MOSFET has been examined broadly. In this work, we study the FE-MOSFET with its analog/RF metrics that are significant for RF applications such as GBP, GFP, TFP and GTFP. It is examined from the simulation that modified FE-MOSFET shows the better results in comparison to the conventional MOSFET structure. FE-MOSFET gives the higher switching ratio, lesser leakage current, better subthreshold swing value etc. with the reduction of the threshold voltage value. Starting from the basic introduction of FETs in Section 1, Section 2 briefs the device structure. The simulation methodology is explained in Section 3, Section 4 deals with the analog/RF performance parameter comparison of these two devices. Further, the digital application of FE-MOSFET is discussed and compared with conventional MOSFET in Section 5. Lastly, the concluding remarks are added in Section 6.

2 Device structure

Fig. 1 illustrates the device structure of conventional MOSFET and ferroelectric based substrate MOSFET (FE-MOSFET). The gate length (L_g) for both device structure is 25 nm and the length of source and drain region (S/D) extensions are 10 nm on both side of the gate. The silicon oxide thickness is fixed to 0.5 nm and also the thickness of high-k dielectric material HfO_2 is kept at 0.5 nm in the gate stack of the proposed device structure. In the conventional MOSFET device structure, SiO_2 with a thickness of 1 nm is used as the gate oxide.

Also, silicon material is used as the substrate region for both the device. The modification is done in the substrate region of the proposed device structure or FE-MOSFET structure. A layer of FE material is sandwiched in between the layers of HfO_2. The silicon substrate region in both the device is doped with p-type species of 1×10^{16} cm^{-3}. Also, the doping concentration for the source/drain region is 1×10^{18} cm^{-3} and of n-type. The material used for the gate terminal is Npoly silicon has a work function (φ_m) at 4.5 eV. The Landau's parameters for the FE material are a= - 1e+07 and β = 8.9e+08 [18].These all default dimensions are tabulated in Table 1

3 Simulation methodology

The VISUAL TCAD simulator has been used to simulate both device configurations. In the complete simulation results, the drain-source voltage (V_{ds}) is set at 0.5 V and gate source voltage (V_{gs}) is varying from 0 to 1 V. Temperature is fixed at 300 K for all the simulation work. Also, many physical models are used to explain different concepts in the simulation work like recombination effects, Shockley-Read-Hall (SRH) model briefs the trap charges existence at the interface. Arora model defined the concentration dependent mobility. The Fermi-Dirac statistics are used for accuracy in the results. The calibration on visual TCAD simulator is done for the existing results of 35nm MOSFET as in published work of G. Roy et al. That's why we used the dimension of 35nm only for the calibration part. The calibrated results of 35 nm MOSFET are presented in Fig. 2 with constant V_{ds} of 50 mV and 850 mV [19].

4 Results and discussion

4.1 Device Scalability

Fig. 3 shows the variation comparison of drain current (I_D) with respect to the gate source voltage (V_{gs}) at constant drain source voltage (V_{ds}) of 0.5 V in both linear and log scale for

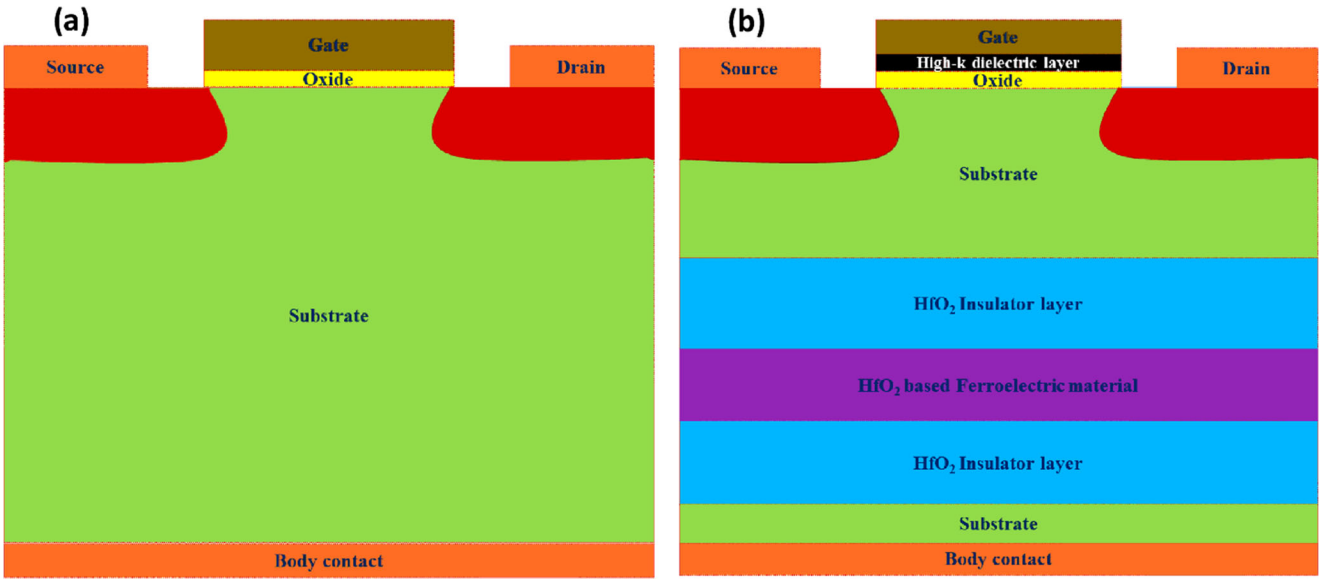


Fig. 1 Systematic architecture of conventional MOSFET (**a**) ferroelectric based substrate MOSFET (**b**)

both the device structures. We can say from the variation comparison that I_D increases exponentially with an increase in the V_{gs} and improvement in I_D is recorded for the modified FE-MOSFET than for conventional MOSFET. As it is found for the modified configuration that the on current (I_{on}) improves with reduction in the off current (I_{off}) or leakage current which gives the higher switching ratio (I_{on}/I_{off}) and enhanced switching speed to the device as shown in Fig. 4. This is attributed due to the reduction in tunneling current and resistance offered by the substrate region modification. Secondly, an important short channel parameter is DIBL which is the second name of channel length modulation. DIBL is responsible for the increase in leakage current at higher constant drain source voltage. So for better device performance, the DIBL effect should be minimized [2–4]. The DIBL value for conventional MOSFET and modified FE-FET is displayed in the same Fig. 4. The threshold voltage (V_{th}) comparison for conventional MOSFET and FE- MOSFET is shown in Fig. 5. V_{th} for a device shows the voltage at which the hole-concentration in the substrate region equate the electron-concentration in the channel region so that the current starts to flow to its maximum value. The threshold voltage for the modified configuration is less than the conventional device structure. This is due to the fact that the ferroelectric material capacitance shows a negative value. This shows that the device can switch from off to on with less value of applied gate source voltage value. The subthreshold swing (SS) value for both the device configuration is displayed in the form of a bar graph for linear and saturation region in Fig. 6. As discussed in Section 1, SS is the important short channel parameter that should be low for better device performance [6]. And from Fig. 6, it is clear that the SS in both linear and saturation region of operation is reduced significantly for FE-MOSFET approaching the ideal value (60mV/dec). It is due to an increase

Table 1 Default device dimension and parameters are taken for the simulation

Parameter	Conventional MOSFET	Modified FE-MOSFET
L_g (nm)	25	25
$L_{S/D}$ (nm)	10	10
T_{ox} (nm)	1	1 (0.5 SiO_2+0.5 HfO_2)
T_{FE} (nm)	0	8
T_{ins} (nm)	0	9
$N_{S/D}$ (cm^{-3})	1×10^{19}	1×10^{19}
N_{ch} (cm^{-3})	1×10^{16}	1×10^{16}
Φ_m (eV)	4.5 (Npoly Si)	4.5 (Npoly Si)

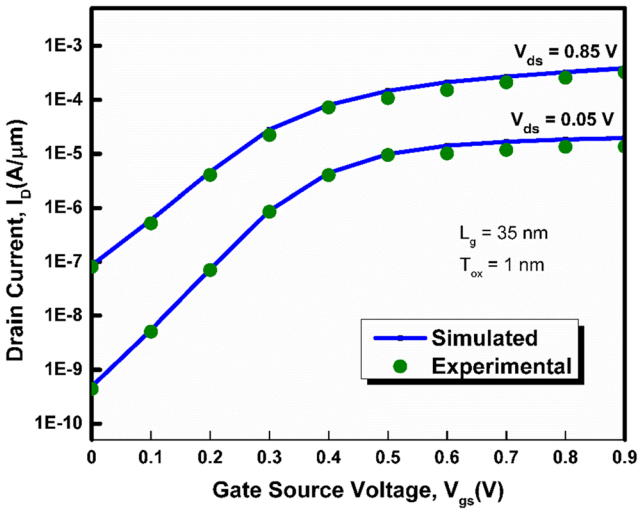


Fig. 2 Calibrated I_D-V_{gs} characteristics of 35nm MOSFET to experimental data

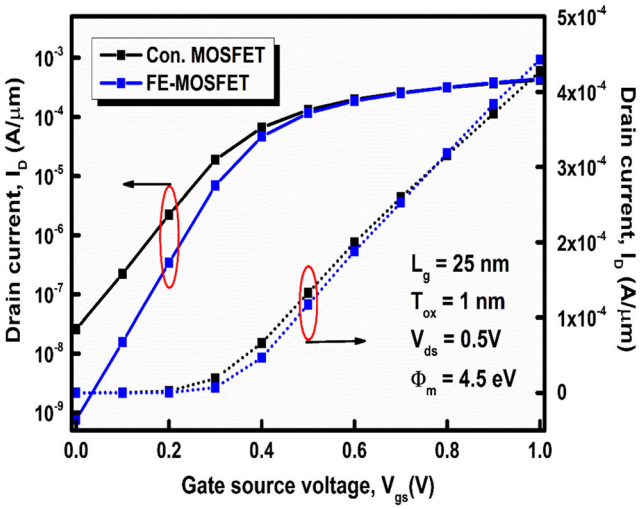


Fig. 3 Variation of I_D with respect to V_{gs} at constant V_{ds} = 0.5 V in logarithmic and linear scale

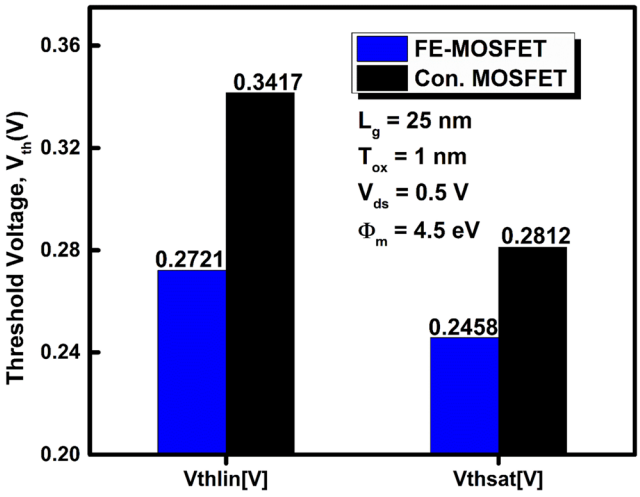


Fig. 5 Threshold voltage comparison for conventional MOSFET and ferroelectric based substrate MOSFET configuration

in gate coupling capacitance with increment in permittivity (k) which improves the gate voltage control on the channel region, thus enhance the SS value.

4.2 Analog metrics

In this part of the paper, from the view of analog performance device application, many important applications like g_m, TGF, V_{EA}, A_V are described with their explanation. Fig. 7 represents the g_m and TGF variation together with respect to the V_{gs} at constant V_{ds} of 0.5 V for both the device structures. Equations (2) and (3) formulates the transconductance and TGF, respectively [20]. From Fig. 7, it is clear that the FE-MOSFET shows the enhanced results with an improved value of g_m and TGF. Transconductance is a very important analog factor for a device as it estimates the overall gain for a device. It can be evaluated from the I_D-V_{gs} characteristics by performing the derivative of I_D with respect to the applied V_{gs} as shown in Eq. (2). Higher value of g_m shows the better control of gate voltage on the channel and hence became the reason of reduced short channel effects [20, 21]. Also, the use of ferroelectric material with a combination of high-k dielectric material enhance the mobile electron concentration and average carrier velocity in the channel region, thus directly increase the g_m. Also, TGF shows the net gain generated per unit power loss for a device as formulated in Eq. (3). A higher value of TGF shows that how efficiently a device can work at low supply voltages. FE-MOSFET configuration reaches the maximum value of TGF. The reason for the higher TGF is the higher drain current and its higher g_m value [22].

$$g_m = \frac{\partial I_D}{\partial V_{gs}} \tag{2}$$

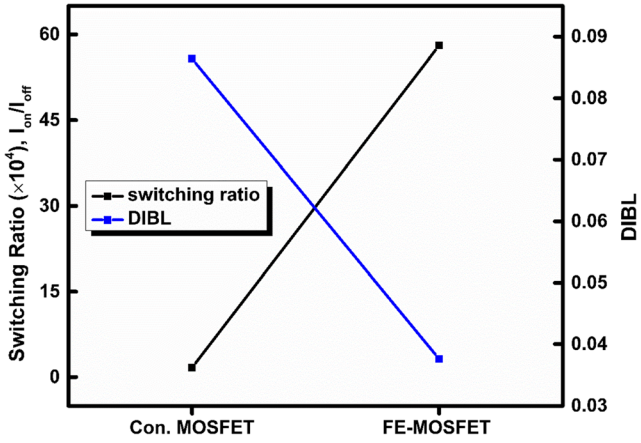


Fig. 4 Comparison of switching ratio and DIBL for conventional MOSFET and ferroelectric based substrate MOSFET structure

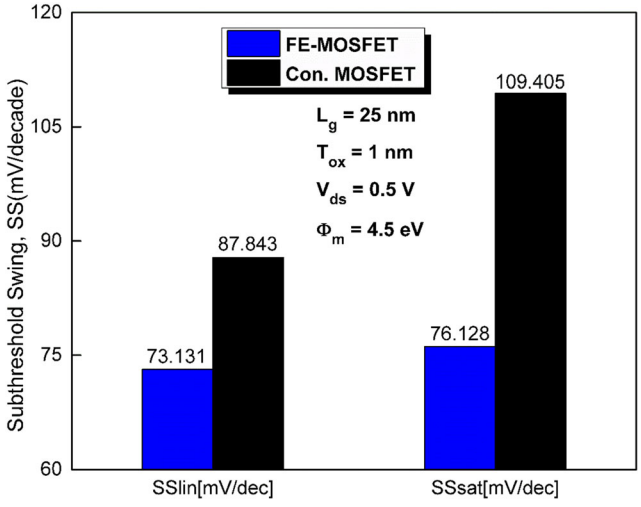


Fig. 6 Subthreshold swing value of conventional MOSFET and FE-MOSFET configuration

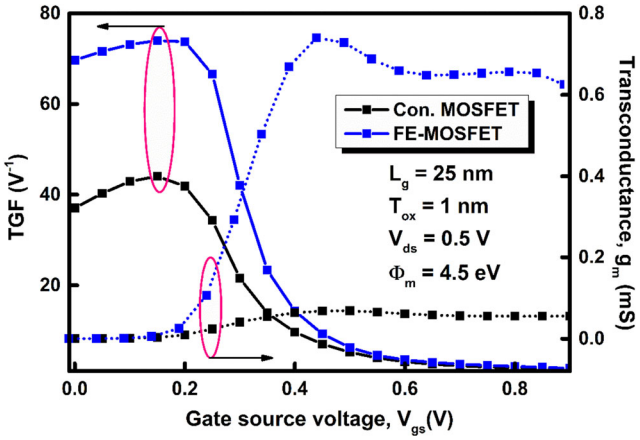


Fig. 7 g_m and TGF variation with V_{gs} for conventional MOSFET and ferroelectric based substrate MOSFET device structure

$$TGF = \frac{g_m}{I_D} \tag{3}$$

$$V_{EA} = \frac{I_D}{g_d} \tag{4}$$

$$A_v = \frac{g_m}{g_d} = \frac{g_m}{I_D} \times V_{EA} \tag{5}$$

The variation of early voltage with the variation of corresponding applied supply voltages is displayed in Fig. 8 for the conventional MOSFET and FE-MOSFET structure. For the improved analog performance of a device configuration, this parameter should be high as possible. Early voltage can find

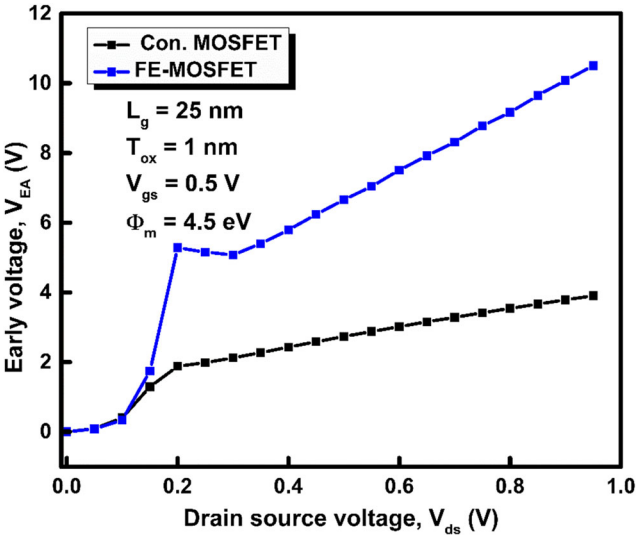


Fig. 8 Comparison of early voltage vs. drain source voltage for conventional MOSFET and ferroelectric based substrate MOSFET structure

out according to Eq. (4) [23, 24]. The highest value of early voltage is recorded for the FE-MOSFET configuration in comparison to conventional MOSFET structure due to the improved SCEs as listed in Table 2. Figure 9 displays the variation of A_V with respect to the variation of V_{gs} at constant V_{ds} at 0.5 V. These value of gain should be high for better performance devices. Also, the maximum value of intrinsic gain is obtained for the FE based substrate MOSFET structure because of the higher V_{EA} and g_m as formulated in Eq. (5). Intrinsic gain defines the overall voltage gain for a device regardless of its bias point [24]. The observed values of SCEs and analog parameters are tabulated in Table 2.

4.3 RF analysis

With keeping the view of RF application, RF parameters play an important role in device performance. Some important RF parameters like unity gain frequency (f_T), GBP, GFP, TFP, and GTFP are evaluated for both conventional MOSFET and ferroelectric based substrate MOSFET device structures. The variation of two parasitic capacitances C_{gs} and C_{gd} as a function of V_{gs} is displayed in Fig. 10. Both the parasitic capacitance plotted in the linear scale. For extraction of these gate to source capacitance and gate to drain capacitance, a small AC signal analysis has been simulated at 1 MHz operating frequency with DC voltage varying from 0 to 1 V with 0.05 step size. In Fig. 10, it is shown that both parasitic capacitances increase gradually with respect to the V_{gs} shows a higher value for the FE-MOSFET than the conventional MOSFET structure due to the high lateral field which directly connects to the increment in charge carrier movement from source to drain side. Another reason for high parasitic capacitances is that there is an increase in gate capacitance with an increase in dielectric permittivity [25, 26]. For gate to source capacitance, the curve shows less increment behavior than the conventional MOSFET but the peak value is shown by FE-MOSFET. Similarly, the curve for drain source capacitance of FE-MOSFET first shows the lower value than the conventional MOSFET but after the certain value around the threshold voltage, it shows the increased behavior in comparison of conventional MOSFET [27].

In next Fig. 11, simulated data of gate capacitance (C_{gg}) is displayed as a function of V_{gs}. The C_{gg} curve shows the same variation as the parasitic capacitances as C_{gg} is a combination of both C_{gs} and C_{gd}. so, a higher value of gate capacitance is shown by the FE-MOSFET than the conventional MOSFET configuration [25–27]. Now, in Fig. 12, the variation of unity gain frequency as a function of V_{gs} is displayed for both the device configuration. Basically, f_T defines the frequency value at which the device obtain unit current gain [28]. The formula used for calculation of f_T is shown in Eq. 6, the maximum value of f_T is shown from the data of FE- MOSFET structure

Table 2 List of analog and electrostatic parameters for conventional MOSFET and ferroelectric based substrate MOSFET

Parameter	Conventional MOSFET	Modified FE-MOSFET
I_{on}(A/ μm)	4.273×10^{-4}	4.427×10^{-4}
I_{off}(A/ μm)	2.587×10^{-8}	7.623×10^{-10}
I_{on}/I_{off}	1.656×10^{4}	0.581×10^{6}
V_{th}(lin) (V)	0.341	0.272
V_{th}(sat) (V)	0.281	0.246
SS_{lin} (mV/dec)	87.844	73.131
SS_{sat} (mV/dec)	109.405	76.128
g_m (mS)	0.068	0.739
TGF (V^{-1})	44.076	73.983
V_{EA} (V)	3.907	10.503
A_V (dB)	13.778	51.477

in a comparison of conventional MOSFET structure as the g_m is higher for the FE-MOSFET structure.

$$f_T = \frac{g_m}{2\pi\left[C_{gs} + C_{gd}\right]} \tag{6}$$

$$GBP = \frac{g_m}{20\pi\left[C_{gd}\right]} \tag{7}$$

$$GFP = \left[f_T\right]\frac{g_m}{g_d} \tag{8}$$

$$TFP = \frac{g_m}{I_D} \times f_T \tag{9}$$

$$GTFP = \frac{g_m}{g_d} \times \frac{g_m}{I_D} \times f_T = Av \times TFP \tag{10}$$

Figure 13 represents the variation of GBP and GFP with V_{gs} at constant V_{ds} of 0.5 V for conventional MOSFET and FE-MOSFET. GBP and GFP are the essential performance parameter that is used for high-frequency applications [29, 30]. Both parameters are formulated in the Eqs. (7) and (8). As the ratio of g_m to C_{gd} is higher and shows the maximum value for the FE-MOSFET in comparison of conventional MOSFET which shows the better device performance .so, the modified FE-MOSFET shows the enhanced results for GBP than the conventional MOSFET structure. The peak value of GBP indicates the voltage at which the gain of the device becomes maximum [30]. Secondly, the GFP factor is the product of unity frequency gain and intrinsic gain as formulated in Eq. (8) which both are high for the FE-MOSFET configuration than the conventional one. So, as displayed in Fig. 13, the GFP shows a gradual increase with the increase in

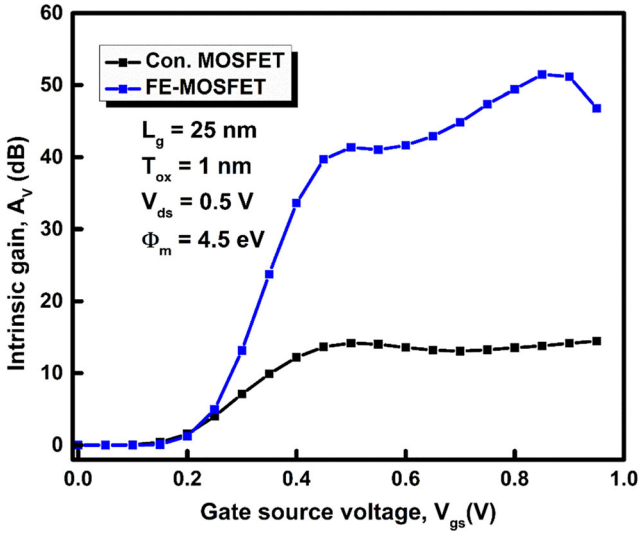


Fig. 9 Variation of A_v with V_{gs} at constant V_{ds} = 0.5 V for conventional MOSFET and ferroelectric based substrate MOSFET structure

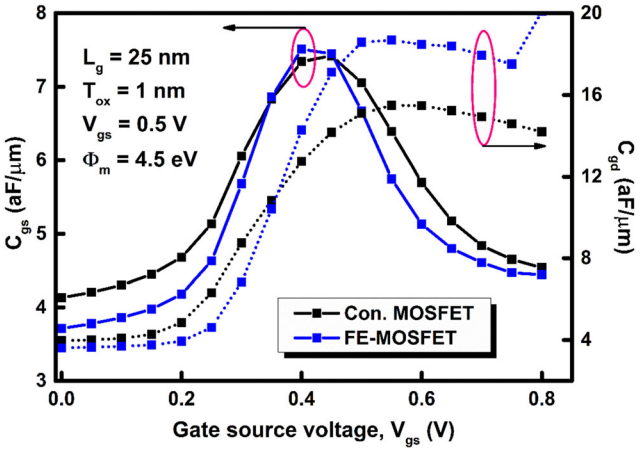


Fig. 10 Variation of parasitic capacitance (C_{gs} and C_{ds}) with V_{gs} at V_{ds} = 0.5 V

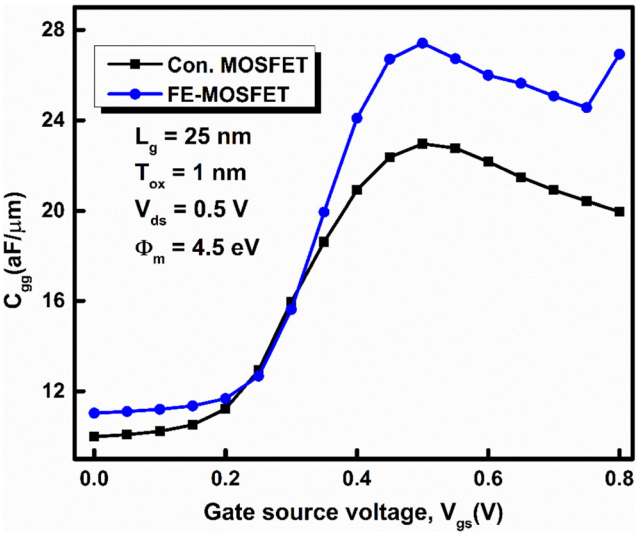


Fig. 11 Variation of gate capacitance with V_{gs} at constant V_{ds} of 0.5 V for conventional MOSFET and FE-MOSFET

gate source voltage and the maximum peak value of GFP is shown by the modified FE-MOSFET structure in a comparison of conventional MOSFET structure.

Other important RF performance parameters are TFP and GTFP as formulated in Eqs. (9) and (10), respectively. Fig. 14 illustrates the variation of TFP and GTFP with an increase in the V_{gs} at constant V_{ds} of 0.5 V for both the configuration. TFP curve shows the gradually increase with respect to the V_{gs} at constant V_{ds} of 0.5 V. As TFP is a product of transconductance generation factor and unity gain frequency as formulated in Eq. (9), which are higher for modified FE-MOSFET configuration in comparison of conventional MOSFET structure. Thus, the maximum value for TFP is shown by the FE-MOSFET device structure.

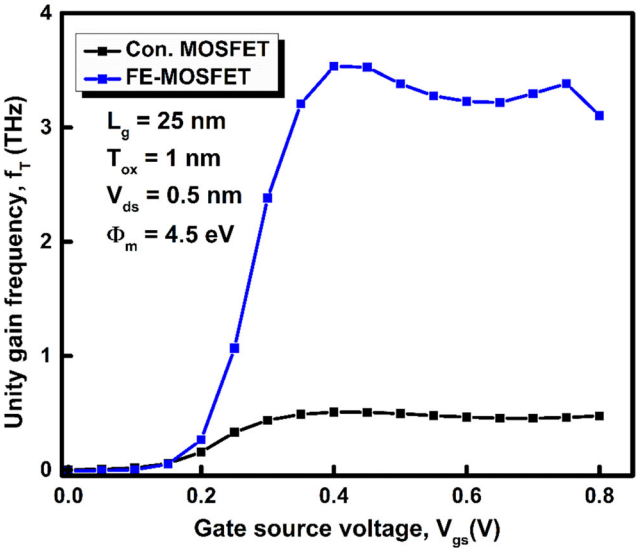


Fig. 12 Comparison of f_T variation vs. V_{gs} at constant V_{ds} = 0.5 V for conventional MOSFET and ferroelectric based substrate MOSFET (FE-MOSFET)

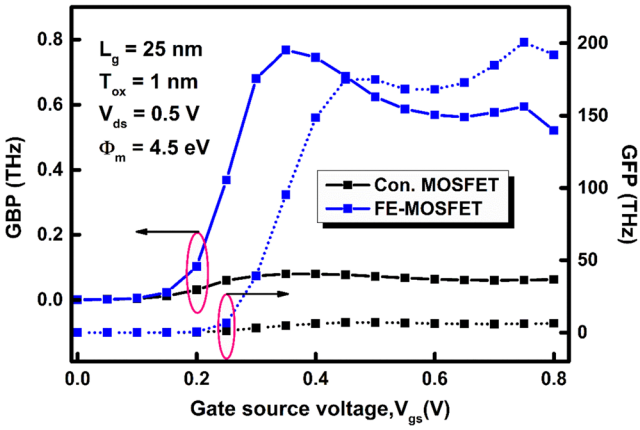


Fig. 13 The GBP and GFP for conventional MOSFET and ferroelectric based substrate MOSFET with V_{gs}

Similarly, from Eq. (10), we can see that the GTFP is dependent on the intrinsic gain and the TFP factor .So, for the high value of A_V and TFP, GTFP would be high. Consequently, we get a higher GTFP value for the FE-MOSFET configuration than the conventional one. Thus, modified FE-MOSFET is a suitable configuration in terms of gain, transconductance and speed applications. All the RF parameters are tabulated in Table 3.

5 Application of ferroelectric based substrate MOSFET

For digital applications, combinational circuits are highly demanded because of their low power consumption and high switching speed. Several designs are explored to meet the objectives. The use of ferroelectric based substrate MOSFET is an example of this effort at the device level. Here, the one brief application of FE-MOSFET is defined and presented using the COGENDA Visual TCAD simulator. And, the performance is compared with conventional MOSFET structure.

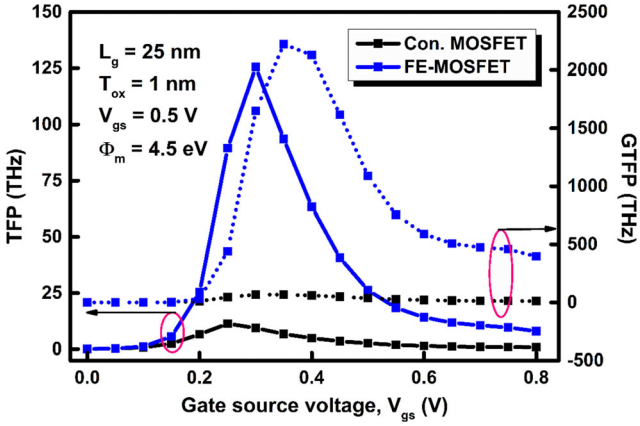


Fig. 14 Variation based comparison of TFP and GTFP with respect to V_{gs} for conventional MOSFET and ferroelectric based substrate MOSFET

Table 3 Summary of RF device performance parameters for conventional MOSFET and FE-MOSFET

Parameter	Conventional MOSFET	Modified FE-MOSFET
C_{gs}(aF/ μm)	7.342	7.507
C_{gd}(aF/ μm)	15.488	18.674
C_{gg}(aF/ μm)	22.959	27.411
f_T(THz)	0.506	4.419
GBP(THz)	0.796	7.675
GFP(PHz)	0.006	0.200
TFP(THz)	11.349	125.490
GTFP(PHz)	0.066	2.221

5.1 Inverter

For showing the application of inverter, n-channel and p-channel FE-MOSFET are designed. For electrical isolation, a spacer of 10 nm silicon oxide is used between the n-channel and p-channel FE-MOSFET. The gate contact of n-channel and p-channel FE-MOSFET are of NpolySi with a work function of 4.5 eV and 4.95 eV, respectively. For p-channel FE-MOSFET, doping material of n-type is used for substrate region and doping material of p-type is used for source and drain region with a concentration of 1×10^{17} cm^{-3} and 1×10^{20} cm^{-3} respectively. The general circuit diagram explaining the FE-MOSFET based inverter is shown in the Fig. 15.

The I_D-V_{gs} characteristics of both channel configuration devices are plotted together to match the V_{th} as shown in Fig. 16. The V_{gs} swept from -1 to 1 V at constant V_{ds} 0.5 V with dual work function metal (DWFM) integration scheme.

The voltage transfer characteristic (VTC) of the designed FE-MOSFET inverter and conventional MOSFET based inverter are displayed in Fig. 17. As from Fig. 17, it is clear that the transition range (TR) of FE-MOSFET based inverter is sharper in comparison to conventional MOSFET inverter which signifies the high speed and fast switching for a device [31].

The two important critical points on the curve where the slope of voltage transfer characteristic becomes -1 are corresponding to logic 0 and logic 1 which are denoted as maximum input voltage (V_{IL}) and minimum input voltage (V_{IH}), respectively [31]. And, the threshold voltage of the inverter configuration is the transition voltage or a voltage point at which $V_{in} = V_{out}$ on the VTC curve. In this case, the switching threshold voltage for the FE-MOSFET inverter is 0.43 V and for the conventional MOSFET based inverter is 0.42 V.

5.2 Noise margin calculation

Noise margin (NM) defines the immunity of noise for a circuit. For a better circuit performance, the noise margin should be high as possible. For low signal level (NM_L) and high signal levels (NM_H), the noise margin is tabulated as:

$$NM_L = V_{IL} - V_{OL} \tag{11}$$

$$NM_H = V_{OH} - V_{IH} \tag{12}$$

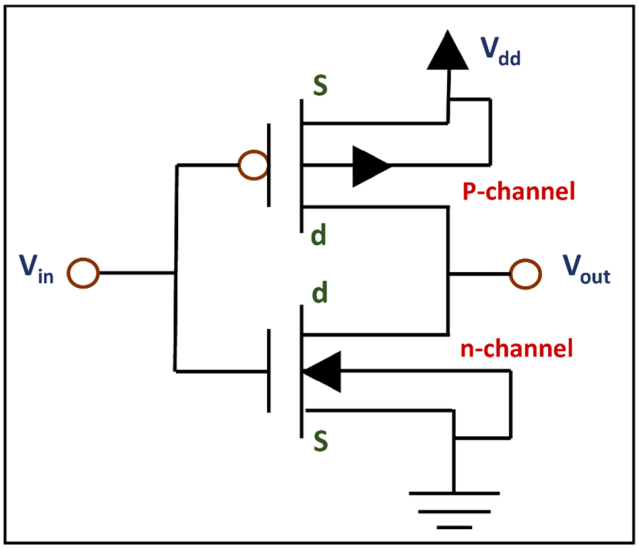


Fig. 15 Circuit diagram of FE-MOSFET based inverter

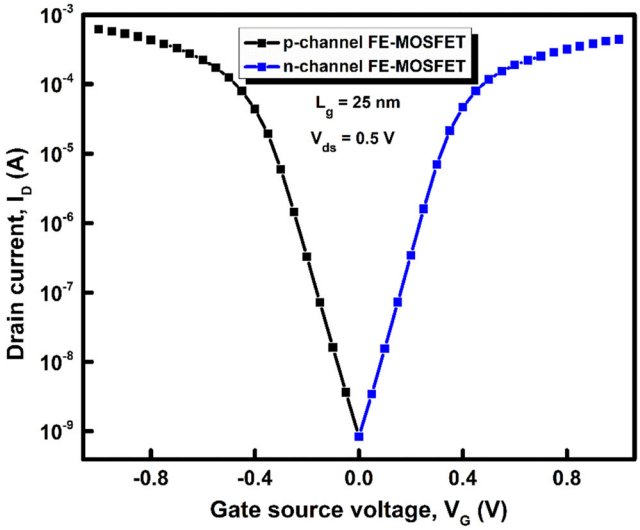


Fig. 16 Threshold voltage matching for n-channel FE-MOSFET and p-channel FE-MOSFET structure

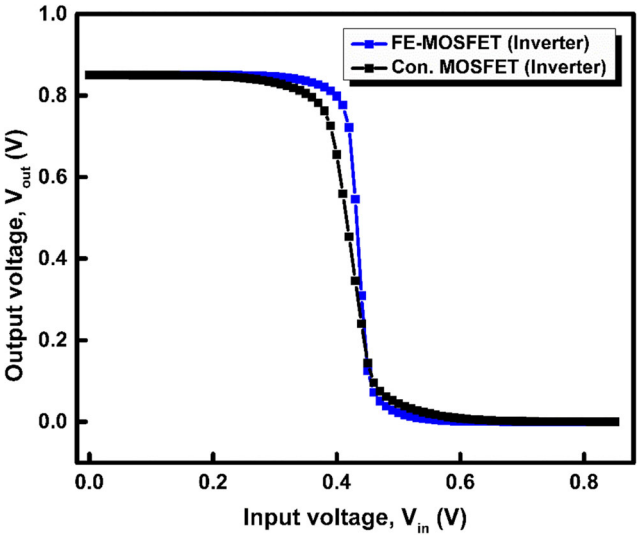


Fig. 17 Voltage Transfer Characteristics of conventional MOSFET and FE-MOSFET based inverter

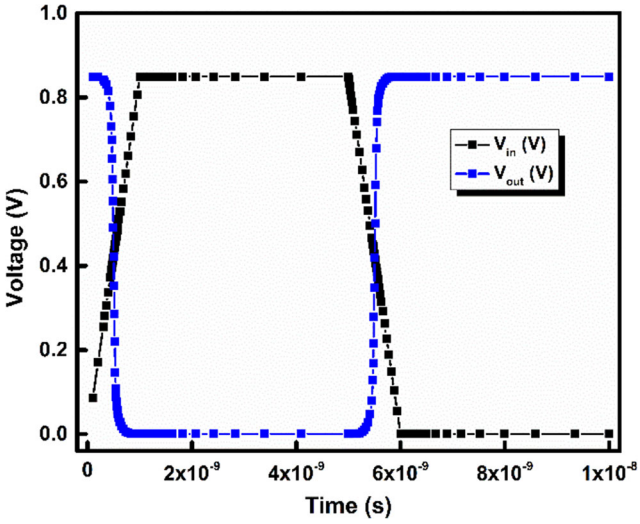


Fig. 18 Transient plot corresponding FE-MOSFET based inverter

Where, V_{OL} and V_{OH} are minimum output voltage and maximum output voltage when corresponding output levels are logic 0 and logic 1, respectively. The parameters that are required to compute the noise margin (NM) and transition region (TR) are tabulated in Table 4.

The transient analysis for FE-MOSFET based inverter is displayed in Fig. 18. All the simulation is done at room temperature (300 K) and using the drift-diffusion method solver level 1(DDML1) technique. The input pulse and V_{DD} of 0.85 V are common for both the analysis. As in inverter, the output of the inverter is at logic 1 when n-channel FE-MOSFET is in cut-off mode and it shows the logic 0 in output when the n-channel FE-MOSFET conducts.

6 Conclusions

In conclusion, the effect of ferroelectric material in the substrate region in between the high-k HfO_2 layers on the conventional MOSFET structure in terms of the analog and RF parameters using the visual TCAD simulator was investigated. The proposed architecture shows the enhanced analog and RF parameters with the reduced SCEs. Modified FE-MOSFET structure shows a better switching ratio as the off current is decreased by 97.05 % with 3.60 % increased I_{on} current. Also, the SCEs like the subthreshold swing is decreased by 16.75 % and 30.42 % in linear and saturation region, respectively. DIBL is reduced by 56.51 % and V_{th} is reduced by 20.37 % and 12.59 % in linear and saturation region, respectively. The analog parameters like g_m and TGF enhanced with 977.57 % and 67.85 %, respectively for the proposed architecture. The V_{EA} and A_V show improved results for the modified configuration with 168.81 % and 273.60 %, respectively.

Furthermore, RF performance parameters like GBP and GFP increased by 864.20 % and 3184.02 %, respectively. The TFP enhanced by 1000.69 % and GTFP gets improved by 3223.05 % for the modified structure. The unity gain frequency (f_T) is increased by 772.28 %. For the modified architecture. Furthermore, from the noise margin parameters, it is concluded that the transition range for the proposed device structure is better than the conventional MOSFET structure as NM_L and NM_H are improved with 27.09 % and 24.81 %. So, with better NM and TR, FE-MOSFET based inverter is a better option for digital applications. Thus, from the above results and discussion, the proposed device structure of FE-MOSFET can be considered as an alternate choice for designing analog and RF circuit devices.

Table 4 Comparison of noise margin for conventional MOSFET and FE-MOSFET based inverters

Parameter	Conventional MOSFET	FE-MOSFET
V_{IL} (V)	0.27	0.34
V_{IH} (V)	0.60	0.54
NM_L (mV)	261.7	332.6
NM_H (mV)	239.4	298.8
TR (V)	0.33	0.20

Acknowledgements The authors are grateful to the Microelectronics Research Laboratory, Delhi Technological University, for supporting the work.

Authors' Contribution All authors contributed to the study's conception and design.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability The authors mentioned above have all the relevant data associated with this research work and will be dedicated to sharing that they will be asked to do so in the future.

Declarations The authors have seen all the Ethical Standards and will supposed to follow them in the future.

Conflict of Interests The authors declare that they have no known conflict of interests or personal relationships that could have appeared to influence the work reported in this paper.

Consent to Participate & for Publication Since the concerned research paper is based on the 'non-life science journal.' So, 'Not Applicable' here. However, the authors have gone through all journal policies and consent to the authorities for further processing.

References

1. Moore GE (1998) Cramming more components onto integrated circuits. Proc IEEE 86:82–85
2. Jeon DY, Park SJ, Mouis M, Barraud S, Kim GT, Ghibaudo G (2013)Low-temperature electrical characterization of junctionless transistors. Solid State Electron 80:135–141
3. Kumar A, Gupta N, Chaujar R (2016) TCAD RF performance investigation of transparent gate recessed channel MOSFET. Microelectron J 49:36–42
4. Doria RT, Pavanello MA, Lee CW, Ferain I, Akhavan ND, Yan R, Razavi P, Yu R, Kranti A, Colinge JP (2010) Analog operation and harmonic distortion temperature dependence of nMOS Junctionless transistors. ECS Trans 31:13–20
5. Colinge JP, Lee CW, Afzalian A, Akhavan ND, Yan R, Ferain I, Razavi P, O'Neill B, Blake A, White M, Kelleher AM, McCarthy B, Murphy R (2010) Nanowire transistors without junctions. Nat Nanotechnol 5:225–229
6. Awadhiya B, Kondekar PN, Yadav S, Upadhyay P (2020) Insight into threshold voltage and drain induced barrier lowering in negative capacitance field effect transistor. Trans Electr Electron Mater
7. Salahuddin S, Datta S (2008) Use of negative capacitance to provide voltage amplification for low power nanoscale devices. Nano Lett 8:405–410
8. Goswami Y, Ghosh B, Asthana PK (2014) Analog performance of Si junctionless tunnel field effect transistor and its improvisation using III-V semiconductor. RSC Adv 4:10761–10765
9. Lo SH, Buchanan DA, Taur Y, Wang W (1997) Quantum mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's. IEEE Electron Device Lett 18: 209–211
10. Ribes G, Mitard J, Denais M, Bruyere S, Monsieur F, Parthasarathy C, Vincent E, Ghibaudo G (2005) Review on high-k dielectrics reliability issues. IEEE Trans Device Mater Reliab 5:5–19
11. Robertson J (2004) High dielectric constant oxides. Eur Phys J Appl Phys 28:265–291
12. Onishi K, Kang CS, Choi R, Cho HJ, Gopalan S, Nieh RE, Krishnan SA, Lee JC (2003) Improvement of surface carrier mobility of HfO2 MOSFETs by high-temperature forming gas annealing. IEEE Trans Electron Devices 50:384–390
13. Kerber A, Cartier E, Pantisano L, Degraeve R, Kauerauf T, Kim Y, Hou A, Groeseneken G, Maes HE (2003) Origin of the threshold voltage instability in SiO2/HfO2 dual layer gate dielectrics. IEEE Electron Device Lett 24:87–89
14. Gupta N, Chaujar R (2016) Optimization of high-k and gate metal work function for improved analog and intermodulation performance of gate stack (GS)-GEWE-SiNW MOSFET. Superlattice Microst 97:630–641
15. Kim HJ, Park MH, Kim YJ, Lee YH, Moon T, Do Kim K, Hyun SD, Hwang CS (2016) A study on the wake-up effect of ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ films by pulse-switching measurement. Nanoscale 8:1383–1389
16. Bhuyan MH (2019) A review of recent research works on negative capacitance field effect transistor 13:10
17. Tu L, Cao R, Wang X, Chen Y, Wu S, Wang F, Wang Z, Shen H, Lin T, Zhou P, Meng X, Hu W, Liu Q, Wang J, Liu M, Chu J (2020) Ultrasensitive negative capacitance phototransistors. Nat Commun 11:4–11
18. Awadhiya B, Kondekar PN, Yadav S, Upadhyay P (2021) Insight into threshold voltage and drain induced barrier lowering in negative capacitance field efect transistor. Trans Electr Electron Mater 22:267–273
19. Roy G, Lema FA, Brown AR, Roy S, Asenov A (2005) Simulation of combined sources of intrinsic parameter fluctuations in a 'real' 35 nm MOSFET. Proc ESSDERC :337–340
20. Pradhan KP, Mohapatra SK, Sahu PK, Behera DK (2014) Impact of high-k gate dielectric on analog and RF performance of nanoscale DG-MOSFET. Microelectron J 45:144–151
21. Kumar A, Tripathi MM, Chaujar R (2018) Comprehensive analysis of sub-20 nm black phosphorus based junctionless-recessed channel MOSFET for analog/RF applications. Superlattices Microstruct 116:171–180
22. Jena B, Dash S, Routray SR, Mishra GP (2019)Inner-gate-engineered GAA MOSFET to enhance the electrostatic integrity. NANO 14:1–8
23. Narendar V, Girdhardas KA (2018) Surface potential modeling of Graded-Channel gate-stack (GCGS)high-K dielectric dual-material double-gate (DMDG) MOSFET and analog/RF performance study. Silicon 10:2865–2875
24. Suddapalli SR, Nistala BR (2021) The analog/RF performance of a strained-Si graded-channel dual-material double-gate MOSFET with interface charges. J Comput Electron 20:492–502
25. Madan J, Gupta RS, Chaujar R (2017) Performance investigation of heterogeneous gate dielectric-gate metal engineered–gate all around-tunnel FET for RF applications. Microsyst Technol 23: 4081–4090
26. Madan J, Chaujar R (2017) Numerical simulation of N^+ source pocket PIN-GAA-Tunnel FET: Impact of interface trap charges and temperature. IEEE Trans Electron Devices 64:1482–1488
27. Malik P, Gupta RS, Chaujar R, Gupta M (2012) AC analysis of nanoscale GME-TRC MOSFET for microwave and RF applications. Microelectron Reliab 52:151–158
28. Mohapatra SK, Pradhan KP, Artola L, Sahu PK (2015) Estimation of analog/RFfigures-of-merit using device design engineering in gate stack double gate MOSFET. Mater Sci Semicond Process 31:455–462
29. Kumar B, Chaujar R (2021) TCAD temperature analysis of gate stack gate all around (GS-GAA) FinFET for improved RF and wireless performance. Silicon
30. Kumar A, Tripathi MM, Chaujar R (2017) Investigation of parasitic capacitances of In_2O_5Sn gate electrode recessed channel MOSFET for ULSI switching applications. Microsyst Technol 23:5867–5874
31. Singh N, Pandey R (2020) Design of 28 nm GAAFET and its digital applications. Int J Adv Sci Technol 29:14074–14088

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ACADEMIA
Accelerating the world's research.

The Merger of Topology Optimisation in Additive Manufacturing

IJRAME Journal, Srashti Saxena

IJRAME PUBLICATIONS

Cite this paper

Downloaded from Academia.edu

Get the citation in MLA, APA, or Chicago styles

Related papers

Download a PDF Pack of the best related papers



TOPOLOGY DESIGN OPTIMIZATION OF RIBS, SPARS, WING ATTACHMENT BULKHEAD AND NO...
IJRAME Journal, A KANNAN

SURFACE ROUGHNESS OPTIMIZATION IN END MILLING USING TAGUCHI METHOD AND ANOVA
IJRAME Journal

Additive manufacturing and non-destructive testing of topology-optimised aluminium components
michael scheerer

IJRAME

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22

The Merger of Topology Optimisation in Additive Manufacturing

***Prof. A. K. Madan*[1]**
Department of Mechanical Engineering
Delhi Technological University,
Delhi 110042

***Srashti Saxena*[2]**
Department of Mechanical Engineering
Delhi Technological University,
Delhi 110042
srashti.saxena99@gmail.com

Abstract: Topology optimization has become a subject of study unto itself in recent years, not only because it provokes the intellectual community's curiosity, but also because it provides significant responses to real-life industrial difficulties. Topology optimization is being used to create a lightweight product while still meeting its functional criteria for a new product or a redesign of an existing product. Additive Manufacturing is defined as the layer-by-layer production of geometrically demanding components, which significantly minimises the complexity restrictions imposed on topology optimization by the old subtractive process. This perspective paper enlists the compassion of these two methods as they work together. The paper compiles the methodology along with the merits and demerits of the optimisation technique. A review has been done on lightweight design and continuum optimisation, which resulted in a reduction of weight by 65% and the use of grid-like ground structures. The goal is to motivate both researchers and engineers to come up with creative solutions to face the modern-day challenge in this field.

Keywords: *Additive Manufacturing, Continuum Optimisation, Lightweight Design, Topological Optimisation.*

1. Introduction

Topology optimization is a comprehensive design technique for making the appropriate structure configuration by allocating materials in a rational way while maintaining given load conditions, performance, and limitations. This method generates an optimal topology by solving a material distribution problem. Using dedicated software, it discards the material that is not subject to stress, therefore making the product effective to produce. The very first breakthroughs in topology optimization relied on subtractive or formative

IJRAME

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22

manufacturing techniques. Because of their distinct manufacturing constraints, these traditional manufacturing processes are limited in their ability to produce complicated form of geometries. But with advancements in manufacturing in the form of additive manufacturing, the complexities in the geometries could be included, making it the finest customer. Additive manufacturing does not require any kind of tool motion for the object to be produced, and therefore, layer-by-layer printing can leash the complexities of a geometry. This method relies on layer-by-layer material deposition, which drastically reduces the geometric complexity barrier. From the perspective of geometric complexity [18], manufacturing performance and costs of production are not affected. Topology optimization

[2]Corresponding Author: Srashti Saxena; srashti.saxena99@gmail.com; 7060428334

creates lightweight parts without sacrificing strength. Topology optimization in conjunction with Additive Manufacturing allows for the manufacturing of low-weight and intricate systems. It relocates the material within the design space and does not entail the integration of exterior barriers, resulting in an organically shaped component that bears little similarity to its heritage counterpart and is considerably lighter in weight. STELIA Aerospace has employed topology optimization to create more stable fuselage panels for aeroplanes, combining rapid prototyping and topology optimisation. More than merely making a physical product, additive manufacturing is about putting design and innovation front and centre. One of the key advantages of additive manufacturing is the ability to be creative without worrying about cost or schedule constraints. Additive manufacturing can considerably cut energy consumption when compared to traditional manufacturing by utilising less material and eliminating the unnecessary steps.

With the continuation of the advancements from the last three decades, topology optimisation has still not been able to get the right track and has been criticised for its organic nature due to challenges being faced during the production. It is challenging to ensure that a topologically optimum design is both manufacturable and aesthetically acceptable (contrasts from what a mechanical part would look like in the typical sense).

IJRAME

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22

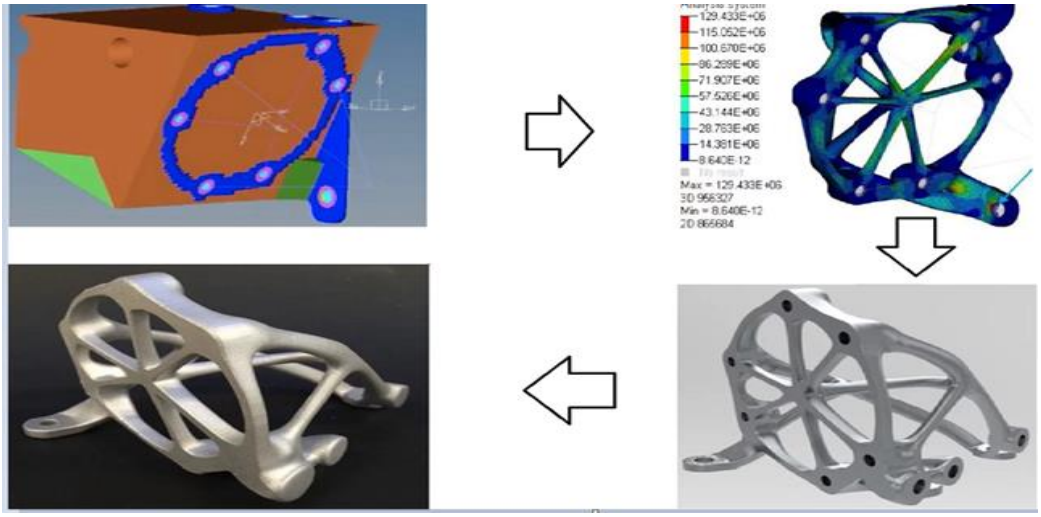


Figure 1. Stages in Topology Optimisation [16]

2. Approach of topology optimisation

2.1 Topology Optimisation Strategic Methodology

CAD has been used in topology optimization to develop a rough/initial model of the object to be optimised, while Finite Element Analysis is used in seeing how stresses and displacements are distributed throughout the product. The Solid Isotropic Material with Penalization (SIMP) method, the Evolutionary Structural Optimization (ESO)/Bidirectional Evolutionary Structural Optimization (BESO) method, and the level set method have emerged as the three main advances in the fields of topology optimization [21]. It is used to eliminate portions of the component that are not effectively sustaining the applied loads, are not experiencing significant deformation, and therefore do not contribute to the overall function [16]. The optimisation process starts with the designing of the model on the CAD software and optimising it until the desired design is achieved. Based on the objectives and limitations established in the design challenge, the topology optimization tools build a complicated natural shape. Following the shape obtained by the topology optimization procedure, the design is next finished on CAD software to produce a smooth and manufacturable part. Finally, the final

IJRAME

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22

optimised design is evaluated using Finite Element Analysis techniques to ensure that the product satisfies its overall performance criteria [16].

2.2 Initial steps of optimization-

1. DESIGN-The process starts as the design has been finalised on the CAD software, which fulfils the initial performance requirements and serves as the basis for the process of optimisation.
2. CONSTRAINTS- The initial conditions of the optimization require boundary conditions; the design space is the area in which the optimization process takes place. This area is the boundary of the existing design. The algorithm eliminates the material within this region only. Part functionality lets the user specify the conditions of the load and the fixed points along with the material of the object to be optimised. As the load and the points vary, part of the functionality process is repeated again.
3. MESHING- The object to be optimised is converted into the mesh. It allows the user to specify different parameters of the elements, the mesh quality as fine or coarse, shape and size of the elements etc. The coarser the mesh is, the less accurate the results are, and the simulation is completed in a short period of time.
4. OPTIMISATION VARIABLE-

    a) Desired Weight- The optimisation process ceases when the desired weight has been achieved. This desired weight is an input variable around which the optimisation process revolves, but fails to work in conditions where a maximum reduction in weight is required.
    b) Number of Runs-The number of runs that can be set up in order to complete the optimization process in a given period of time. Unless the time taken is unpredictable for the optimisation.
    c) Convergence- The convergence criteria allow the optimisation to stop within two to three iterations if the result of the two output weights is less than the value predefined.
    d) Factor of Safety- It specifies the maximum or minimum stress which the desired product can bear.

IJRAME

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22

5. FINAL PRODUCT- The given steps one to four are run in order to get the optimised parts. Finite Element Analysis is done on the optimised part in order to ensure whether the stresses induced and deflections are within the specified limit or not.



Figure 2. Flow Chart of Product Manufacturing

3. Case Study-

3.1 Continuum Optimisation-

In this study, a review has been done on structural steel design using topology optimisation for its approach up to shape and size only [12]. Topology optimisation is easy to define in shape and size, whereas it has restrictions to variable cross section, the geometry of the product and to some pre-defined restraint variable. A key concern in structural optimization is reliability-based optimization, which considers inconsistencies in geometric dimensions, properties of the material, forces, operating conditions, and so forth that occur in real engineering issues. Until now, numerous reliable ways such as probability methods and interval methods have been proposed [19]. It does, however, make every effort to improve on the existing design. It is significant to mention that Topology Optimisation is not confined to materially homogeneous volumes. It could be used for a heterogeneous material, such as composite or microstructurally created materials. The weight of steel is reduced by 18% to 75%, that has a massive effect on steel structure. Nevertheless, when it concerns to interconnections in structural steel, the reduction is usually 12 % to 25%, which is the majority of the total weight of the steel structure, and its design concept has a significant impact on its weight and effectiveness, making it particularly prone to Topology Optimisation. Over the last several decades, the continuum topology optimization method

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22

has been fruitfully developed. It can also be used in grid-like "ground structures" composed of one-dimensional elements. Continuum structures are ground structures with one dimensional element subjecting itself to Discrete Optimisation. With the objective of finding out the optimal number, size, shape, coordinates of the nodes, and structural element connectivity, the optimisation is restricted only to the shape and size. Due to this reason, it is also known as Truss Topology Optimisation or Skeletal Structures Topology Optimisation. The optimisation of nodes is considered one of the major issues [12]. These optimisation techniques are applicable for solids and shells with definite boundaries.

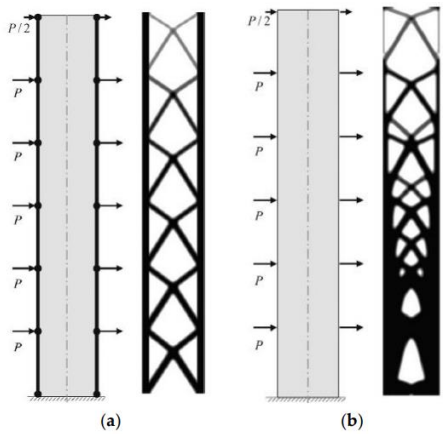


Figure 3. (a) Discrete optimisation (b) Continua Optimisation [12]

3.2 Lightweight Design-

In this study, a review has been done on a jet engine bracket to depict its weight reduction merits and the adoption of additive manufacturing techniques for production purposes [16, 25]. Finite element analysis has been done to investigate the distribution of the stress and the Von Misses distribution contour has been plotted. The software shows an area in blue, which indicates the material which needs to be removed has an insignificant effect on the performance of the parts. The bracket is subjected to four load conditions and the objective of the optimisation technique is to reduce its weight up to a satisfactory level. The characteristics of the material used for the bracket are-

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22

1) The material used is Ti-6Al-4V with a strength of 904MPa at a 23 degree C temperature.
2) The diameter of the pin is 1:19 mm.
3) The pin's stiffness is infinite.
4) The maximum diameter of the nut face is 10.287 mm (inner diameter) and 14.173 mm (minimum outer diameter).

The optimisation is subject to minimising the weight of the bracket. Subject to

1. 30% mass fraction
2. Minimum feature size: > 1.13 mm.

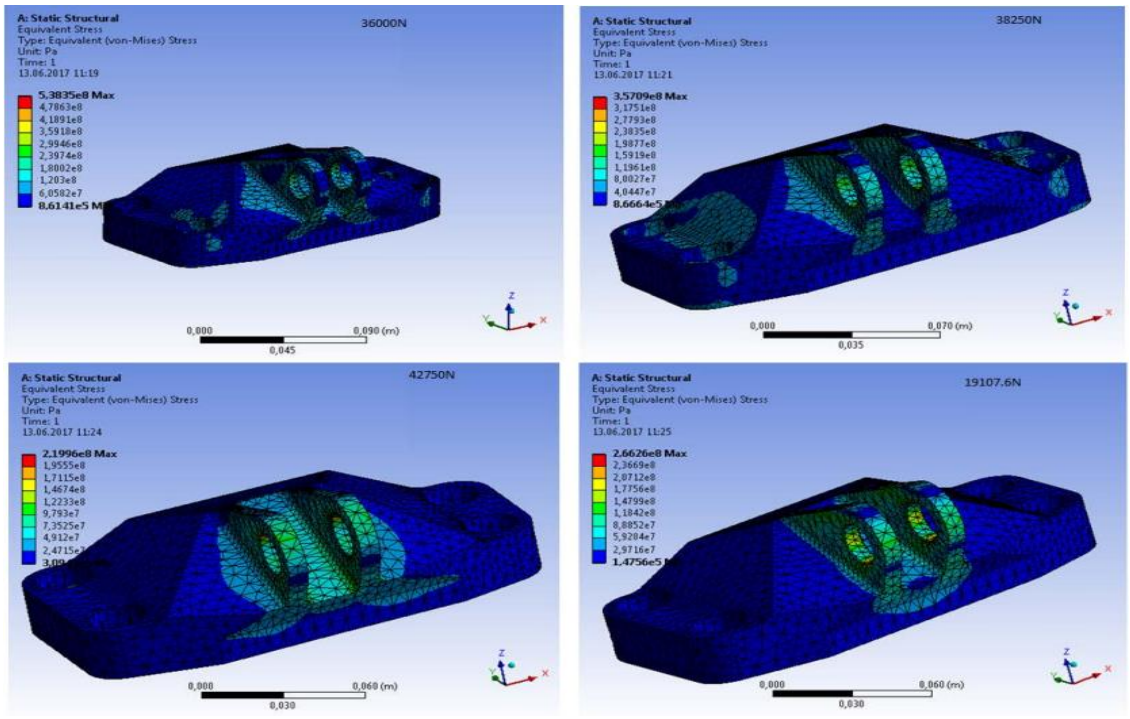


Figure 4. Von Mises distribution contour plot marked by blue [16]

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22

The optimisation starts by dividing the design into two parts, the first one being the design space and the second one being the non-design space where the topology optimisation tools work. These spaces are then meshed using the tool present in the Altair Hypemesh 14 Optistruct software for commercial uses and a fine mesh is used of element size 0.25 mm. The non-design spaces are then subjected to loading conditions and constraints, and instead of torque, two equal couple forces are applied in different directions. The topology optimisation tool evaluates the distribution of stresses applied and the displacement for all the loads distinctly. Depending on this stress and displacement distribution, the regions where insignificant loads are acting are eliminated by the software. A rough geometry is formed that is then remodelled to give a pleasant and an aesthetic look. Figure 10 [16] shows the design which has been finalised and is created in Autodesk Inventor Professional. This final design bears the mechanical load as per the prerequisite condition. This structural verification is performed on the software ANSYS R17 Academia Educational Tool where the von Mises stress value is checked and compared with the yield strength of the material used. The design, when fully examined for stress, is then manufactured by a rapid prototyping method using the technique of fused deposition modelling (FDM). Figure 12 [16] shows the final product, and it can be concluded that the final product of the jet engine bracket resulted in a 65% reduction in weight. The mass is reduced from 2.067 kg to 0.72 kg.

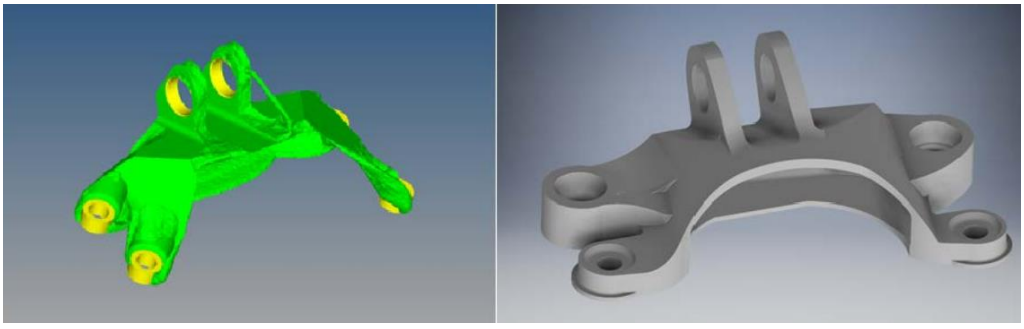


Figure 5. Optimal Geometry and Remodeled design [25]

IJRAME

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22



Figure 6. Engine Bracket after Fused Deposition Modelling (Additive Manufacturing) [16]

4. Conclusion

Here a detailed account of Topology Optimisation methodology was explored, as well as its conjunction with the additive manufacturing. The optimising technique's strategic planning has been reviewed in detail so that it may be simply comprehended. This paper has examined its benefits and drawbacks. The case study shows that the continuous optimization is restricted to changing cross section and dimension, however various efforts have been done to improve the current product. It can now be applied to both homogeneous and heterogeneous materials. The second case study explains how the product's redundant stress-free portion can be discarded hence the weight minimized. The weight of the jet engine bracket has been lowered by 65%, from 2.067 kg to 0.72 kg. Overall, it is reasonable to state that topology optimization, in association with additive manufacturing, is important in the design and manufacture of lightweight goods for usage in a variety of industries.

5. Future Scope

The anticipated use of Additive Manufacturing is the most significant driving force for implementing Topology Optimisation in design practice. This technique should be used in the majority sector of manufacturing for waste management. It can find its role in Space Science for the manufacturing of satellites and rockets. Also,

IJRAME

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22

for more complexity, the software tools should be able to complete the optimisation technique within a given span of time.

REFERENCES

[1] Lee, Jian-Yuan & Arun Prasanth, Nagalingam & Yeo, S. H.. (2022). Design Consideration for Additively Manufactured Components Through Topology Optimization and Generative Design for Weight Reduction. 10.1007/978-981-16-5763-4_49.

[2] Almasri W., Bettebghor D., Ababsa F., Danglade F., Adjed F. (2021) Deep Learning Architecture for Topological Optimized Mechanical Design Generation with Complex Shape Criterion. In: Fujita H., Selamat A., Lin J.CW., Ali M. (eds) Advances and Trends in Artificial Intelligence. Artificial Intelligence Practices. IEA/AIE 2021. Lecture Notes in Computer Science, vol 12798. Springer, Cham. https://doi.org/10.1007/978-3-030-79457-6_19

[3] Diab W. Abueidda, Seid Koric, Nahil A. Sobh. Topology optimization of 2D structures with nonlinearities using deep learning, Computers & Structures,Volume 237, 2020, https://doi.org/10.1016/j.compstruc.2020.106283.

[4] Jianan Lu & Yonghua Chen (2013) Topology optimisation and customisation of a prosthetic knee joint design, International Journal of Computer Integrated Manufacturing, 26:10, 968-976, DOI: 10.1080/0951192X.2011.652178

[5] Ningrong Lei, Xiling Yao, Seung Ki Moon & Guijun Bi (2016) An additive manufacturing process model for product family design, Journal of Engineering Design, 27:11, 751-767, DOI: 10.1080/09544828.2016.1228101

[6] Sascha Senck, Michael Happl, Michael Reiter, Michael Scheerer, Manuel Kendel, Jonathan Glinz & Johann Kastner (2020) Additive manufacturing and non-destructive testing of topology-optimised aluminium components, Nondestructive Testing and Evaluation, 35:3, 315-327, DOI: 10.1080/10589759.2020.1774582

IJRAME

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22

[7] Jhunyuan Zhang, Zhongyu Li, Qiang Fang & Chao Chen (2020) Topological optimisation design of passenger car seat backrest frame based on multiple-loading conditions, International Journal of Crashworthiness, 25:5, 581-590, DOI: 10.1080/13588265.2019.1634355

[8] Yun-Fei Fu, Bernard Rolfe, Louis N. S. Chiu, Yanan Wang, Xiaodong Huang & Kazem Ghabraie (2019) Design and experimental validation of self-supporting topologies for additive manufacturing, Virtual and Physical Prototyping, 14:4, 382-394, DOI: 10.1080/17452759.2019.1637023

[9] Zude Zhou, Benyuan Chen, Hong Wang & Zhun Fan (2010) Study on the evolutionary optimisation of the topology of network control systems, Enterprise Information Systems, 4:3, 247-264, DOI: 10.1080/17517570903527861

[10]Erva Ulu, Rusheng Zhang & Levent Burak Kara (2016) A data-driven investigation and estimation of optimal topologies under variable loading configurations, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 4:2, 61-72, DOI: 10.1080/21681163.2015.1030775

[11]Yi Min Xie (Mike) (Prof.), Zhi Hao Zuo (Dr.), Xiaodong Huang (Civil Eng.), Tim Black (Civil Eng.) & Peter Felicetti (Civil Eng.) (2014) Application of Topological Optimisation Technology to Bridge Design, Structural Engineering International, 24:2, 185-191, DOI: 10.2749/101686614X13830790993366

[12]Ribeiro, T.P.; Bernardo, L.F.A.; Andrade, J.M.A. Topology Optimisation in Structural Steel Design for Additive Manufacturing. *Appl. Sci.* **2021**, *11*, 2112. https://doi.org/10.3390/app11052112

[13]Prakash, S. G. (2020). Application Based Design for Additive Manufacturing : Development of a systematic methodolgy for part selection and design for Additive Manufacturing (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-287190

[14]Prakhar Sharma (2020). Computational Modelling & Topology Optimisation of Car Door Hinge

[15]Melissa Orme, Ivan Madera, Michael Gschweitl and Michael Ferrari. Topology Optimization for Additive Manufacturing as an Enabler for Light Weight Flight Hardware. Designs 2018, 2, 51; doi:10.3390/designs2040051

[16]A W Gebisa and H G Lemu 2017. A case study on topology optimized design for additive manufacturing. IOP Conf. Ser.: Mater. Sci. Eng. 276 012026

IJRAME

INTERNATIONAL JOURNAL OF RESEARCH IN AERONAUTICAL AND MECHANICAL ENGINEERING
WWW.IJRAME.COM
ISSN (ONLINE): 2321-3051

Vol.9 Issue 10,
OCTOBER 2021
Pg: -11-22

[17]Prashant Kumar Srivastava, Simant, Sanjay Shukla. Topology Optimization: An Effort to Reduce the Weight of Bottom Centre Pivot. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2, July 2019.

[18]Liu, Jikai & Gaynor, Andrew & Chen, Shikui & Kang, Zhan & Suresh, Krishnan & Takezawa, Akihiro & Li, Lei & Kato, Junji & Tang, Jinyuan & Wang, Charlie & Cheng, Lin & Liang, Xuan & To, Albert. (2018). Current and future trends in topology optimization for additive manufacturing. Structural and Multidisciplinary Optimization. 57. 10.1007/s00158-018-1994-3.

[19]Ji-Hong Zhu, Pierre Beckers, Marc Dahan, Jun Yan, and Chao Jiang. Shape and Topology Optimization for Complicated Engineering Structures. Mathematical Problems in Engineering Volume 2015, Article ID 723897, 2 page http://dx.doi.org/10.1155/2015/723897

[20]Sang J.B., Liu B., Xing S.F., Yang L.C., Qie Y.H. (2006) Application and Research of Structure Topology Optimization of Scraper Conveyer with MSC.Nastran.optishape. In: Computational Methods in Engineering & Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-48260-4_155

[21]Chen, Y.; Wang, Q.; Wang, C.; Gong, P.; Shi, Y.; Yu, Y.; Liu, Z. Topology Optimization Design and Experimental Research of a 3D-Printed Metal Aerospace Bracket Considering Fatigue Performance. Appl. Sci. 2021, 11, 6671. https:// doi.org/10.3390/app11156671

[22]A. Kentli, "Topology optimization applications on engineering structures," Truss and Frames—Recent Advances and New Perspectives, 1–23, 2020, doi: 10.5772/intechopen.90474.

[23]Martin Bendsoe(2008). Topology optimization using the finite volume method.

[24]Tyflopoulos, Evangelos & Flem, David & Steinert, Martin & Olsen, Anna. (2018). State of the art of generative design and topology optimization and potential research needs.

[25]Grabcad Community, https://grabcad.com/challenges/ge-jet-engine-bracket-challenge.

International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934



ORIGINAL ARTICLE

Thermodynamic exergy analysis of two-stage vapour compression refrigeration systems using HFO refrigerants for replacing R134a

R. S. Mishra

Department of Mechanical Engineering, Delhi Technological University Delhi, India

Article Information

Received: 13 December 2020
Revised: 02 January 2021
Accepted: 09 January 2021
Available online: 19 January 2021

Keywords:

HFO refrigerants;
Thermodynamic performance evaluation;
Energy-Exergy Evaluation;
Compression-Absorption; Cascade system;

Abstract

Global warming is one of the unkind environmental concerns that our earth planet is facing nowadays. Main of its causes is the second & third generation of refrigerants that, upon release, remain in the atmosphere for longer periods and contribute towards global warming. This issue could potentially be solved by replacing the earlier generation's high global warming potential (GWP) refrigerants with environmentally friendly refrigerants. In this paper comparative analysis of HFO refrigerants (i.e. R1225ye(Z), R1224yd(Z), R1336mzz(Z), R 1233zd(E), R1243zf, R1234yf, R1234ze(Z) and R1234ze(E)) in two stage vapour compression refrigeration system has been carried out using thermodynamic (energetic and exergetic) performances. The effect of performance parameters such as first law efficiency (COP), second law efficiency (exergetic efficiency) and exergy destruction ratio using ecofriendly refrigerants were investigated. It was found that both energy and exergy efficiencies of R1234ze(Z) is highest and using R1234yf is lowest at -40 ^{o}C evaporating and 40 ^{o}C condensing temperatures. The first law efficiency using R1234ze(Z) is 9.3 % higher than using R134a and R1234yf is 9.10 % lower than using R134a Similarly The second law efficiency using R1234ze(Z) is 8.087% higher than using R134a and R1234yf is 8.44% lower than using R-134a It was also found that for replacing high global warming potential R134a refrigerant. These ecofriendly HFO refrigerants work satisfactory.
©2021 ijrei.com. All rights reserved

1. Introduction

An eminent philosopher has defined Science as "Tapasaya" and Technology as "Bhog" whereas science is pure selfless and open to all search for truth whereas technology is the exploitation of scientific knowledge for selfish purposes and profit motive and hence it tends to be secretive in this competitive world of today with a throat cutting environment. Sustainable Development is a process in which development can be sustained for generations. It also focuses attention on inter-generational fairness in the exploitation of development opportunities while social development is a function of technological advancement, and also the technological advancement, in turn is a function of scientific know how for a streamlined development of the society [1]. Technology is one of the crucial determinants of sustainable development. Technological import through collaborations has been one of the most important sources of technological inputs for Indian conditions. The use of technologies originating in rich countries often ten to create many social, ecological and resource problems in poor countries. The exploitation of the vast natural resources through progressive development of science, engineering and technology that has brought about the vast changes in the civilization and society from the stone age to the present high technology era. In

Corresponding author: R.S. Mishra
Email Address: hod.mechanical.rsm@dtu.ac.in
https://doi.org/10.36037/IJREI.2021.5202

facts, the mad race for industrialization and economic development has resulted in over exploitation of natural resources, leading to a situation where the two worlds of mankind- the biosphere, lithosphere and hydrosphere of his inheritance and the techno sphere of his creation, are out of balance with each other, indeed on a collision path. To facilitate optimal utilization of finite natural resources for ensuring a sustainable benefit steam for better quality of life on the one hand and to simultaneously keep in mind the conservation of natural resources on the other hand, it is essential that the technology conservation process must be made as efficient as possible. Therefore, sustainable economic development depends on the careful choice of technologies and judicious management of resources for productive activities to satisfy the changing human needs without degrading the environment or depleting the natural resources base.

Refrigeration technology based on the principle of rejection of heat to the surrounding at higher temperature and absorption of heat at low temperature evaporator [1], expansion valve, condenser and compressor are the main four components of single stage vapour compression system. Vapour compression refrigeration systems consume large amount of electricity. This difficulty can be removed by improve the performance parameters of system. Coefficient of performance and exergetic efficiency are main two parameters to calculate the performance of refrigeration systems [2]

Coefficient of performance can be enhanced either by minimizing power consumption of compressor or increasing of refrigeration effect. Refrigeration effect can be increased by adoption of multi-stage throttling. On the other hand, power consumption of compressor can be enhanced by incorporation of multi-stage compression and water coolers/ flash chamber. Collective effect of these two factors improves overall performance of vapour compression system [3] It is presented that irreversibility in system components take place due to large temperature difference between system and surrounding. In order to improve the system performance Irreversibility should be measured in the cycle because Exergy losses are responsible for degradation of system performance Coefficient of performance is commonly used to calculate the performance of vapour compression system but COP provides no information regarding thermodynamic losses in the system components. Using exergy analysis one can be quantify the exergy losses in vapour compression refrigeration systems. Exergy losses increase with increasing of temperature difference between systems and surrounding. Exergy is the available or useful energy and loss of energy means loss of exergy in the system. Exergy losses are useful to improve the performance of system and better utilization of energy input given to the system which is beneficial for environmental conditions and economics of energy technologies. Utilization of green energy can be increased by this method [4]. In past decades, refrigerants such as R12, R12, R22 etc. are used in vapour compression refrigeration system responsible for increasing of global warming and ozone depletion potential. An international society named Montreal protocol discussed and signed on the refrigerants having higher global warming and ozone depletion potential values for all countries. In order to control the emission of greenhouse gases one more committee was formed named as Kyoto protocol [5].

After 90's a program was ran to phase out the higher GWP and ODP refrigerants (CFC and HCFC) for the purpose of environmental problems. To replace "old" refrigerants with "new" refrigerants lots of researches has been carried out [6]. Selladurai and Saravanakumar [7] evaluated thermal performance parameters such as COP and exergetic efficiency using hydrocarbons mixtures on a domestic refrigerator, which was designed to work with HFC-134a and observed that the thermodynamic performance is higher as compared to R134a. and found condenser, expansion valve and evaporator showing lower exergy destruction compared to compressor. Reddy et al. [8] carried out thermodynamic analysis of vapour compression refrigeration system using HFC refrigerants and computed the effect of HFC refrigerants on coefficient of performance and second law efficiency. Also found the effect of the variation of superheating of evaporator outlet, evaporator temperature and degree of sub cooling at condenser outlet, vapour liquid heat exchanger effectiveness and condenser temperature on COP and exergetic efficiency they observed that the R134a and R407C show highest and lowest performance in all respect. Kumar et al. [9] carried out energy and exergy analysis of single stage vapour compression refrigeration system using R11 and R12 as working fluids and computed COP, exergetic efficiency and exergy losses in different components. Nikolaidis and Probert [10] study, effect of condenser and evaporator temperatures on two-stage vapour compression refrigeration system using R22 was studied and suggested that there is requirement to optimize the condenser and evaporator conditions. Many researchers carried out researches on different proportion of hydrocarbons as working fluid in vapour compression refrigeration systems.

Based on the literature it was observed that researchers have gone through detailed first law analysis in terms of coefficient of performance and second law analysis in term of exergetic efficiency of simple vapour compression refrigeration system with single evaporator. Researchers did not go through the irreversibility analysis or second law analysis of multiple evaporators systems with multi-stage expansion in vapour compression refrigeration systems. Researchers did not go through irreversibility and second law analysis of multi-stage vapour compression refrigeration systems by using HFO refrigerants. The present works analyze the system in terms of energy and exergy efficiencies and explain the effect of exergy losses on two-stage vapour compression refrigeration system with HFO refrigerants for replacing R134a.

2. Modelling of Two stages Vapour compression refrigeration system

Some mathematical calculations are required to analyze the two-stage vapour compression refrigeration system based on energy

and exergy method. Two stage vapour compression refrigeration system consist of low and high pressure compressor, condenser, evaporator, expansion valves, water-intercooler and flash chamber. Energy and exergy efficiencies are different for different HFO ecofriendly refrigerants for same system. Following assumptions are taken for thermodynamic analysis of the system:

- Temperature and pressure losses are not considered.
- All components are running under steady state conditions.
- Energy and exergy losses due to potential and kinetic energy are neglected.
- Mechanical efficiencies of low and high pressure compressors are assumed to be 80%.

Two stage vapour compression refrigeration system and its P-H plot shown in Fig. 1 and 2 respectively.

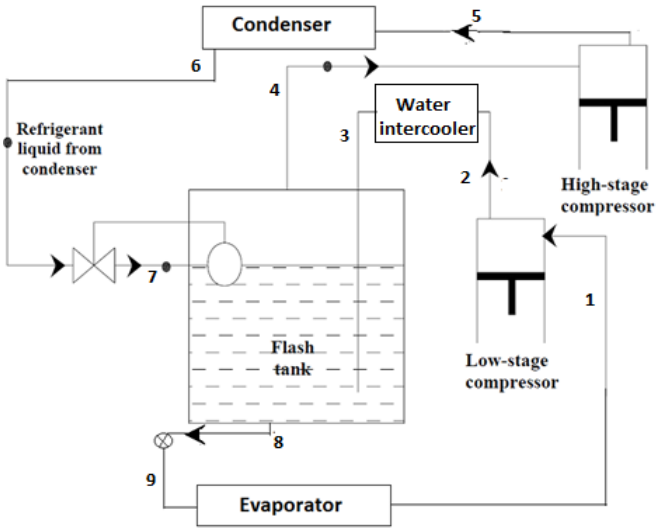


Figure 1: Two-stage vapour compression refrigeration system

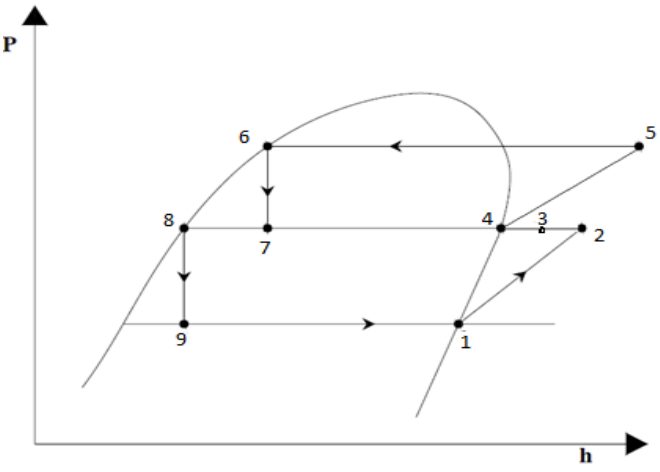


Figure 2: Pressure enthalpy diagram of two-stage VCRS

Exergy energy analysis can be done as follow:

Exergy at any point,

$$EX = (h - h_{amb}) - T_{amb}(s - s_{amb}) \quad (1)$$

For high and low temperature compressors

$$\left(T_{amb}\dot{S}_{gen}\right)_{HP} = \dot{W_{HP}} + \dot{m}_{HP}((h_1 - h_2) - T_{amb}(s_1 - s_2)) \quad (2)$$

$$\left(T_{amb}\dot{S}_{gen}\right)_{LP} = \dot{W_{LP}} + \dot{m}_{LP}((h_4 - h_5) - T_{amb}(s_4 - s_5)) \quad (3)$$

For evaporator

$$\left(T_{amb}\dot{S}_{gen}\right)_{Evap} = \dot{m}_{LP}((h_9 - h_1) - T_{amb}(s_9 - s_1)) + Q\left(1 - \frac{T_{amb}}{T_L}\right) \quad (4)$$

For condenser

$$\left(T_{amb}\dot{S}_{gen}\right)_{Cond} = \dot{m}_{HP}((h_5 - h_6) - T_{amb}(s_5 - s_6)) - Q_{Cond}\left(1 - \frac{T_{amb}}{T_H}\right) \quad (5)$$

For expansion valves

$$\left(T_{amb}\dot{S}_{gen}\right)_{EV} = \dot{m}_{HP}((h_6 - h_7) - T_{amb}(s_6 - s_7)) + \dot{m}_{LP}((h_8 - h_9) - T_{amb}(s_8 - s_9)) \quad (6)$$

For water-intercooler

$$\left(T_{amb}\dot{S}_{gen}\right)_{WI} = \dot{m}_{LP}((h_2 - h_3) - T_{amb}(s_2 - s_3)) \quad (7)$$

For Water Intercooler

$$\left(T_{amb}\dot{S}_{gen}\right)_{FC} = \dot{m}_{LP}((h_3 - h_4) - T_{amb}(s_3 - s_4)) + (\dot{m}_{HP}h_7 - \dot{m}_{LP}h_8) - T_{amb}(\dot{m}_{HP}s_7 - \dot{m}_{LP}s_8) \quad (8)$$

Exergetic efficiency

$$\eta_{ex} = \frac{Q\left(1 - \frac{T_{amb}}{T_L}\right)}{W_{LP} + W_{HP}} \quad (9)$$

Coefficient of Performance (COP)

$$COP = \frac{\dot{m}_{LP}((h_1 - h_9)}{W_{LP} + W_{HP}}$$

Total entropy generation

$$\left(T_{amb}\dot{S}_{gen}\right)_{TOTAL} =$$

$$\sum\left(T_o\dot{S}_{gen}\right)_{HP} + \left(T_o\dot{S}_{gen}\right)_{LP} + \left(T_o\dot{S}_{gen}\right)_{Evap} + \left(T_o\dot{S}_{gen}\right)_{Cond} + \left(T_o\dot{S}_{gen}\right)_{EV} + \left(T_o\dot{S}_{gen}\right)_{WI} + \left(T_o\dot{S}_{gen}\right)_{FC} \quad (10)$$

Exergy Destruction Ratio (EDR)

$$\eta_{ex} = \frac{1}{1+EDR} \quad (11)$$

Where

Sustainability index (SI)

$$SI = \frac{1}{1-\eta_{ex}} \quad (12)$$

3. Results and Discussion

3.1 Performances of vapour compression refrigeration system using HFO refrigerants

Thermodynamic (energy-exergy) performances of two stage vapour compression refrigeration system using HFO refrigerants is shown in Table-1 respectively. As cleared that the first law efficiency (i.e. coefficient of performance) of two stages vapour compression refrigeration system is of R1234ze(Z) is 9.327% higher than using HFC-134a and the second law efficiency (i.e. exergetic efficiency) using R1234ze(Z) is 8.087% higher than using HFC-134a respectively. It means that R134a consumes more electricity than R1234ze(Z)

Although first law efficiency (i.e. coefficient of performance) using R1233zd(E) is 7.35%, using R1336mzz(Z) is 1.64% higher than using HFC-134a. Similarly, first law efficiency (i.e. coefficient of performance) using R1234yf is 9.1%, using R1243zf is 1.75% and using R1225ye(z) is 3.3% lower than using HFC-134a. The second law efficiency using R1234ze(Z) is 8.087%, R1233zd(E) is 5.88% higher than using HFC-134a respectively

Similarly, second law efficiency using R1234yf is -8.44%, using R1243zf is 1.75% and R-1234ze(E) is-3.25 and R1225ye(z) is 3.037% lower than using HFC-134a. Similarly, ambient condition plays an important role in electricity consumption of vapour compression refrigeration systems because higher the temperature difference between system and surrounding, results higher will be compressor work that's why COP of vapour compression refrigeration system increase with increase in evaporator temperature and decrease with decrease in evaporator temperature.

Table-1 Effect of system performance parameters using ecofriendly HFO refrigerants on the thermodynamic performances of vapour compression refrigeration systems for isentropic compressors efficiencies =80%. at condenser temperature $T_{_Cond}$ =40°C and evaporator temperature -40°C, and evaporator load =35kW, ambient temperature=25°C

Performance Parameters	R1234 ze (Z)	R1234 ze(E)	R1225 ye (Z)	R1224 yd(Z)	R1233 zd (E)	R1336 mzz(Z)	R1243 zf	R1234 yf
First law Efficiency ($COP_{_Actual}$)	1.934	1.720	1.71	1.866	1.899	1.798	1.738	1.608
Work required to run HPCompressor (W_{CompHP}) "kW"	8.159	12.01	12.1	12.01	7.477	5.836	11.94	13.52
Work required to run LPCompressor (W_{CompLP})"kW"	9.942	8.333	8.372	8.333	10.95	13.63	8.194	8.246
Work required to run whole system (W_{Comp}) "kW"	18.1	20.35	20.47	20.35	18.43	19.47	20.14	21.77
Condenser Heat "kW"	54.47	61.5	61.43	56.83	55.21	57.77	60.21	64.69
Evaporator load" kW"	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0
HP Mass flow rate (kg/sec)	0.267	0.3891	0.4568	0.360	0.2956	0.359	0.344	0.4745
LP Mass flow rate (kg/sec)	0.188	0.2213	0.258	0.246	0.2114	0.267	0.1982	0.2485
Total System Exergy Destruction ($ExD_{_Total}$) (%)	45.2	50.95	50.84	47.05	46.32	49.43	50.06	53.58
Exergy Destruction in condenser ($ExD_{_Cond}$) (%)	14.93	15.0	15.02	14.91	14.71	14.48	15.03	14.9
Exergy Destruction in compressor ($ExD_{_comp}$)(%)	19.72	20.41	20.5	20.16	19.89	20.09	20.32	20.7
Exergy Destruction in HP compressor(ExD_{CompHP} (%)	8.583	11.46	11.54	8.43	7.793	5.753	11.49	12.22
Exergy Destruction in LP compressor($ExD_{_Comp_LP}$%)	11.14	8.954	8.956	11.73	12.1	14.34	8.83	8.475
Exergy Destruction in Valve $ExD_{_Valve}$ (%)	10.67	15.0	15.25	12.25	11.88	15.17	14.74	17.6
Exergy Destruction in HP Valve ($ExD_{_Valve_HP}$(%)	6.372	9.688	10.03	4.46	3.647	2.271	9.824	12.31
Exergy Destruction in LP Valve ($ExD_{_Valve_LP}$ (%)	4.298	5.317	5.224	7.791	8.23	12.09	4.916	5.287
Exergy Destruction in Evaporator ($ExD_{_Eva_}$ (%)	0.16	0.5182	0.1465	0.1161	0.1578	0.1507	0.1481	0.135
Exergy Destruction in Water intercooler ExD_{Wic}(%)	0.272	0.0181	0.0791	0.3928	0.3177	0.4538	0.1792	0.2526
Rational System Efficiency(%)	54.80	49.05	49.16	52.95	53.68	50.57	49.94	46.42

3.2 Performances of vapour compression refrigeration system using HFC refrigerants

Thermodynamic (energy-exergy) Performances of two stage vapour compression refrigeration system using HFO refrigerants is shown in Table-2 respectively. As cleared that the first law efficiency (i.e. coefficient of performance) of two stages vapour compression refrigeration system is of R717is 12.38% higher than using HFC-134a.and the second law efficiency using R717 is 10.592% higher than using HFC-134a respectively. It means that R134a consumes more electricity than R717. Although first law efficiency (i.e. coefficient of performance) using R152a is

7.066%, using R245fa is 5.879%, using R32 is 3.73%, higher than using HFC-134a. Similarly, first law efficiency using R125 is 19.2%lower than using HFC-134a. The second law efficiency using R152a is 6.509%, R245fa is 4.674%, and R32 is 2.978%, higher than using HFC-134a respectively Similarly second law efficiency using R1234yf is 18.44, % lower than using HFC-134a.

Table-2 Effect of system performance parameters using ecofriendly refrigerants on the thermodynamic performance of vapour compression refrigeration systems for isentropic compressors efficiencies =80%. at condenser temperature $T_{_Cond}$ =40°C and evaporator temperature -40°C, and evaporator load =35kW, ambient temperature=25°C

Performance Parameters	R717	R134a	R152a	R245fa	R32a	R125
First law Efficiency (COP_ Actual)	1.988	1.769	1.894	1.873	1.835	1.429
HP Mass flow rate (kg/sec)	0.03703	0.3342	0.1898	0.3079	0.1746	0.6815
LP Mass flow rate (kg/sec)	0.02854	0.1967	0.1242	0.2103	0.1129	0.2974
Work required to run Compressor W_{CompLP}"kW"	7.933	8.04	8.052	10.56	7.762	8.055
Work required to run whole systemW_{Comp}"kW"	17.6	19.79	18.48	18.68	19.07	24.5
Condenser Heat _ HTC "kW"	49.16	58.71	54.69	56.51	53.0	68.33
Evaporator load"kW"	35.0	35.0	35.0	35.0	35.0	35.0
Total System Exergy Destruction ($ExD_{_Total}$) (%)	43.93	49.34	46.0	46.93	47.79	58.65
Exergy Destruction in condenser ($ExD_{_Cond}$) (%)	18.51	15.13	15.55	14.78	17.33	14.21
Exergy Destruction in compressor ($ExD_{_comp}$) (%)	16.32	20.01	19.31	20.04	18.04	20.45
Exergy Destruction in HP compressor(ExD_{CompHP}) (%)	8.633	11.32	10.42	8.395	10.02	13.06
Exergy Destruction in LP compressor ($ExD_{_Comp_LP}$)(%)	2.482	8.699	8.886	11.68	7.832	7.392
Exergy Destruction in Valve $ExD_{_Valve}$ (%)	6.574	14.25	11.22	12.26	11.92	23.55
Exergy Destruction in HP Valve ($ExD_{_Valve_HP}$(%)	4.092	9.52	7.099	4.444	8.345	18.37
Exergy Destruction in LP Valve ($ExD_{_Valve_LP}$ (%)	2.482	4.727	4.119	7.813	3.48	5.18
Exergy Destruction in Evaporator ($ExD_{_Eva_}$ (%)	0.1571	0.1630	0.1651	0.2063	0.1522	0.1065
Exergy Destruction in Water intercooler ExD_{Wic}(%)	2.371	0.2586	0.2339	0.3877	0.3619	0.3304
Rational System Efficiency(%)	56.07	50.7	54.0	53.07	52.21	41.35

Thermodynamic (energy-exergy) performances of two stage vapour compression refrigeration system using HFO refrigerants is shown in Table-3 respectively. As cleared that the first law efficiency of two stages vapour compression refrigeration system is using R236fa is 2.78%, using R227ea is 15.77% and using R404a is 12.83%, using R407c is 8.3% and R143a is 9.723% lower than using HFC-134a. The second law efficiency using R236fa is 3.826%, R227ea is 14.67% R404a is 11.95% and R-410a is2.94%, R407c is 8.126%, R507a is 12.367% and R143a is 9.171% lower than using HFC-134a respectively.

Table-3 Effect of system performance parameters using ecofriendly refrigerants on the thermodynamic performances of vapour compression refrigeration systems for isentropic compressors efficiencies =80%. at condenser temperature $T_{_Cond}$ =40°C and evaporator temperature -40°C and evaporator load =35kW, ambient temperature=25°C

Performance Parameters	R236fa	R227ea	R404a	R410a	R407c	R507a	R143a
First law Efficiency (COP_ Actual)	1.701	1.49	1.542	1.718	1.622	1.53	1.597
Work required to run Compressor W_{CompLP}"kW"	11.18	14.58	14.68	12.63	13.29	14.78	13.99
Work required to run whole systemW_{Comp}"kW"	9.395	8.904	8.013	7.745	8.292	7.98	7.929
Condenser Heat _ HTC "kW"	62.7	70.29	64.24	57.27	58.66	64.68	62.69
Evaporator load"kW"	35.0	35.0	35.0	35.0	35.0	35.0	35.0
Total System Exergy Destruction ($ExD_{_Total}$) (%)	51.24	56.74	55.36	50.79	53.42	55.57	53.95
Exergy Destruction in condenser ($ExD_{_Cond}$) (%)	15.15	15.07	14.82	16.02	16.96	14.67	14.92
Exergy Destruction in compressor ($ExD_{_comp}$) (%)	20.69	23.49	20.25	19.15	19.26	20.26	20.08
Exergy Destruction in HP compressor(ExD_{CompHP}) (%)	10.63	14.58	12.45	11.29	11.27	12.51	12.19
Exergy Destruction in LP compressor ($ExD_{_Comp_LP}$)(%)	10.07	8.904	7.797	7.862	8.01	7.75	7.895
Exergy Destruction in Valve $ExD_{_Valve}$ (%)	15.17	19.69	20.04	15.77	14.2	20.05	18.94
Exergy Destruction in HP Valve ($ExD_{_Valve_HP}$(%)	8.273	13.29	15.1	11.62	10.41	15.46	14.12
Exergy Destruction in LP Valve ($ExD_{_Valve_LP}$ (%)	6.899	6.396	4.941	4.154	3.784	5.04	4.828
Exergy Destruction in Evaporator ($ExD_{_Eva_}$ (%)	0.1492	0.017	0.2668	0.1229	3.272	0.115	0.1298
Exergy Destruction in Water intercooler ExD_{Wic}(%)	0.0747	1.04	0.0187	0.279	0.2769	0.032	0.1251
Rational System Efficiency(%)	48.76	43.26	44.64	49.21	46.58	44.43	46.05

Thermodynamic (energy-exergy) performances of two stage vapour compression refrigeration system using HFO refrigerants is shown in Table-4 respectively. As cleared that the first law efficiency of two stages vapour compression refrigeration

system is of R123 is 9.158% higher than using HFC-134a.and the second law efficiency using R123 is 6.9 % higher than using HFC-134a respectively. It means that R134a consumes more electricity than R123. Although first law efficiency using R600a is 0.8087%, using R600a is 0.848% and using R141b is 13.6% higher than using HFC-134a. Similarly, first law efficiency using R1234yf is -2.205%, lower than using HFC-134a. The second law efficiency using R141b is 11.08%, higher than using HFC-134a respectively Similarly second law efficiency using R290 is 1.93%, using R124 is 0.1% lower than using HFC-134a.

Table-4 Effect of system performance parameters using ecofriendly refrigerants on the thermodynamic performance of two stages vapour compression refrigeration systems for isentropic compressors efficiencies =80%. at condenser temperature $T_{_Cond}$ =40°C and evaporator temperature -40°C, ambient temperature=25°C and evaporator load =35kW

Performance Parameters	R123	R124	R600a	R290	R141b
First law Efficiency ($COP_{_Actual}$)	1.931	1.78	1.784	1.73	2.01
Work required to run Compressor $W_{Comp\ HP}$ "kW"	6.034	10.91	11.08	12.34	5.029
Work required to run Compressor $W_{Comp\ LP}$ "kW"	12.81	8.49	8.804	7.898	12.39
Work required to run whole systemW_{Comp}"kW"	18.12	19.67	19.62	20.23	17.42
Condenser Heat $__{HTC}$ "kW"	53.33	59.81	59.85	59.74	50.2
Evaporator load"kW"	35.0	35.0	35.0	35.0	35.0
HP Mass flow rate (kg/sec)	0.3175	0.425	0.1817	0.1803	0.2224
LP Mass flow rate (kg/sec)	0.2425	0.2559	0.116	0.1038	0.1813
Exergy Destruction in condenser ($ExD_{_Cond}$) (%)	14.37	15.09	15.16	15.09	14.1
Exergy Destruction in compressor ($ExD_{_comp}$) (%)	19.38	20.31	20.49	20.08	18.75
Exergy Destruction in HP compressor(ExD_{CompHP}) (%)	6.322	10.88	10.79	11.66	5.438
Exergy Destruction in LP compressor ($ExD_{_Comp_LP}$)(%)	13.05	9.435	9.695	8.419	13.31
Exergy Destruction in Valve $ExD_{_Valve}$ (%)	11.79	13.63	13.21	15.18	9.758
Exergy Destruction in HP Valve ($ExD_{_Valve_HP}$(%)	2.472	8.141	7.719	10.63	1.533
Exergy Destruction in LP Valve ($ExD_{_Valve_LP}$ (%)	9.313	5.485	5.553	4.545	8.225
Exergy Destruction in Evaporator ($ExD_{_Eva_}$ (%)	0.3251	0.4717	0.1526	0.1499	0.1678
Exergy Destruction in Water intercooler ExD_{Wic}(%)	0.0605	0.1526	0.1826	0.2217	0.9072
Rational System Efficiency(%)	54.2	50.65	51.11	49.72	56.32

Table-5 shows the improvement in two stage vapour compression refrigerants using ecofriendly refrigerants as compared with the thermodynamic performances using R134a and it was found that R1234ze(Z) gives best thermodynamic performances and R1234yf gives lowest thermodynamic performances. It was also found that second law performances using R1336mzz(z) and R124 is nearly same but R-600a gives slightly higher but less than 1%.

Table-5: Improvement in first and second law efficiency using Ecofriendly refrigerants as compared to R134a Rational System Efficiency

Eco-friendly refrigerants	Improvement in First Law Efficiency (%)	Improvement in Rational System Efficiency (%)
R1234ze (Z)	9.327	8.087
R1224 yd(Z)	5.483	4.438
R1233zd (E)	7.35	5.88
R1336 mzz(Z)	1.64	-0.256
R717	12.38	10.592
R152a	7.066	6.509
R245fa	5.879	4.674
R123	9.158	6.903
R124	0.622	-0.099
R600a	0.848	0.8087
R141b	13.6	11.085
R245fa	5.879	4.674
R32a	3.73	2.978

Table-6 shows the improvement in two stage vapour compression refrigerants using ecofriendly refrigerants as compared with the thermodynamic performances using R134a and it was found that R125 gives lowest thermodynamic performances

Table-6: Decrement in first and second law efficiency using Ecofriendly refrigerants as compared to R134a Rational System Efficiency

Eco-friendly refrigerants	Improvement in First law Efficiency(%)	Improvement in Rational System Efficiency(%)
R1234ze(E)	-2.77	-3.254
R1225ye (Z)	-3.335	-3.037
R1243zf	-1.752	-1.499
R1234yf	-9.10	-8.442
R125	-19.22	-18.44
R236fa	-2.78	-3.826
R227ea	-15.77	-14.67
R404a	-12.83	-11.95
R410a	-2.883	-2.94
R407c	-8.31	-8.126
R507a	-13.51	-12.367
R143a	-9.723	-9.171
R290	-2.205	-1.933

The output parameters computed are compressors work as exergy input ($W_{_Comp_HP}$ and $W_{_Comp_LP}$), and Compressor and the coefficient of performance (COP) cycle and energy performance of two stages vapour compression refrigeration

systems ($COP_{\text{two stages}}$) have been carried out. The second law performances (exergetic efficiency and exergy destruction ratios) of two stages vapour compression refrigeration system have been computed Results showed that R1234ze (Z) and R1233zd(E) gives better thermodynamic performances. Than using HFC -134a. However, R1224yd (Z) of seven GWP overcome favourable thermodynamic (energy and exergy) performances, above -10°C of evaporator can suitably replace R134a. Similarly, HFO-1336mzz(Z) also gives better energy-exergy performances as compared to HFC-134a. However, R1234yf although gives 4% to 10% lesser thermodynamic (energy and exergy) performances than using high GWP ecofriendly HFC-134a refrigerant. Therefore, these ultra-low GWP ecofriendly HFO refrigerants can serve as new alternative refrigerants for replacing HFC-134a in the vapour compression refrigeration systems for a sustainable environment.

3. Conclusions

Exergetic analysis of two stage refrigeration system was carried out with different HFO refrigerants and following conclusion and recommendation are presented below:

- R1234ze (Z) and R1233zd(E) gives better thermodynamic performances. Than using HFC -134a. However, R1224yd (Z) of seven GWP overcome favorable thermodynamic (energy and exergy) performances, above -10°C of evaporator can suitably replace R134a.
- Similarly, HFO-1336mzz(Z) also gives better energy-exergy performances as compared to HFC-134a. However, R1234yf although gives 4% to 10% lesser thermodynamic (energy and exergy) performances than using high GWP ecofriendly HFC-134a refrigerant
- R1234yf shows lowest thermodynamic (energy and exergetic performances) among selected HFO refrigerants
- Exergy destruction for R1234yf is higher than R1234ze(Z).
- Exergetic and energetic efficiency of R152a is highest among selected refrigerants.
- Compressors responsible for highest exergy destruction for all refrigerants taken under consideration.
- Therefore, these ultra-low GWP ecofriendly HFO refrigerants can serve as new alternative refrigerants for replacing HFC-134a in the vapour compression refrigeration systems for a sustainable environment
- Exergetic analysis of two stage refrigeration system was carried out with different HFO refrigerants and found that the R1234yf shows lowest thermodynamic (energy and exergetic performances) among selected HFO refrigerants
- Exergy destruction for R1234yf is higher than R1234ze(Z)
- Compressors responsible for highest exergy destruction for all refrigerants taken under consideration.
- Lowest exergy destruction for all refrigerants found in evaporator for all considered ecofriendly refrigerants.

References

[1] R.S. Mishra [2014] Vapour compression refrigeration technology for sustainable development, International Journal of Advance Research and Innovation, ISSN 2347 - 3258 , Volume-3, Issue 4 (2015) 647-652.

[2] Kapil Chopra, V.Sahni, R.S Mishra. Thermodynamic analyses of multiple evaporators vapour compression refrigeration systems with R410A, R290, R1234YF, R502, R404A and R152A.International Journal of Air-conditioning and Refrigeration 21(1) (2014) 1-14

[3] Kapil Chopra, V.Sahni , R.S Mishra -Thermodynamic analysis of a multi-evaporators vapour compression refrigeration system using ecofriendly refrigerants (R1234yf, R1234ze and R134a). International Conference on Smart Technologies for Mechanical Engineers.2013.Paper No-42.

[4] Kapil Chopra, V.Sahni , R.S Mishra - Energy and exergy analysis of multiple evaporators and compressors with individual or multiple expansion valves vapour compression refrigeration system using ecofriendly refrigerants. International Conference on Smart Technologies for Mechanical Engineers.2013.Paper No-121.

[5] Johnson. E. Global warming from HFC. Environ. Impact Asses.1998; 18:485-492.

[6] QiyuChen,R.C Prasad. Simulation of a vapour compression refrigeration cycles HFC134A and CFC12.Int Comm.Heat Mass Transfer.1999; 26:513-521.

[7] R.Saravanakumar, V.Selladurai-Exergy analysis of a domestic refrigerator using eco-friendly R290/R600a refrigerant mixture as an alternative to R134a.Int J Therm Anal Calorim.2013;

[8] V. Siva Reddy, N.L Panwar, S.C Kaushik-Exergy analysis of a vapour compression refrigeration system with R134a, R143a, R152a, R404A, R407C, R410A, R502 and R507A.Clean Techn Environ Policy. 2012; 14: 47-53.

[9] S.Kumar, M.Prevost, R.Bugarel- Exergy analysis of a vapour compression refrigeration system. Heat Recovery Systems & CHP.1989;9:151-157.

[10] C. Nikolaidis, D. Probert-Exergy method analysis of a two-stage vapour-compression refrigeration-plants performance. Int J Applied Thermal Engineering.1998; 60:241-256.

[11] R.S. Mishra Thermal Analysis & Optimization of Two Stages Vapour Compression Refrigeration Systems using Sixteen Ecofriendly Refrigerants, International Journal of Advance Research and Innovation , Volume 3, Issue 1 (2015) 88-93 ISSN 2347 – 3258.

Cite this article as: R.S. Mishra, Thermodynamic exergy analysis of two-stage vapour compression refrigeration systems using HFO refrigerants for replacing R134a, International journal of research in engineering and innovation (IJREI), vol 5, issue 2 (2020), 96-102. *https://doi.org/10.36037/IJREI.2021.5202*.

Trajectory Tracking of a 2-DOF Helicopter System using Fuzzy Controller Approach

Abhishek Chaudhary
Electrical Engineering Department
Delhi Technological University, DTU
New Delhi, India
abhishek@dtu.ac.in

Bharat Bhushan
Electrical Engineering Department
Delhi Technological University, DTU
New Delhi, India
bharat@dce.ac.in

2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI) | 978-1-6654-1559-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICETCI51973.2021.9574049

***Abstract*—In this paper, the desired pitch and yaw axis trajectories of a 2-DOF helicopter system are tracked. For this purpose, fuzzy system is implemented and its performance is compared with a conventional control approach, namely an LQR controller. In proposed fuzzy system, in order to obtain an optimal fuzzy model for benchmark Unmanned Air Vehicle (2-DOF Helicopter), the parameters of the membership functions at the antecedent part of the fuzzy rules are tuned.**

Keywords— 2-DOF Helicopter; Unmanned Air Vehicles (UAVs); Fuzzy Logic Rules; LQR; trajectory control; pitch angle; yaw angle.

I. INTRODUCTION

In recent years, the controllers of unmanned air vehicles (UAVs) is one emerging research area. Helicopter is one of the most commonly used vehicle in this area. Helicopters are mostly used in fields like transportation, traffic control, firefighting, air-sea rescue, etc. [1]. Helicopters are benchmark nonlinear systems due to their instability and cross-coupling effects between their yaw and pitch axes, which makes their position and trajectory control a difficult task to be handled.

In the literature, the control of trajectory tracking in helicopter systems is still a significant problem [2]–[4]. In [2], by using two controllers; a feedforward controller for trajectory tracking and a feedback controller for the stabilization of the trajectory tracking error, the trajectory tracking control of a 3-DOF (Degrees-of-Freedom) helicopter is realized. The results obtained illustrate the good tracking performance of the controllers when designed controller algorithm is tested during experiment. In [3], for a small scale unmanned helicopter a heading reference trajectory and an inertial position tracking are studied. The helicopter dynamics are composed in two subsystems namely; heading-heave movement and longitudinal-lateral of the system and these subsystems are controlled by feedback loops for trajectory tracking errors. Performance of this proposed controller is tested by using a realistic simulator, namely X-Plane flight simulator and the simulation results show the effectiveness of approach proposed. In [4], a helicopter is stabilized based on visual simultaneous mapping and localization. The attitude and the position control of the helicopter is achieved by using parallel tracking and mapping (PTAM) method by using wireless camera and the results shows the efficiency of this proposed approached.

Linear quadratic regulator (LQR) was reformed along with PID to diminish the tracking error in the pitch and yaw output by enhancing the quadratic function using the state feedback law [5]. A gradient descent algorithm was proposed in achieving the design of robust adaptive fuzzy controller [6]. Various other control techniques like model predictive control [7], adaptive disturbance rejection controller [8] and adaptive second-order sliding mode controller [9] have also been used by different researchers. During the real time operations of helicopter system, various failure mechanisms occurs that affects the performance of system and causes instability. To regulate the higher order flight control actuator failures, a reorganized FTC system has been considered [10].

Fuzzy logic is a commonly used soft computing method which is applied in various fields such as: pattern recognition, telecommunication, control etc. Since the introduction of fuzzy sets by Prof. Lotfi A. Zadeh in 1965, fuzzy has been rapidly developed to a promising research tool. By using linguistic rules, knowledge and human spirit can be integrated into any fuzzy logic system. With the support of fuzzy set philosophy and mathematics, a fuzzy logic system can perform thinking according to the labelled linguistic rules. Fuzzy control is a warm research topic, which has strained a great deal of attention from researchers working from fundamental research to industrial applications in the recent time. It has been perceived that fuzzy control has advanced from model-free approach (Mamdani, 1993) to model-based approach.

A dual fuzzy-neural network control is experimented in an intelligent landing system of an aircraft used in air transportation. The simulation studies of the same shows that the considered approach works quite well under the ambiguous and changing weather conditions [11]. In [12], the flight control of an UAV is examined. For this purpose, three ANFIS (Adaptive Neuro Fuzzy Interface System) controllers are deployed for the three axes; pitch, yaw, and roll in order to control the speed, the altitude and the heading of the UAV. It is observed that the employment of the method is able to meet the anticipated specifications.

The developed approach addressed the issues in configuring the control laws of fuzzy set rules. The results demonstrated the boundedness of all the signals in the system and the asymptotical convergence of the tracking error of pitch angle and yaw angle at minimum value. The proposed approach has a self-organizing

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:54:18 UTC from IEEE Xplore. Restrictions apply.

capability and it is applied in documentation of nonlinear dynamic systems. The obtained simulation results indicate high tracking performance and fast response of the control system. In this study, the fuzzy logic systems is used. The aim is to design a fuzzy system in order to track the desired trajectories of a 2-DOF helicopter. The conventional approaches in the training of fuzzy systems is to tune the parameters of the antecedent member- ship functions are used. These operators are very simple and they do not have the associativity or commutativity properties [13]–[15]. In [16], an intelligent technique is used to control the non-linear coupling between two rotor motors of the 2-DOF Helicopter system using PID and Fuzzy controller. The results obtained are compared with a conventional control approach LQR.

This study is planned as follows; in section II, the mathematical portrayal of the 2-DOF helicopter system is given. In section III, fuzzy logic is discussed. In section IV, the simulation results are given. In section V, the obtained results are analyzed.

II. DESCRIPTION OF THE 2-DOF HELICOPTER SYSTEM

Fuzzy Controller with Triangular and Gaussian Membership functions and a conventional control approach, LQR are deployed on the 2-DOF helicopter shown in Fig. 1. It is a extremely nonlinear multi-input-multi-output (MIMO) system with robust cross-couplings between the pitch and the yaw axes due to the drive torques performing on the axes. Helicopter is attached on a fixed base and its yaw and pitch (the back and the front) propellers are energetic by two DC motors. The front propeller handles the vertical motion of the 2-DOF helicopter system about its pitch axis. This angle is well-defined as positive when the front propeller sources a motion in the upward direction. In case the propeller motors are not excited, i.e. when the helicopter is at resting position, the pitch angle is about $-40.5^0$and its gesture is limited between $-40.5^0$and 40.5^0. The back propeller handles the horizontal motion about the yaw axis.

The 2-DOF helicopter is able to freely rotate all around 360^0in the yaw axis. The yaw angle is well-defined as negative in the anti-clockwise and positive in the clockwise direction. The two thrust forces are Fp and Fy shown in Fig.1, produced at the distances r_p and r_y from the pitch axis and the yaw axis respectively [17]. The voltages provided to the front propeller and the back propeller motors are the inputs of the given helicopter system and the pitch (θ) and the yaw (Ψ) angles in radians are the two outputs of this benchmark 2-DOF helicopter system. The aim is to propose a controller to track down the desired trajectories in the pitch axis and the yaw axis of the system. The dynamic nonlinear equations of the system acquired by using Lagrangian mechanics are shown for the pitch and the yaw axes as follows [17]:

Non-Linear Equation for pitch axis:

$$(J_{eq,p} + m_{heli}\, l_{cm}^2)\ddot{\theta} = K_{pp}\, V_{m,p} + K_{py}\, V_{m,y} - m_{heli} g l_{cm} \cos\theta - B_p\, \dot{\theta} - m_{heli}\, l_{cm}^2 sin\theta cos\theta \dot{\Psi}^2$$

Non-Linear Equation for yaw axis:

$$(J_{eq,y} + m_{heli}\, l_{cm}^2 cos\theta^2)\ddot{\Psi} = K_{yp}\, V_{m,p} + K_{yy}\, V_{m,y} - B_y \dot{\Psi} + 2m_{heli}\, l_{cm}^2 sin\theta cos\theta \dot{\Psi}\dot{\theta}$$

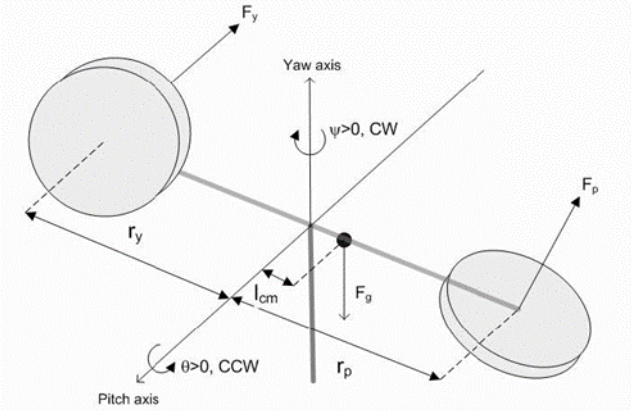


Fig. 1. Simple free body diagram of 2-DOF helicopter [17]

The thrust forces acting on the pitch axis and the yaw axis from the front motor and back motor are then explained. Using the Euler-Lagrange formula, the nonlinear equations of motion of the 2 DOF Helicopter system are obtained. These equations are linearized about zero and the linear state-space model to find the position dynamics of the system. Given the state-space representation

$$\dot{x} = Ax + Bu$$

state vector for the 2 DOF Helicopter is:

$$y = Cx + Du$$

$$x^T = [\theta(t), \psi(t)]$$

output vector is:

$$y^T = [\theta(t),\ \psi(t)]$$

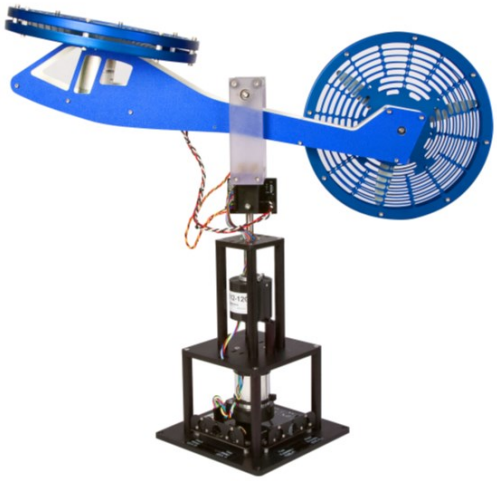


Fig 2 Quanser 2-DOF Helicopter [17]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:54:18 UTC from IEEE Xplore. Restrictions apply.

The Laboratory model view of Quanser 2-DOF Helicopter which is having all these above mentioned specifications is shown in Figure2 and the description of all non-linear parameters is given in Table 1.

TABLE 1 : Non-linear model parameters

	Description	Values
$J_{eq,p}$	The total moment of inertia about pitch axis	$0.0384 kg.m^2$
l_{cm}	Center of mass length along 2-DOF helicopter body from pitch axis	$0.1855m$
$J_{eq,y}$	The total moment of inertia about yaw axis	$0.0431 kg.m^2$
K_{pp}	Thrust torque constant acting on the pitch axis from pitch propeller	$0.2041 N.m/V$
K_{py}	Thrust torque constant acting on the pitch axis from the yaw propeller	$0.0068 N.m/V$
K_{yp}	Thrust torque constant acting on the yaw axis from the pitch propeller	$0.0219 N.m/V$
K_{yy}	Thrust torque constant acting on yaw axis from yaw propeller	$0.072 N.m/V$
g	Gravitational constant	$9.81 m/s^2$
B_p	Viscous damping about the pitch axis	$0.8 N/V$
B_y	Viscous damping about the yaw axis	$0.318 N/V$
m_{heli}	The total moving mass of the helicopter	$1.3872 kg$
$V_{m,p}$	Applied Voltage to the pitch motor	$\pm 24 V$
$V_{m,y}$	Applied Voltage to the yaw motor	$\pm 15 V$

III. FUZZY LOGIC CONTROLLER

Fuzzy logic controllers are able to outline mankind knowledge about a system with the help of membership functions and using IF-THEN type of fuzzy rules. This methods of making any machine behave like a human in a fuzzy manner is coined as fuzzy logic controller. Fuzzy deals with issues such as uncertainty, imprecise and qualitative decision making problems. In the fuzzy application, we need to always remember that it is a range to point control or vice versa. By describing the associated membership functions, output can be extracted by the fuzzification of both output and input. This whole implementation of fuzzy logic controller is carried out in mainly three different steps: - fuzzification, Fuzzy interference process and de-fuzzification, as shown in Figure 3. One measured input and derivative of this measured input is applied to the fuzzy logic controller and then fed to the 2-DOF helicopter and finally, the output returned back to the input using closed loop feedback process of the same unmanned vehicle (2-DOF helicopter).

For the monitoring of a system, a proper human thinking base knowledge is obtained by this approach for manipulation, representation and implementation of fuzzy logic controller. To obtain an ideal fuzzy control model, the conventional approach is to tune the membership function of the fuzzy logic controller. It is one of the arrangements which is more likely to be closer to that of the human thinking and more natural language than that of the available traditional logical systems. The linguistic control strategy is converted into automatic control strategy which is obtained by experiential knowledge.

Fuzzy logic controller validates all the parameters as a pleasing tool, and its reaction can diverge depending upon the nature of membership functions of input-output & also on the different set of rules being used in the FLC logic. For helicopter controlling, T-S fuzzy model is used, and two inputs and one output is provided for pitch angle error as shown in figure 4 as well as for yaw angle error as shown in fig 5. The one input is change in pitch error and the other input is the pitch error. Gaussian type and Triangular type membership functions are chosen in order to have continuous differentiable curve. Seven different linguistic variables have been chosen for every single input state variable named as NS, ZE, PS, PM, NB, NM and PB with appropriate rule base. These are same for the pitch and yaw input functions

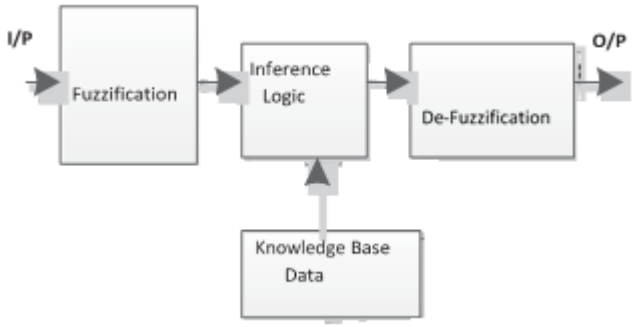


Fig. 3 Fuzzy Logic Controller Architecture

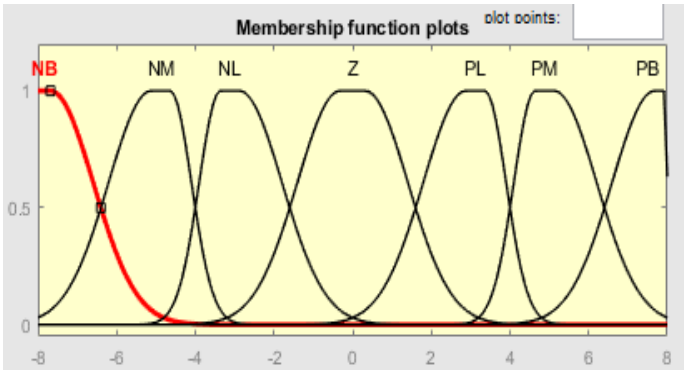


Fig. 4 Membership functions of pitch control

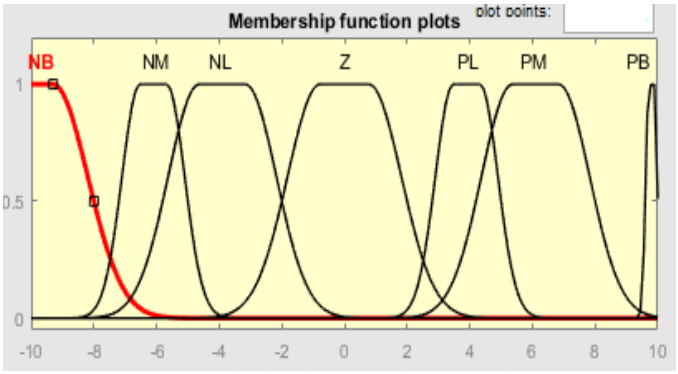


Fig. 5 Membership functions of yaw control

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:54:18 UTC from IEEE Xplore. Restrictions apply.

IV. THE SIMULATION ANALYSIS

2-DOF helicopter system is controlled by implementing fuzzy logic controller to track an anticipated reference trajectory. The 2-DOF helicopter system, a benchmark control experimental setup is designed for laboratory researches, includes one fixed base and two different propellers. The base provides freedom for movement along its center. The propellers are being rotated by dc motors which are fixed separately at both the ends of these propellers. The main objective of these two propellers is to control the elevation angle and the yaw angle of 2-DOF helicopter in the middle of operation. Yaw angle is being controlled by one back end propeller while elevation angle is being controlled by front end propeller. These propellers helps in providing stable side motion.

The controlling challenge here is to deal with the cross coupling effect which is formed by pitch and yaw motion altogether. The block diagram of this Fuzzy system is given in Fig. 6. The 2-DOF helicopter is a benchmark MIMO nonlinear system and two controllers are designed, one for the pitch axis and other for the yaw axis. The error $e_p(t)$ is considered as the difference between the anticipated pitch angle and the actual pitch angle in radians. This error $e_p(t)$ and the derivative of this error $\Delta e_p(t)$ are given as the tow inputs to the fuzzy logic controller for the pitch axis of 2-DOF Helicopter system. Similarly, the error $e_y(t)$ that is considered as the difference between the desired and the actual yaw angle in radians. The error $e_y(t)$ derivative of this error $\Delta e_y(t)$ are given as the two inputs to the fuzzy logic controller for the yaw axis of 2-DOF Helicopter system. Now in our implementation, each controller has two inputs and each input is fuzzified by using different seven Gaussian membership functions. These membership functions are distributed in the closed intervals and have been depicted in Fig. 4 and Fig.5 for fuzzy logic controllers.

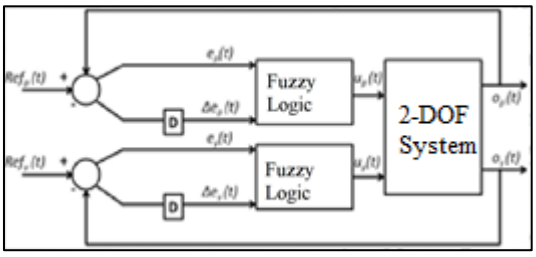


Fig.6 Block Diagram of the proposed fuzzy System

The consequential part of the IF-THEN logic type fuzzy rules is tuned to obtain the optimal fuzzy logic control model. The membership functions at the antecedents are tuned in this process. Several number of simulation results are carried out. The results obtained for different reference trajectories, during simulation, are given in Figs.7-9. In Fig. 7, a sinusoidal reference trajectory is given to the pitch axis and a step reference trajectory is given to the yaw axis. In Fig. 8, a step reference trajectory given to the pitch axis and a sinusoidal reference trajectory given to the yaw axis.

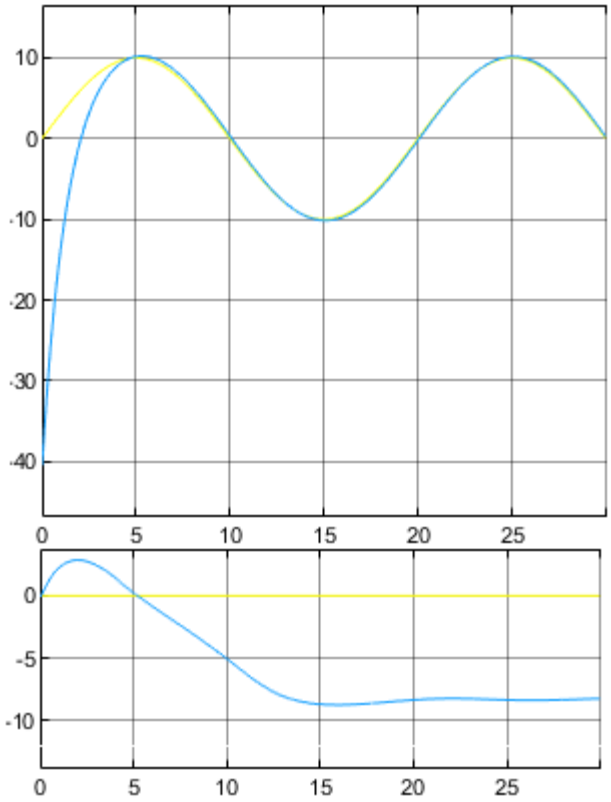


Fig.7 Sinusoidal reference trajectory for pitch axis and step reference trajectory for yaw axis.

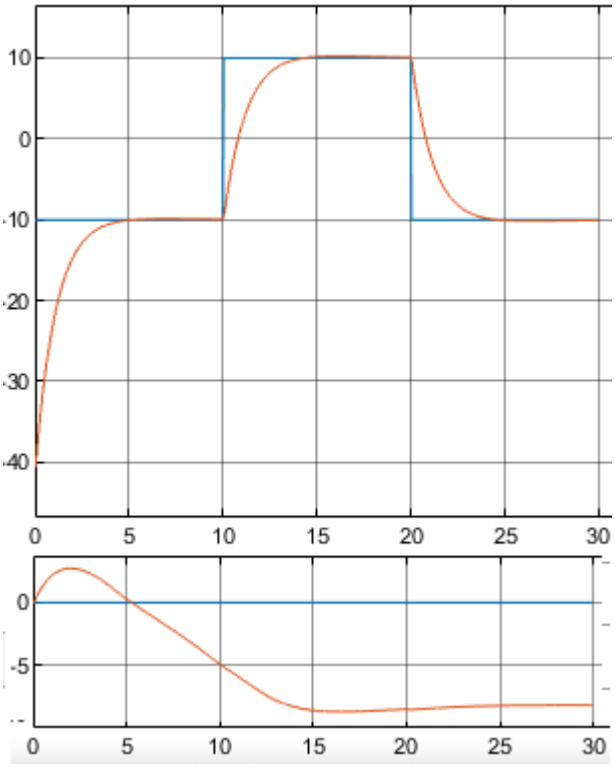


Fig.8 Step trajectory for pitch axis and Sinusoidal reference trajectory for yaw axis.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:54:18 UTC from IEEE Xplore. Restrictions apply.

V. RESULT ANALYSIS

The study of helicopter model has been done by implementing conventional LQR controller and fuzzy controller using Simulink/MATLAB. The sinusoidal and step input trajectories are considered. Table2 represents time response analysis in order to have the performance verification of the implemented controllers. The desired pitch and yaw axis trajectories of a 2-DOF helicopter system are tracked and compared in this section.

TABLE 2 : Time Response Results

	Settling Time (milli seconds)	Peak Over-shoot (in %)
LQR	4.72	8.99
FUZZY	0.95	0.5

The response after implementing LQR and Fuzzy controllers on 2-DOF helicopter system are discussed in Fig. 9 and Fig. 10 respectively.

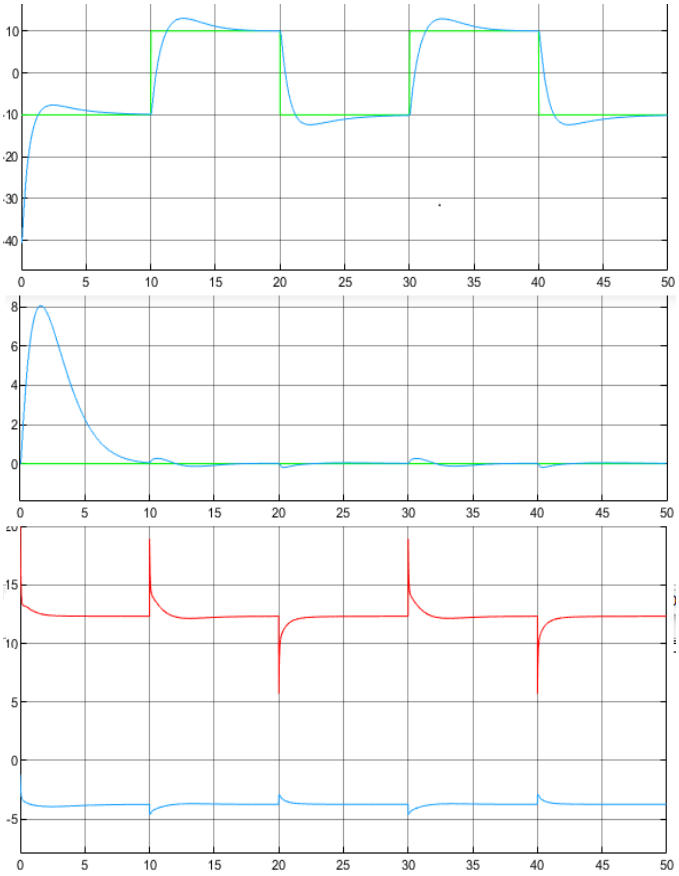


Fig.9 Response after implementing LQR Controller on 2-DOF Helicopter System.

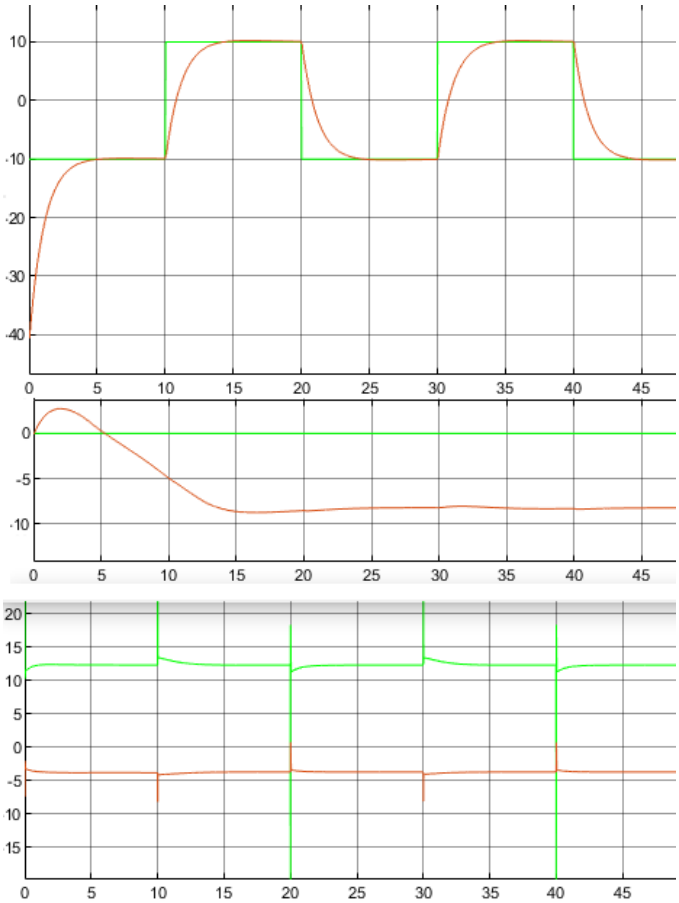


Fig.10 Response after implementing fuzzy Controller on 2-DOF Helicopter System.

VI. CONCLUSION

This paper describes the implementation of Fuzzy controller on helicopter system. The conventional LQR results are compared with intelligent fuzzy control. Analysis of implemented control techniques have been done for yaw and pitch angles. Fuzzy is employed to provide desired trajectory and eliminates cross-coupling between the rotors. The performance of the LQR & fuzzy is assessed by using Simulink/MATLAB. For the comparison of the controller's results the time response analysis has been done. Fuzzy controller shows its excellence when introduced to steady-state error and settling time with respect to LQR control.

REFERENCES

[1] M. Hernandez-Gonzalez, A.Y. Alanis, and E.A. Hernandez-Vargas, "Decentralized discrete-time neural control for a Quanser 2-DOF helicopter," Applied Soft Computing, vol. 12, no. 8, pp. 2462-2469, 2012.

[2] T. Kiefer, K. Graichen, and A. Kugi, "Trajectory Tracking of a 3DOF Laboratory Helicopter Under Input and State Constraints," IEEE Trans. on Control Systems Technology, vol. 18, no. 4, pp. 944-952, 2010.

[3] I.A. Raptis, K.P. Valavanis, and G.J. Vachtsevanos, "Linear Tracking Control for Small-Scale Unmanned Helicopters," IEEE Trans. on Control Systems Technology, vol. 20, no. 4, pp. 995-1010, 2012.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:54:18 UTC from IEEE Xplore. Restrictions apply.

[4] K. Tanaka, H. Ohtake, M. Tanaka, and H.O. Wang, "Wireless Vision-Based Stabilization of Indoor Microhelicopter," IEEE/ASME Trans. on Mechatronics, vol. 17, no. 3, pp. 519-524, 2012.

[5] C. Liu, J. Pan, and Y. Chang, "PID and LQR trajectory tracking control for an unmanned quadrotor helicopter: Experimental studies," in Proc. 35th Chin. Control Conf. (CCC), Jul. 2016, pp. 10845–10850.

[6] M. Jahed and M. Farrokhi, "Robust adaptive fuzzy control of twin rotor MIMO system," Soft Comput., vol. 17, no. 10, pp. 1847–1860, Oct. 2013.

[7] J. Gao, P. Wu, T. Li, and A. Proctor, "Optimization-based model reference adaptive control for dynamic positioning of a fully actuated underwater vehicle," Nonlinear Dyn., vol. 87, no. 4, pp. 2611–2623, 2017.

[8] S. Ghabraei, H. Moradi, and G. Vossoughi, "Multivariable robust adaptive sliding mode control of an industrial boiler–turbine in the presence of modeling imprecisions and external disturbances: A comparison with type-I servo controller," ISA Trans., vol. 58, pp. 398–408, Sep. 2015.

[9] S. Mondal and C. Mahanta, "Adaptive second-order sliding mode controller for a twin rotor multi-input-multi-output system," IET Control Theory Appl., vol. 6, no. 14, pp. 2157–2167, Sep. 2012

[10] J. D. Boskovic and R. K. Mehra, "A decentralized fault-tolerant control system for accommodation of failures in higher-order flight control actuators," IEEE Trans. Control Syst. Technol., vol. 18, no. 5, pp. 1103–1115, Sep. 2010.

[11] K. Xu, G. Zhang, and Y. Xu, "Dual Fuzzy Neural Network Control in Civil Aviation Intelligent Landing System," AICI 2011, Part II, LNAI 7003, pp. 166-174, Springer-Verlag Berlin Heidelberg, 2011.

[12] S. Kurnaz, O. Cetin, and O. Kaynak, "Adaptive neuro-fuzzy inference system based autonomous flight control of unmanned air vehicles," Expert Systems with Applications, vol. 37, no. 2, pp. 1229-1234, 2010.

[13] I. Batyrshin and O. Kaynak, "Parametric classes of generalized conjunction and disjunction operations for fuzzy modeling," IEEE Trans. on Fuzzy Systems, vol. 7, no. 5, pp. 586-596, 1999.

[14] I. Batyrshin, O. Kaynak, and I. Rudas, "Fuzzy modeling based on generalized conjunction operations," IEEE Trans. on Fuzzy Systems, vol. 10, no. 5, pp. 678-683, 2002.

[15] A.C. Aras, O. Kaynak, and R. Abiyev, "Slip control of a quarter car model based on type-1 fuzzy neural system with parameterized conjunctions," IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, pp. 2488-2493, 2012.

[16] R. Singh, B. Bhushan, A. Varshney, "Real-Time Simulation and Control of Helicopter Systems," 2019 International Conference on Power Electronics, Control and Automation (ICPECA).

[17] Quanser Inc., Quanser 2-DOF Helicopter Manual, *Tech. rep.*, Quanser 2010.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on December 09,2021 at 10:54:18 UTC from IEEE Xplore. Restrictions apply.

Journal of Engg. Research ICARI Special Issue pp. 215-224 DOI : 10.36909/jer.ICARI.15317

Wear and Residual Stress Analysis of Waste Sea Shell and B_4C Particles Reinforced Green Hybrid Aluminium Metal Composite

Vipin Kumar Sharma*, Sri Kant Rana**, Roop Lal*** and Ramakant Rana*

* *Department of Mechanical and Automation Engineering, Maharaja Agrasen Institute of Technology, Rohini, Delhi-110086, India*
** *Gina Cody School of Engineering and Computer Science, Concordia University, Montreal, Canada*
*** *Department of Mechanical Engineering, Delhi Technological University, Delhi*
*** *Corresponding Author : rooplal@dtu.ac.in*

ABSTRACT

Present work, evaluates the effects of Sea shell and B4C powder on the mechanical behavior of the aluminium material (Al 6082). Stir casting method was used to fabricate a hybrid composite of Al 6082 with sea shell and B4C. A linear reciprocating tribometer was used to evaluate the wear and friction behavior. The addition of sea shell and B4C particles, resulted in 7-13 % reduction in coefficient of friction and 32-43 % improvement in wear resistance as compared to the Al 6082 alloy. The average Vicker hardness was also improved by 20-70 %. The residual stresses developed during the mechanical testing were also measured to inspect the generation of residual stresses in the fabricated composite. Optical micrographs and scanning electron microscope (SEM) were obtained to analyze the prepared composites for the wear behavior. Waste sea shells were reinforced with B4C in Al 6082 alloys. Micro-hardness along with microstructure and residual stress of the developed green hybrid aluminium metal composite are compared and presented. The wear and friction data have also been shown in this paper for the use of green hybrid aluminium composite in tribological applications.

Key words: Sea shell; B4C; Wear; Friction; Residual stress .

INTRODUCTION

In the recent past, numerous efforts have been made to find out the natural fibers which could be used as reinforcement in the aluminium matrix (Athijayamani et al. 2015). Natural fibers are not man-made fibers. These are produced from animals or plants. From the last few years, the usage of natural fibers has gained a tremendous increase in the composite materials. Bamboo, bagasse, kenaf, flax, jute sisal are some of the natural fibers which are produced in abundance and has been used by researchers in the fabrication of composite materials. These reinforcements contain some important metallic compounds in them which helped in improving the mechanical properties of the aluminium. The low weight, cost, easy usage, good mechanical properties (tensile modulus, surface finish) are some of the factors due to which natural fibers are gaining so much attention in the fabrication of composite materials.

The marine environment has many living organisms like fishes, oysters, and sharks. These organisms protect themselves from the difficult condition by a covering known as shells and scales. These sea shells contribute to carbon-di-oxide pollution and need to be disposed of properly. The calcium carbonate ($CaCO_3$) reacts with the atmospheric air and generates carbon-di-oxide. The utilization of sea shells in the material industry is a positive move which protects the environment. The sea shells consist of calcareous material which has good thermal and

mechanical properties. The outer layer of the shells is made of composite material which is brittle in nature and helps in providing higher adsorption force to the external loads. Due to these important factors, sea shells are finding their use in the composite material and coating applications. The calcinated sea shells were used in the NiP matrix by Jagatheeshwaran et al. (2015). The sea shell and NiP coatings were produced on the steel substrate and pin on disc experiments were performed for the determination of wear properties. The sea shell mixed composite coating resulted with better wear resistance as compared to the NiP coated and uncoated steel specimens. However, the sea shell mixed composite coating resulted in a higher friction force. This limitation of the sea shell mixed composite coating was overcome by Jagatheeshwaran et al. (2017) by incorporating ZnO particles in the NiP and sea shell. The wear and friction properties of the prepared coating were much improved as compared to the uncoated specimens. Chong et al. (2005) prepared the plastic materials using the sea shells for the fire retardation applications. The presence of $CaCO_3$ in the sea shell played an important role in providing the fire retardation properties of the plastic material. The $CaCO_3$ decomposes to calcium oxide and CO_2 at higher temperatures and hinders the oxygen supply required for the fire. Alvarenga et al. (2012) provided a detailed review of the usage of the olyster shells. Authors provided a detailed study of the impacts of sea shell wastes on the environment. The conversion of sea shells to powdered $CaCO_3$ also requires a lot of machine work which also affects the environment. In their study, the impact of sea shell when these are deposited to landfills and impact when these sea shells were transformed into raw material as $CaCO_3$ powder was considered. The depositing of sea shells to land had an impact 93mPt whereas powdered $CaCO_3$ had an impact of 58.97 mPt. Moustafa et al. (2015) used the sea shell for the fabrication of the acrylonitrile-butadiene-styrene copolymer. The effective utilization of the $CaCO_3$ in the sea shells was reported by the authors. For the comparison of different properties, commercially available $CaCO_3$ was also used to fabrication the composite. Later, the fabricated materials were compared for various mechanical properties. It is reported that the presence of sea shells enhanced the mechanical properties of the fabricated composite material.

From the literature review, it is found that sea shell particles have been used to fabricate different composite coating materials. The sea shell has helped in improving the wear resistance of the products, however the friction force increases because of the presence of soft phase of the sea shell. To reduce the friction force as well as the wear from the aluminium material, attempts have been made in the present research, by incorporating the B_4C particles along with sea shell. A hybrid aluminium metal composite has been fabricated using the sea shells and B_4C particles using the stir casting method. The prepared composite was investigated for wear, friction, hardness and residual stresses for analyzing the effects of reinforcement particles.

2 MATERIALS AND METHOD

2.1 Matrix Material

Aluminium alloy (Al 6082) was used as the matrix material for the fabrication of the hybrid AMC. Al 6082 is the member of the Al-Mg-Si family and is mostly used in the construction industry. The chemical composition of the Al 6082 is given in Table 1.

Table 1. Chemical composition of Al 6082.

Element	Cr	Cu	Fe	Mg	Si	Zn	Al
%	0.25	0.1	0.5	0.8	0.6	0.9	Balance

2.2 Reinforcements

In the present work, Sea Shell (SSH) and Boron Carbide (B_4C) particles were used as reinforcements. The SSHs were produced by the exoskeleton of an invertebrate, and it generally has calcium carbonate. The presence of these elements makes it the potential reinforcement which could enhance the properties of the materials. The other reinforcement used is B_4C. B_4C is a very hard boron-carbon ceramic. The presence of these phases in the composite enhances the wear and friction properties of the composite. The average particle size for B_4C was 1-3 μm and for SSH it was 0.5-2 μm.

2.3 Fabrication of Hybrid Composite

There are many techniques which were used by researchers for the fabrication of hybrid aluminium composites (Erbayrak and Erbayrak 2020). Stir casting (Sharma et al. (2017)) and powder metallurgy (Nassar and Nassar (2017), Singh et al. (2018)) are among those which are mostly used. In this work, a stir casting setup was used to fabricate the composite. Initially, the SSH and B_4C powder were processed by using a ball milling setup to obtain the micro-sized particles and a proper mix as per the requirement. In the ball milling process, the SSH and B_4C were added with hardened steel balls in the ratio of 1:10 by weight. The Ball mill was rotated for 10 hours at 100 rpm to obtain uniform sized particles. The powdered samples of B_4C and SSH were examined under a scanning electron microscope for the homogeneous size distribution. In this work, two compositions of the hybrid composites were used. In the first composite, 8 w/w% of SSH and 4w/w% of B_4C particles were reinforced into Al 6082 and in the second composite 4w/w % of SSH and 8w/w% of B_4C were reinforced. The typical setup of the stir casting method consists of three main things. The first is the graphite crucible. This crucible is used to to keep the raw material and helps in melting it. The other part is the heating chamber which provides the required heat to melt the raw material and last part is the stirring system to mix the reinforcements homogeneously in the matrix material. The fabrication of the composite materials was done in a close atmosphere, and the reinforcements were introduced in to the molten metal at a constant rate as per the requirement measurements. For homogeneous mixing, the matrix and reinforcements are mixed with the help of a stirrer. It is done by using a mild stell blades stirrer. The chamber of the furnace was first heated to a temperature of 500ºC and then the crucible was filled with aluminum metal. The temperature of the chamber was increased to 900ºC to allow the aluminum to melt completely. The reinforcements were preheated in another heater to a temperature of about 200ºC. The stirring was started at a time when the metal reaches to fully molten sate. After stirring for 15 minutes at a temperature of 800ºC, the slurry was reheated and held at a temperature of 950ºC for 15 minutes so that, the slurry was in a completely molten state. A preformed metal mold was used to fabricate the samples. The slurry was then poured into the mold at a uniform rate to avoid trapping of gasses. Then the mold was allowed to cool and the metal was being ejected from the mold.

2.4 Measurement of Properties

After the successful fabrication of the composites, several properties were evaluated. Wear and friction, micro hardness, and residual stress were the main properties which were evaluated for the prepared composites.

2.4.1 Wear and Friction

Rotary Pin on disc (Lal and Singh (2018), Singh et al. (2019), Lal and Singh (2019), Rana et al. (2020)) and reciprocating pin on disc (Sharma et al. (2019) are the two methods which has been generally used by researchers for evaluating wear and friction behavior. In this study, a linear reciprocating tribometer (LRT) was used for the measurement of the coefficient of friction of the prepared composite. LRT consists of a fixed flat plate and a reciprocating pin. The pin reciprocates on the flat plate with some frequency and stroke length under the effects of the load. The friction force between the flat plate and the pin was measured with the help of friction force sensors.

This friction force was later converted to the coefficient of friction by diving the obtained friction force by the applied load. The amount of wear from the plate and pin surfaces was measured by weighing the flat plate specimen by using a weighing balance with 0.0001 g accuracy. In this work, the prepared hybrid composites were used as the flat plate and hardened steel as pin specimen. The experiments were performed at 10 Hz of the frequency with 10 mm as stroke length. A load of 30 N was applied on the pin. Before starting the experimentation, the pin and flat plate surfaces were cleaned with the help of acetone and were dried well.

2.4.2 Micro-hardness and Microstructure

An optical microscope and scanning electron microscope (SEM) were used to inspect the distribution of the reinforcements in the aluminium matrix. The specimen surfaces were cleaned and polished by using different grades (400, 800, 1200, 1600 grit) of SiC abrasive papers. After this the wet polishing was performed with 0.05µm alumina powder. The prepared surfaces were etched and observed under an optical microscope at 50X magnification.

The micro-hardness of the material directly affects the wear and friction properties of the material (Jeyapandiarajan and Anthony 2019, Ullen et al. 2020). A Vicker hardness tester was used to measure the micro-hardness as per ASTM E-384 standards. A diamond indenter was used to indent the composite surface with a load of 3000 mN for 20 seconds. Three observation of hardness were recorded and average was taken to draw the results.

2.4.3 Residual Stress

With the change in the natural shape of the material, some residual stresses remain in the materials. These residual stresses are generally relieved by using the heat treatment process. The residual stresses in the material may have negative effects on the properties and lifetime of the material. So, it is important to measure the residual stresses to predict the quality of the materials. The residual stress in the prepared composites was measured with the help of a residual stress analyzer (Pulstec limited model-µ X-360n). The residual stress analyzer works the principle of Bragg's law. For the measurement, X-ray irradiation time was fixed at 30 seconds with tube current of 1 mA and 30kV voltage. The incidence angle and X-ray wavelength was set at 35º and 2.29 A. The diffraction data were calculated by using the diffraction peak positions. The residual stress in the material changes the lattice spacing. This change in lattice space was used to calculate the diffraction data and later residual stress.

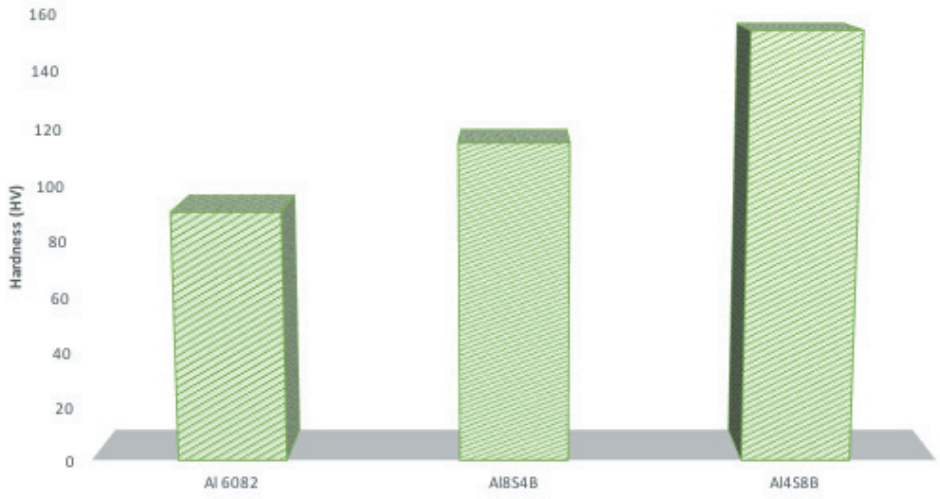
3. RESULTS AND DISCUSSION

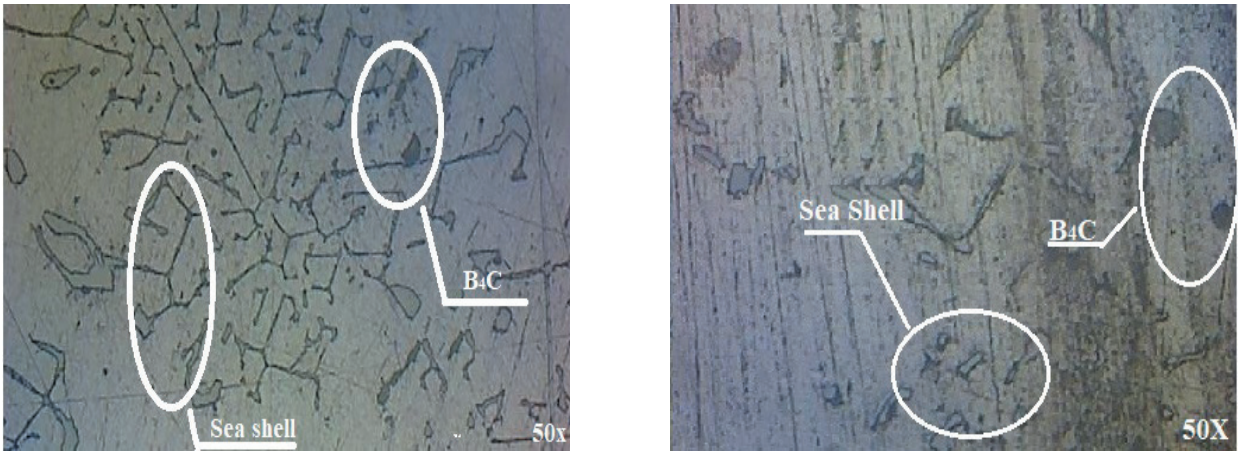
3.1 Microstructure and Hardness

The Vicker hardness measurement results are presented in Figure 3. It is inferred from the results that B_4C and SSH enhanced the hardness of the Al 6082 material. The Al 6082 had 90.13 HV Vicker hardness which increased to 114.63 HV with the addition of 8% SSH and 4% B_4C. With the further increase in B_4C contents, the Vickers hardness improved to 153.63 HV for 4%SSH and 8% B_4C sample.

The proper distribution of the reinforcement in the matrix material greatly enhanced the mechanical properties. For the inspection of the distribution of reinforcements in the Al 6082 alloy, optical micrographs along with scanning electron microscope (SEM) images were obtained. Figure 4 (a) presents the optical micrographs and SEM image for the Al8S4B. It was observed from the image that carbonate particles wear uniformly distributed throughout the matrix of the Al 6082. There were no protruded particles on the surface of composite. This distribution of SSH particles enhanced the machinability of the aluminium. The SEM image of Figure 4 (b) shows the B_4C protruding from the hybrid composite surface. These particles provided the wear resistance to the aluminium material.

Figure 3. Vicker micro-hardness observations





(a) (b)

Figure 4. Micrographs of (a) Al8S4B (b) Al4S8B

3.2 Wear and Friction

A linear reciprocating tribometer was used to measure the wear and friction properties of the fabricated composites. The composite samples were used as a flat plate on which a die steel pin was made to reciprocate under a normal load. Figure 5 presents the variation of the coefficient of friction. For the easy understanding of the COF curve, it was divided into four parts, these were "running in", "mixed zone", "steady-state" and "wear out". During the "running in" zone, the asperities of the plate and pin surface made contact, this in result increased the COF value. After the initial sliding, the asperities get removed and COF starts decreasing in the "mixed zone" and attained a near-constant value in the "steady-state" region as indicated in Figure 5. The "wear-out" region started at the end of the steady-state region and COF started to increase again. The average COF obtained for Al 6082, Al8S4B and Al4S8B were 0.89, 0.72 and 0.73 respectively. It was revealed that presence of SSH helped in reducing the COF values.

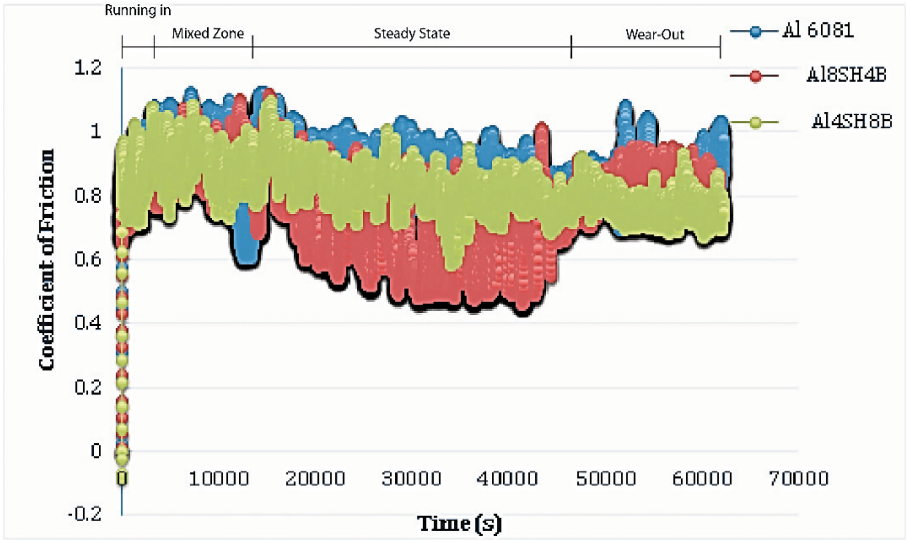


Figure 5. Variation of the coefficient of friction

The amount of wear from the plate samples was measured using a weighing balance. Figure 6 presents the wear in the form of mass loss. It is clearly seen from Figure 6 that, an average mass loss for the fabricated composites is lower, compared to the Al 6082 material. The mass loss for AlS8B4, Al4S8B and Al 6082 are 0.0459 g, 0.0393 g and 0.0759 g respectively. The presence of B_4C particles in the aluminium matrix greatly improved the wear resistance. The hardness of B_4C particles helped in improving the wear resistance. The B_4C particles prevented the wear of the aluminium matrix. The SSH present in the composite lowered the COF which also helped in reducing the wear. It was clearly observed from Figure 6 that, as more amount of B_4C was added in the composite, the mass loss further reduced. For the better understanding of the worn nature of the fabricated composite materials, SEM images were obtained for the wear samples. Figure 7 (a) presents the SEM image for the worn Al8S4B composite. Deep grooves and thick wear lines were observed on the worn surface. Wear debris detached from the composite were also observed on the surface. This indicated that the worn nature was mostly abrasive. The SEM image of Al4S8B indicated low amount wear wedges on the worn surface. Thin wear lines, as compared with the wear in Al8S4B, also showed lower wear for the AL48B composite. The presence of more B_4C helped reduce the mass loss for the Al4S8B composite (Figure 7(b)).

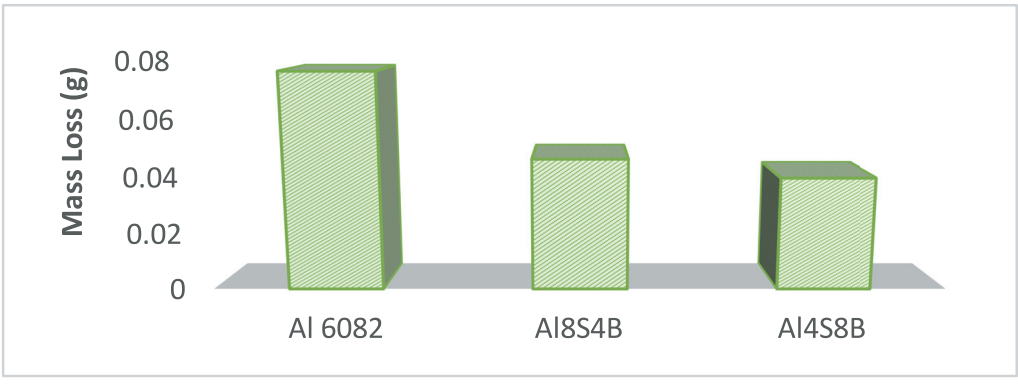
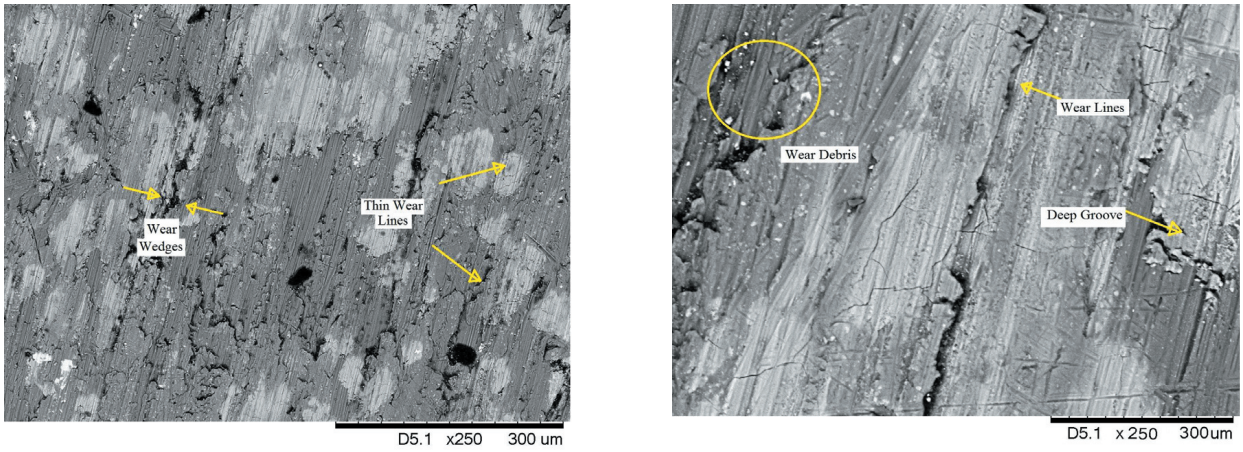


Table 6. Mass loss from the fabricated materials



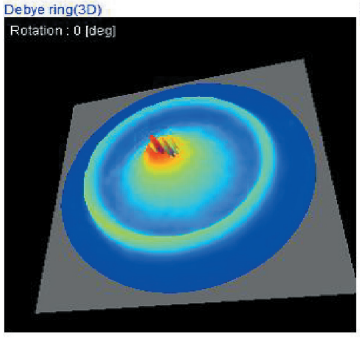
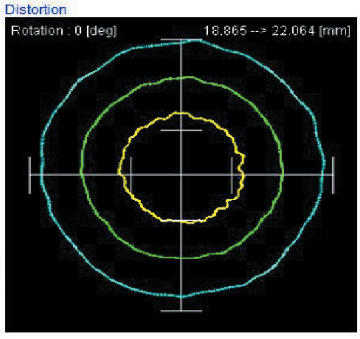
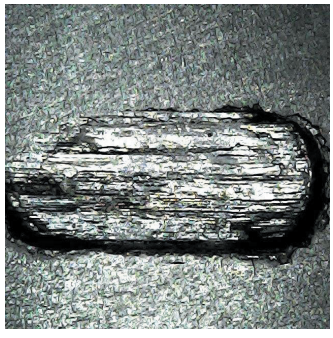
(a) (b)

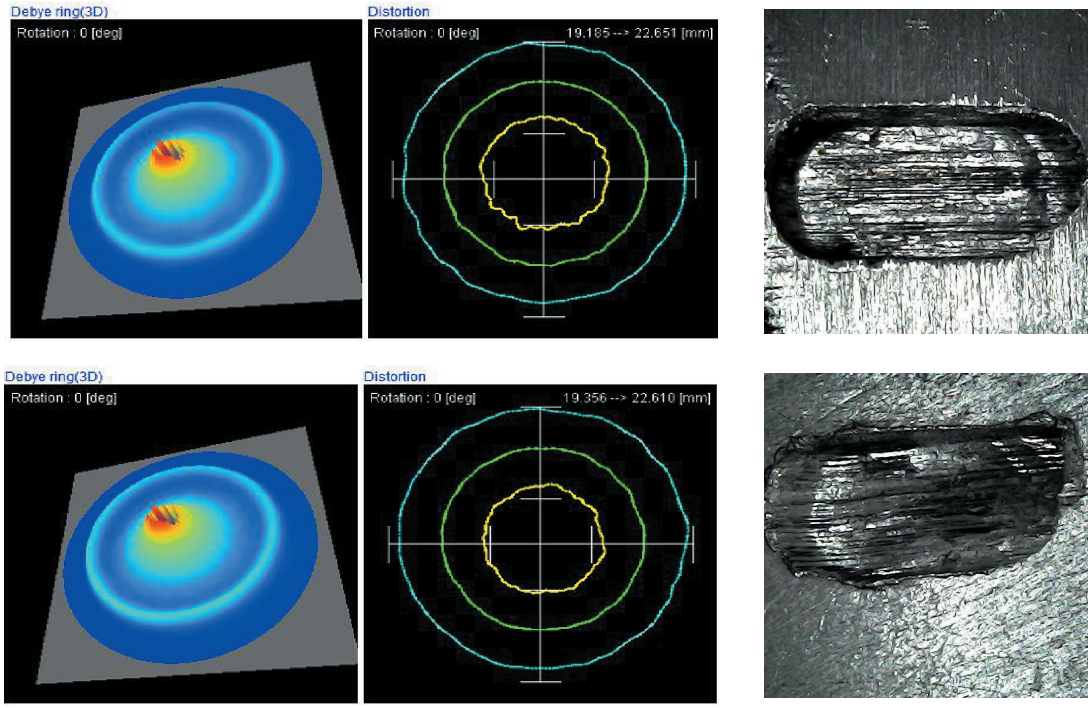
Figure 7. Scanning electron microscope image for (a) Al8S4B (b) Al4S8B composite

3.3 Residual Stress Analysis

A residual stress analyzer was used to measure the residual stress in the fabricated composites.. It was seen that the Al 6082 had a minimum amount of residual stress as compared to the fabricated composites. The Al 6082 had 99 MPa as residual stress. Al8S4B and Al4S8B had 129 MPa and 127 MPa residual stress. The introduction of reinforcement materials into the matrix of Al 6082 generated this residual stress which needs to be released before commercially using the materials.

For a better understanding of the residual stresses, debye ring and distortion profiles were also generated for all the samples. The Debye ring as shown in Table 5 represents the concentration of the residual stress in the tested materials. The value of the residual stress increases from the red region to the blue region. The region of the worn sample where, the pin makes the first contact with the composite material resulted with maximum residual stress, and as the pin moved on the surface, the residual stress kept on decreasing. The distortion rings indicated the damage occurred on the material which lead to the generation of residual stresses. The distortion images shown in Figure indicates the same amount of damage to all the three considered material. No abrupt or sudden damage was reported.

Debye Curve	Distortion Curve	Wear Area Inspected	Material
			Al 6082



Al8S4B

Al4S8B

CONCLUSIONS

In the present experimental work, Sea shells and B4C particles were reinforced into the Al 6082 alloy for improving its wear and friction properties. A stir casting setup was used to fabricate two-hybrid composites. Al 6082with 8%w/w sea shell and 4%w/w B4C and Al 6082 with 5%w/w sea shell and 8%w/w B4C composites were prepared. These composites were tested for various mechanical properties and the following conclusion could be drawn from this experimental study.

1. The sea shell and B4C particles were successfully used to fabricate the composites. Scanning electron microscope images indicated a uniform distribution of the sea shell and B4C particles in the Al 6082 matrix.

2. The micro-hardness tests revealed that reinforcements enhanced the hardness of the composite. B4C is the major contributing factor which improved the hardness. There was an increase of upto 70 % as compared with the hardness of Al 6082 material.

3. The reinforcements also improved the wear and friction properties of the composite. The average coefficient of friction enhanced to13% with comparison to Al 6082. The wear loss from the Al 6082 also gets reduced with the introduction of sea shell and B4C particles. The wear reduced upto 43%. There was a reduction of mass loss to 34% for Al8S4B material and 43% with Al4S8B material. However, the reinforcement in the Al 6082 resulted in some increase in the residual stresses. The residual stresses in the fabricated composites increase to 30%.

REFERENCES

Alvarenga, R. A. F. de, Galindro, B. M., Helpa, C. de F., & Soares, S. R. 2012. The recycling of oyster shells: An environmental analysis using Life Cycle Assessment. Journal of Environmental Management.106: 102–109.

Chong, M. H., Chun, B. C., Chung, Y.-C., & Cho, B. G. 2006. Fire-retardant plastic material from oyster-shell powder and recycled polyethylene. Journal of Applied Polymer Science. 99(4): 1583–1589. https://doi.

org/10.1002/app.22484

El-Daly, A. A., Abdelhameed, M., Hashish, M., & Daoush, W. M. 2013. Fabrication of silicon carbide reinforced aluminum matrix nanocomposites and characterization of its mechanical properties using non-destructive technique. Materials Science and Engineering: A. 559: 384–393. https://doi.org/10.1016/j.msea.2012.08.114

Jagatheeshwaran, M. S., Elayaperumal, A., & Arulvel, S. 2017. Impact of nano zinc oxide on the friction – Wear property of electroless nickel-phosphorus sea shell composite coatings. Materials Science and Engineering: B. 225: 160–172.

Jagatheeshwaran, M. S., Elayaperumal, A., & Arulvel, S. 2015. Wear characteristics of electroless NiP/bio-composite coatings on En8 steel. Journal of Manufacturing Processes. 20: 206–214. https://doi.org/10.1016/j.jmapro.2015.08.002

L. Lancaster, M. H. Lung, and D. S. 2013. Utilization of Agro-Industrial Waste in Metal Matrix Composites: Towards Sustainability. International Journal of Environmental and Ecological Engineering. 7(1): 35–43.

Lal Roop and Singh R.C. 2019. Investigations of tribodynamic characteristics of chrome steel pin against plain and textured surface cast iron discs in lubricated conditions. World Journal of Engineering. 16(4):560-568.

Lal, R. and Singh, R. 2018. Experimental comparative study of chrome steel pin with and without chrome plated cast iron disc in situ fully flooded interface lubrication. Surface Topography: Metrology and Properties. 6(3): 035001.

Madakson B, Yawas D.S., Apasi A. 2012. Characterization of Coconut Shell Ash for Potential Utilization in Metal Matrix Composites for Automotive Applications. International Journal of Engineering Science and Technology. 4(3): 1190–1198.

Moustafa, H., Youssef, A., Duquesne, S. and Darwish, N. 2015. Characterization of bio-filler derived from seashell wastes and its effect on the mechanical, thermal, and flame retardant properties of ABS composites. Polymer Composites 38(12):2788-2797.

Nassar Amal E., Nassar Eman E. 2017. Properties of aluminium matrix Nano composites prepared by powder metallurgy processing. Journal of King Saud University – Engineering Sciences. 29: 295–299

Rana R., Murtaza Q., Walia R.S. 2020. GA based optimization of tri-bological behaviour of diamond coated tungsten carbide. World Journal of Engineering. 17(3): 335-346.

Sharma, V. K., Singh, R. C., & Chaudhary, R. 2017. Effect of flyash particles with aluminium melt on the wear of aluminium metal matrix composites. Engineering Science and Technology, an International Journal. 20(4): 1318–1323.

Sharma, V. K., Singh, R. C., Chaudhary, R., Saxena, M., & Anand, M. 2019. Effects of flyash addition on the dry sliding tribological behavior of aluminum composites. Materials Research Express. 6(8): 0865f4. https://doi.org/10.1088/2053-1591/ab24fd

Sharma, V., Singh, R. and Chaudhary, R. 2018. Wear and friction behaviour of aluminium metal composite reinforced with graphite particles.International Journal of Surface Science and Engineering. 12(5/6): 419-432.

Singh, R., Chaudhary, R. and Sharma, V. 2019. Fabrication and sliding wear behavior of some lead-free bearing materials. Materials Research Express. 6(6): 066533.

A Athijayamani, C Manickam, J Kumar, N Diwahar 2015. Mechanical and wear behaviors of untreated and alkali treated roselle fiber-reinforced vinyl ester composite. Journal of Engineering research.3 (3):

P Jeyapandiarajan and Xavior M Anthony 2019. Influence of cutting condition on machinability aspects of Inconel 718: A review paper. Journal of Engg. Research. 7(2): 315-332

N. Beköz Ullen, S.M.A. Hasak, M.H. Dirikolu 2020. Factors Influencing the Machinability during Turning Sinter-Hardened Cu-Ni-Mo Based Steel: Dependency on Cutting Speed, Feed Rate, and Cutting Depth. Journal of Engg. Research. 8 (4): 236-257

Seda Erbayrak and Engin Erbayrak 2020. Determination of the Impact Damage Threshold Point of Composite Material Using Fuzzy-Base Taguchi Method. Journal of Engg. Research. 8 (3): 153-171