SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(70th Edition)

OCTOBER, 2018

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 **Seventy** 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 **October, 2018**. 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. Analog and RF assessment of sub-20 nm 4H-SiC trench gate MOSFET for high frequency applications, ***3.Rishu Chaujar***, Applied Physics, DTU

2. Analysing the Influence of Improved Situation, Capability Level of Actors and Flexible Process Workflow on Public Value of EGovernance Projects in India, ***6.Parul Jain Gupta*** and ***3.Pradeep Kumar Suri***, DSM, DTU

3. Application of Linear Quadratic Methods to Stabilize Cart Inverted Pendulum Systems, ***3.Sudarshan K Valluru***, ***3.Madhusudan Singh*** and ***6.Mayank Singh***, Electrical, DTU

4. Bifurcation Curves for Electrical DC Motors, ***3.Sudarshan K Valluru***, ***3.Madhusudan Singh***, ***8.Aditya Verma*** and ***8.Anshul Gupta***, Electrical, DTU

5. Binary whale optimization algorithm and its application to unit commitment problem, ***6. Vijay Kumar[1]*** and ***3.Dinesh Kumar[2]***, [1]Computer Science and [2]ECE, DTU

6. Bird Species Classi cation using Transfer Learning with Multistage Training, Sourya Dipta Das and ***6.Akash Kumar***, DTU

7. Cement Industry and Strategies for Mitigating Carbon Emissions: An Overview, ***8.Ahana Ghosh*** and ***3.Anubha Mandal***, Environmental, DTU

8. Comparative Study of Mechanical Properties of TIG Welded Joints of Similar and Dissimilar Grades of Stainless Steel Material, ***3.Roop Lal***, ***6.Mohd Shuaib*** and ***6.Vikal Paliwal***, Mechanical, DTU

9. Comparative study of surface coating for automotive application-A review, Ramesh ChandraYadaw, Shailesh Kumar Singh, Somanth Chattopadhyaya, Sanjeev Kumar and ***3.R.C.Singh***, Mechanical, DTU

10. Determination and Allocation of Optimal Size of TCSC and STATCOM for obstruction alleviation in Power System, ***6.Jitender Kumar*** and ***3.Narendra Kumar***, Electrical, DTU

11. Effect of various culture conditions on shoot multiplication and GC–MS analysis of Plumbago zeylanica accessions for plumbagin production, ***6.Arpita Roy*** and ***3.Navneeta Bharadvaja***, Biotechnology, DTU

12. Enhancement of 1.54 mm emission in Ce3þ-Er3þ codoped Ca4Si2O7F2 phosphor, Rupesh A. Talewar, Sk. Mahamuda, Arpita Vyas, ***3.A.S. Rao*** and S.V. Moharil, Applied Physics,, DTU

13. ensmallen: a flexible C++ library for efficient function optimization, ***8.Shikhar Bhardwaj***, Ryan R. Curtin, Marcus Edel, Yannis Mentekidis and Conrad Sanderson, Mathematics and Computer, DTU

14. Experimental Investigation on the Mechanical Properties and Water Absorption Behavior of Randomly Oriented Short Pineapple/Coir Fiber-Reinforced Hybrid Epoxy Composites, ***6.Mohit Mittal*** and ***3.Rajiv Chaudhary***, Mechanical, Production, and Industrial, DTU

15. Geothermal Energy Resource of North-western Himalayas, ***7.Piyush Rawat*** and ***3.JP Kesari***, Mechanical, DTU

16. Green Power Allocation for Cognitive Radio Networks with Spectrum Sensing, Ramnaresh Yadav, ***6.1.Ashwani Kumar*** and Keshav Singh, ECE, DTU

17. Health and economic growth: Evidence from dynamic panel data of 143 years, ***3.Rajesh Sharma***, DSM, DTU

18. Manned Mars Mission Risks Evaluation, Laura Bettiol, Andrea De La Torre, Divyesh Patel, Funmilola Oluwafemi, Guzel Kamaletdinova, ***6.Robin Kumar Singh***, Uthpala Heshani, Yasith Lakmal, Aureliano Rivolta and Andrey Sorokin, DTU

19. Modelling of vapour compression refrigeration system (VCRS) by using alternative refrigerants with TiO2 and without nano materials, ***3.R.S. Mishra***, Mechanical & Production, DTU

20. Multi-Objective Genetic and Adaptive Particle Swarm Optimization Algorithms: A Performance Analysis with Benchmark functions, ***3.Sudarshan K.Valluru*** and ***3.Madhusudan Singh***, Electrical, DTU

21. Negative Capacitance Transistor to Address the Fundamental Limitations in Technology Scaling: Processor Performance, Hussam Amrouch, ***8.1.Girish Pahwa***, Amol D. Gaidhane, J¨org Henkel and Yogesh S. Chauhan, ECE, DTU

22. NFC Secure Element-based Mutual Authentication and Attestation for IoT access, ***3.Divyashikha Sethia***, ***3.Daya Gupta*** and Huzur Saran, Computer Science, DTU

23. Numerical simulations: Toward the design of 18.6% efficient and stable perovskite solar cell using reduced cerium oxide based ETL, ***6.Rahul Pandeya***, ***6.Anand Prakash Saini*** and ***3.Rishu Chaujar***, Applied Physics, DTU

24. Performance Analysis of 4 Stroke Compression Ignition Engine by Using Diesel and Biodiesel Blend of Hingot Oil, Mr. Akash Singh Raikwar, Mr. Amit Agarwal and ***3.Dr. Amit Pal***, Mechanical, DTU

25. Predictive Current Control of Feedback Linearized Induction Motor Drive, ***6.Ambrish Devanshu***, ***3.Madhusudan Singh*** and ***3.Narendra Kumar***, Electrical, DTU

26. Probabilistic Approach Towards Darjeeling Himalaya's Landslides- A Case Study, ***3.Raju Sarkar***, Civil, DTU

27. Qualitative and Quantitative Phytochemical Analysis of Centella asiatica, ***6.Arpita Roy***, ***6.M Laxmi Krishnan*** and ***3.Navneeta Bharadvaja***, Biotechnological, DTU

28. Robust Repetitive Control Design for 3P4W Shunt Active Power Filter, Gitanjali Pandove and ***3.Mukhtiar Singh***, Electrical, DTU

29. Source/Gate Material-Engineered Double Gate TFET for improved RF and linearity performance: a numerical simulation, ***8.Skanda Shekhar***, ***6.Jaya Madan*** and ***3.Rishu Chaujar***, Applied Physics, DTU

30. S-Shaped Metamaterial Ultra-Wideband Directive Patch Antenna, Parul Dawar, ***3.Asok De*** and ***3.N. S. Raghava***, Electronics and Communications, DTU

31. Study for Reduction of Pollution Level in Diesel Engines, Petrol Engines and Generator Sets by Bio signal Ring, ***3.SK Singh***, ***6.Surbhi Jain***, ***6.Tejasvini Ahuja***, ***8.Yukti Sharma*** and ***8.Nidhi Pathak***, Environmental, DTU

32. TECHNOLOGICAL ASPECT IN REAL TIME BIDDING: A PROBABILISTIC APPROACH, ***3.KAPIL SHARMA***, IT, DTU

33. Theoretical model for the effect of dust grains on self-filamentation of a gaussian electromagnetic beam in a fully ionized plasma, ***7.Ruchi Sharma*** and ***3.Suresh C. Sharma***, Applied Physics, DTU

34. Thermodynamic analysis of vapour compression refrigeration system with mixing of Al2O3Nano material in fourteen ecofriendly refrigerants, ***3.R.S. Mishra***, Mechanical & Production, DTU

35. Thermodynamic analysis of vapour compression refrigeration systems using alternative refrigerants, ***3.R.S. Mishra***, Mechanical & Production, DTU

36. To Determine the Effect of Dimpled Poppet Valve on the Flow in Engine Cylinder by Port Flow Simulation, ***8.Anirudh Singh***, ***8.Apoorv Dubey*** and ***3.Raj Kumar Singh***, Mechanical, DTU

37. UV radiation emitting Gd3+ activated Sr2SiO4 host system prepared by sol–gel procedure: structural, electron paramagnetic resonance, and luminescence studies, Vijay Singh, N. Singh, M. S. Pathak, ***6.Sumandeep Kaur***, ***3.M. Jayasimhadri***, S. Watanabe and T. K. Gundu Rao, Applied Physics, DTU

38. Video-based Person Re-identification Using Spatial-Temporal Attention Networks, ***6.Shivansh Rao***, Tanzila Rahman, Mrigank Rochan and Yang Wang, DTU

1. *Chancellor*

2. *Pro Vice Chancellor*	***2.1.*** *Ex Pro Vice Chancellor*
3. *Faculty*	***3.1.*** *Ex Faculty*
4. *Teaching-cum-Research Fellow*	***4.1.****Asst. Librarian*
5. *Alumni*	
6. *Research Scholar*	***6.1.*** *Ex Research Scholar*
7. *PG Scholar*	***7.1.*** *Ex PG Scholar*
8. *Undergraduate Student*	***8.1.*** *Ex Undergraduate Student*

Int. J. Electron. Commun. (AEÜ) 98 (2019) 51–57

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

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

Regular paper

Analog and RF assessment of sub-20 nm 4H-SiC trench gate MOSFET for high frequency applications

Rishu Chaujar

Applied Physics Department, Delhi Technological University, Delhi, India

ARTICLE INFO

Article history:
Received 10 July 2018
Accepted 27 October 2018

Keywords:
Analog
Frequency
Gain
Stray capacitance
RF
4H-SiC-RC-MOSFET

ABSTRACT

In this work, 4H-SiC Trench Gate (Recessed Channel) (4H-SiC-RC) MOSFET structure is demonstrated to have superior analog and RF Figure of Merits (FOMs) by numerical simulation. The integration of 4H-SiC with phosphorene contact on to the trench region results in higher drain current (2.3 mA) and outstanding switching ratio (10^{16}) owing to reduced off-current which leads to the lower sub-threshold slope (SS). Other electrical parameters such as electric field, electron mobility, and electron velocity are remarkably improved in 4H-SiC-RC MOSFET. Further, the high-frequency RF FOMs have also been studied, and it is observed that cut-off frequency (f_T) gets doubled and maximum oscillator frequency (f_{MAX}) increased by five times in the proposed design as compared to conventional design. Thus, the 4H-SiC-RC design is suitable for high switching, high power, and high-frequency application.

© 2018 Elsevier GmbH. All rights reserved.

1. Introduction

Low intrinsic carrier concentration, wide band gap (3.26 eV) and high critical electric field (2.2×10^6 V/cm) properties of 4H-SiC material make it suitable for high power MOSFETs [1]. 4H-SiC based MOSFETs are also suitable for high-speed application due to their fast switching capability [2]. Silicon based power MOSFETs are hindered by the swift in on-resistance with increasing blocking voltage as well as high avalanche breakdown electric field, and high electron saturation velocity [1,3]. GaN is also a wide bandgap material, but it is not advantageous as SiC because it has silicon dioxide as its native oxide by which metal-oxidesemiconductor structures are easy to produce by thermally growing an oxide at the SiC surface [1]. When the MOS devices are miniaturized at sub-20 nm, stray capacitances are increased which reduce the gate controllability over the channel [4]. Due to miniaturization at sub-20 nm scale, Short channel effects (SCEs) come into existence, and the may cause to hinder the device and its analog and RF performances [5]. So there is a need to eliminate SCEs [6,7] for high-performance transistor application at sub 20-nm regime [8]. Many gate engineering schemes and gate materials have been used in previously published work [9,10] to overcome these effects, but to enhancing the switching speed and RF performance [11,12] of RC-MOSFET some new engineering, schemes have been implemented in this work.

In order to reduce SCEs, first engineering scheme is implemented on to the gate and gate material is replaced with the black phosphorus which is used as gate contact due to higher mobility at room temperature ($\sim$1000 $cm^2 V^{-1} s^{-1}$) and a layered structure [13]. Black phosphorus has tunable band gap from bulk (0.3 eV) to a single layer (1.0–1.5 eV) [14]. Due to many more electrical properties, it is the most appropriate material in comparison to other materials. As RC device has some disadvantage such as corner effects thus; to diminish these effects another scheme is introduced in the channel region and silicon (Si) is replaced with 4H-SiC and used as a wafer. The 4H-SiC wafer is used due to superior physical properties of SiC as compared to silicon. 4H and 6H polytypes wafers of SiC are commercially available. Between them, 4H–SiC reflects more isotropic electrical properties, higher carrier mobility, and band gap energy. These properties make it more attractive for use in the fabrication of vertical MOSFET structures than 6H–SiC. 4H-SiC is a compatible material for semiconductor devices. When 4H-SiC is used with silicon in the channel then it may lead some problem at the interface due to different chemical composition of 4H-SiC and silicon channel. It may create trapped charges at the interface due to the defects during the manufacturing process and it may leads gate leakage current. When 4H-SiC is used instead of silicon in the channel, then smaller gate oxide thickness is used due to its material properties to avoid interface charges.

In the proposed 4H-SiC-RC-MOSFET electrical characteristics of a MOSFET are improved as the enhanced value of switching ratio is found, and the lower sub-threshold slope is observed. 4H-SiC is a promising material for high power electronics applications as the

E-mail address: chaujar.rishu@dtu.ac.in

https://doi.org/10.1016/j.aeue.2018.10.035
1434-8411/© 2018 Elsevier GmbH. All rights reserved.

performance of a semiconductor device depends upon the device material characteristics.

2. Device schematics and its parameters

Fig. 1(a) and (b) shows the device schematics of 4H-SiC-RC-MOSFET and conventional recessed channel (CRC) MOSFET respectively. The used parameters are listed in Table 1. A polysilicon gated recessed channel (PS-RC) MOSFET has same parameters except gate material (φ_{PS} = 4.17 eV). In this work, PS-RC-MOSFET device has been used for the comparison of electrical properties with CRC and 4H-SiC-RC MOSFETs.

3. Simulation methodology and calibration

For TCAD simulations, various models have been taken into effect such as, for mobility of carriers under the influence of changing electric field, we have included Parallel Electric Field Dependence (fldmob model) the model takes care of all the parameters affecting the mobility of carriers including velocity saturation effect. For carrier generation-recombination, we have implemented (klasrh model) concentration-dependent Klaassen Shockley-Read-Hall Recombination model as it can include a concentration-dependent lifetime of carriers [15]. As electron-hole pair generation-recombination is a key phenomenon in predicting the behavior of MOSFET, it can be most accurately modeled via (klasrh model). The above-selected models account for almost the entire device physics associated with the MOSFET, thereby producing extremely accurate results. During the measurement, gate bias is swept from 0 to 1.5 V.

To validate the simulation models of a 4H-SiC trench MOSFET the experimental statistics have been drawn-out from [16]. Thus, the experimental data and simulated data is calibrated and plotted as evident in Fig. 2. For 50 nm gate length the simulated data almost matched with the experimental data as evident from Fig. 2. Thus the simulation models are validating.

4. The fabrication process of BP-JL-RC-MOSFET

In order to fabricate the 4H-SiC-RC-MOSFET, some steps are considered which are shown in Fig. 3. However, the 4H-SiC Trench MOSFET of 50 nm has already been fabricated by some researchers such as Song et al. [16]. Similar structure with grooved gate has also been fabricated by Appenzeller et al. [17], fabricated 36 nm groove gate and Xiao-Hua et al. [18], fabricated 140 nm groove gate MOSFET.

In the process of fabrication, first of all, a silicon carbide (4H-SiC) wafer (substrate) is taken, and then the diffusion process is used for the creation of n-well region. Before Drive-In diffusion and gate oxidation process, the active region is identified. Source-drain implantation and anneal process are implemented before groove etching by the combination of reactive ion etching and electron beam lithography, and then the gate oxide is grown by thermal oxidation at the bottom of the groove. Contact etches, and metal deposition (using electron beam evaporation followed by lift-off) process is done in the last fabrication process. Thus, the fabrication of 4H-SiC-RC-MOSFET can do by using the process flow of device design schemes. The proposed device is the aids of the transparent gate engineering scheme which has been combined with groove gate for accomplishing the improved characteristics.

Table 1
Device design parameters of 4H-SiC-RC-MOSFET and CRC-MOSFET.

Device parameters	4H-SiC-RC-MOSFET	CRC-MOSFET
Gate length (L_g)	20 nm	20 nm
Channel length (L_{ch})	44 nm	44 nm
Source and the drain region length	30 nm	30 nm
Width of the device (W)	200 nm	200 nm
Groove depth	38 nm	38 nm
Negative junction depth (NJD)	10 nm	10 nm
Source/drain doping (N_D^+)	5×10^{16} cm^{-3}(N_D^+)	1×10^{19} cm^{-3}(N_D^+)
Physical oxide thickness (t_{ox})	2 nm	2 nm
Gate metal work function (Φ)	5.16 eV	4.4 eV
Gate to source voltage (V_{gs})	1.5 V	1.5 V
Drain to source voltage (V_{ds})	0.2 V	0.2 V

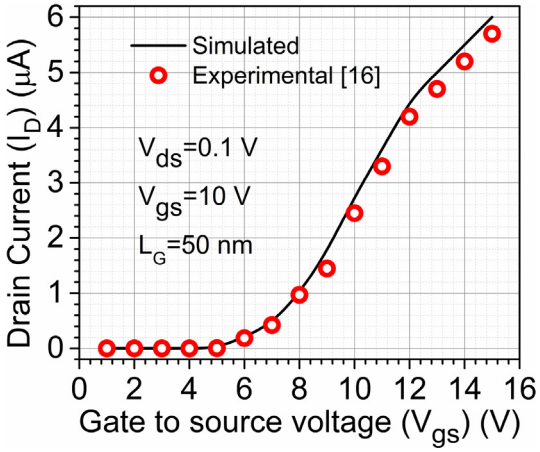


Fig. 2. Experimental and simulation of I_{ds}-V_{gs} characteristics of 50 nm gate length 4H-SiC Trench MOSFET.

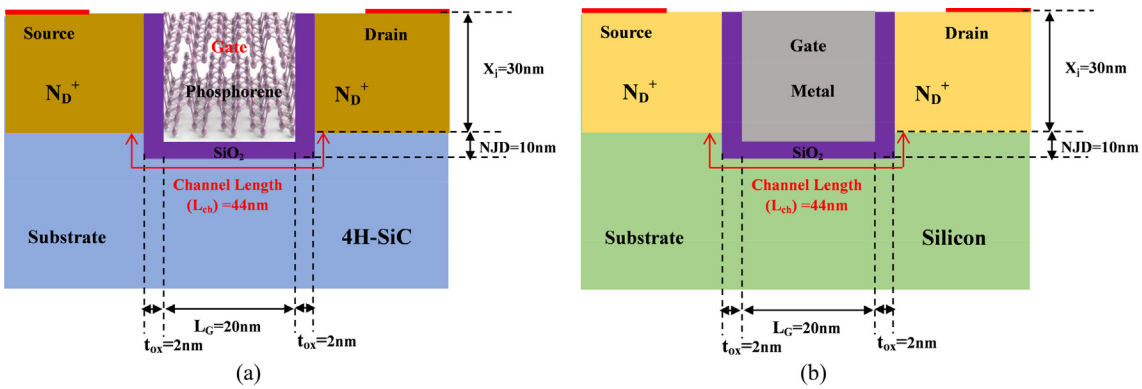


Fig. 1. Schematic structure of (a) 4H-SiC-RC-MOSFET and (b) CRC-MOSFET.

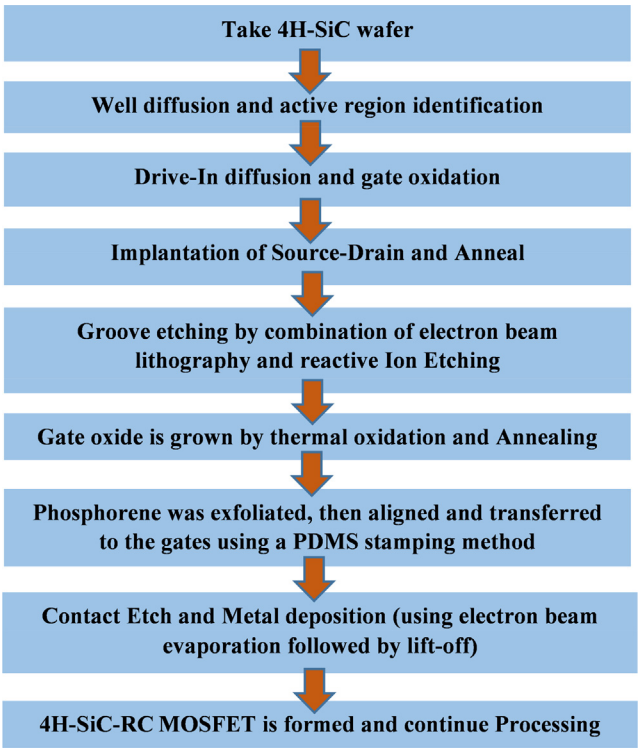


Fig. 3. Fabrication flowchart of BP-JL-RC-MOSFET.

5. Results and discussion

5.1. Analog performance assessment:

I_{ds}-V_{gs} characteristics (transfer characteristics) (reflected in Fig. 4(a)) of CRC, PS-RC, and 4H-SiC-MOSFET have been evaluated. Gate bias (V_{gs}) is 5.0 V and drain voltage (V_{ds}) is 0.2 V only have been taken in to account for the simulation and plotted in log scale to show the OFF-current (I_{off}). When the gate bias is applied from 0 to 5 V, then the drain current enhances significantly and reaches at the nearly same value of maximum drain current ($\sim 10^{-4}$ A), considered as in CRC, PS-RC, and 4H-SiC-MOSFET while I_{off} has a major difference in CRC($\sim 10^{-5}$ A), PS-RC ($\sim 10^{-5}$ A), and 4H-SiC-RC ($\sim 10^{-19}$ A) devices. Fig. 4(b) reflects the transconductance for CRC, PS-RC and 4H-SiC-MOSFET and it is observed that the peak value of transconductance is higher in the 4H-SiC device due to higher drain current as shown in Fig. 4(a). Moreover, due to the lower I_{off}, switching ratio is higher for 4H-SiC-RC-MOSFET ($\sim 10^{15}$) as compared CRC-MOSFET ($\sim 10^{2}$) and PS-RC-MOSFET ($\sim 10^{2}$) as shown in Fig. 5(a). Higher I_{on} and lower I_{off} make a device in which the gate controllability is more hence reduces hot carrier effects (HCEs) and short channel effects (SCEs). Thus, the device will be favorable for the switching application and IC design. The sub-threshold slope [19,20] has been evaluated at sub-20 nm gate length for CRC(151 mV/dec), PS-RC (160 mV/dec) and 4H-SiC-MOSFET (103 mV/dec) as shown in Fig. 5(a). It is a feature of transfer characteristics of MOSFET and related to the gate controllability of the device. As the gate controllability enhances in the 4H-SiC device, sub-threshold slope improves (lower in magnitude) which reflects that how the gate material plays a vital role in the device

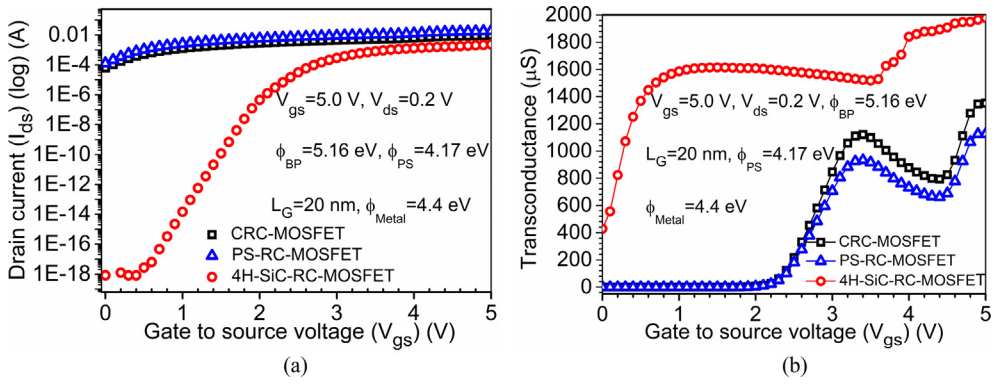


Fig. 4. (a): I_{ds}-V_{gs} characteristics and (b): g_m; of CRC-MOSFET, PS-RC-MOSFET and 4H-SiC-RC-MOSFET.

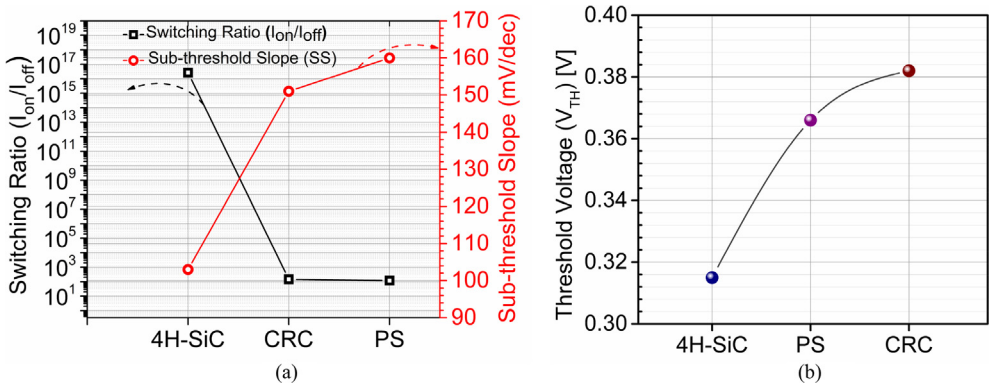


Fig. 5. (a) I_{on}/I_{off} ratio (left y-axis) and sub-threshold slope (right y-axis), (b) threshold voltage; of 4H-SiC-RC-MOSFET, CRC-MOSFET and, PS-RC-MOSFET.

performance and reducing the HCEs and SCEs. Moreover, the threshold voltage has been evaluated (shown in Fig. 5(b)) and observed that, 4H-SiC-RC-MOSFET has lower threshold voltage (0.31 V) as compared to CRC-MOSFET (0.3 V) and PS-RC-MOSFET (0.31 V) but the threshold voltage of PS-RC-MOSFET is less than CRC-MOSFET due to the same composition of polysilicon and silicon channel beneath the gate oxide.

Further, the surface potential of CRC, PS-RC, and 4H-SiC-RC-MOSFETs have been evaluated (shown in Fig. 6(a)) as the surface potential is the difference between the voltage at the surface of the MOSFET capacitor and the voltage in the bulk of it [21]. It is observed from Fig. 6(a) that, the surface potential of 4H-SiC-RC-MOSFET is higher than conventional devices (CRC and PS-RC) due to the higher electron mobility of BP material, used as a gate material. To maintain the charge neutrality, the surface potential must be high, and so, the capacitance will be low. It can also be observed from Fig. 6(b) that the lowering of the potential barrier is more near the source side as compared to drain side which indicated a reduction in Drain Induced Barrier Lowering (DIBL). The device performance phenomenon can be better understood by energy band diagram. So in order to explain device performance, energy band profiles have been plotted for CRC, PS-RC and 4H-SiC-RC-MOSFETs as depicted in Fig. 6(b). In the MOS structure, the energy band diagram contains the electron energy levels. Both the conduction band energy and valence band energy have been plotted in Fig. 6(b) for CRC, PS-RC, and 4H-SiC-RC-MOSFETs. From the results, it is evident that the 4H-SiC structure has more energy (magnitude in eV) in conduction band as well as valence band as compared to the conventional device structure.

Furthermore, electron mobility (μ) and electric field (E) have been evaluated as these two parameters are related to each other and for higher drift velocity (v_d), electron mobility and electric field both should be high. If one parameter reduces then the second parameter must be higher to maintain higher drift velocity. For the comparison of two different devices, electric field and electron mobility have been evaluated along the channel as reflected in Fig. 7(a) and (b) respectively. The electric field is observed and evaluated along the channel as shown in Fig. 7(a). In the source side, an electric field is more as compared to drain side of the channel region. Channel electric field is enhanced by 2.5 times in 4H-SiC-RC-MOSFET (25×10^5 V/cm) as compared to CRC-MOSFET (10×10^5 V/cm) and PS-RC-MOSFET (8×10^5 V/cm). The higher electric field is due to the high critical electric field of SiC and higher electron velocity since; electron mobility is high in the channel region of 4H-SiC-RC-MOSFET. Higher electric field and mobility of electron are few reasons for higher drain current, and thus the switching ratio is high. At the source side, the mobility of electrons is higher as compared to drain side for the 4H-SiC device (shown in Fig. 7(b)), but the impact of 4H-SiC is observed in the source/drain region and channel as well. Due to higher electron mobility of BP, the electron mobility is higher in the channel region and this improvement can also be observed in the source region as compared to CRC and PS-RC MOSFETs.

Further, the velocity of electron in the channel is evaluated for 4H-SiC-RC, PS-RC and CRC devices, and it is observed that in the 4H-SiC-RC device electron velocity is very high (7.5×10^8 cm/s) in the channel as compared to CRC ($\sim 5 \times 10^8$ cm/s) and PS-RC (4×10^8 cm/s) (shown in Fig. 8) due to higher electric field (as shown in Fig. 7(a)) as electron velocity is directly proportioned to electric field. Higher electron velocity results in higher switching ratio thus; 4H-SiC-RC-MOSFET is a suitable device for excellent switching applications.

Fig. 9 (contour plot) explores the variation of electron temperature [22] along the channel for 4H-SiC-RC, PS-RC and CRC

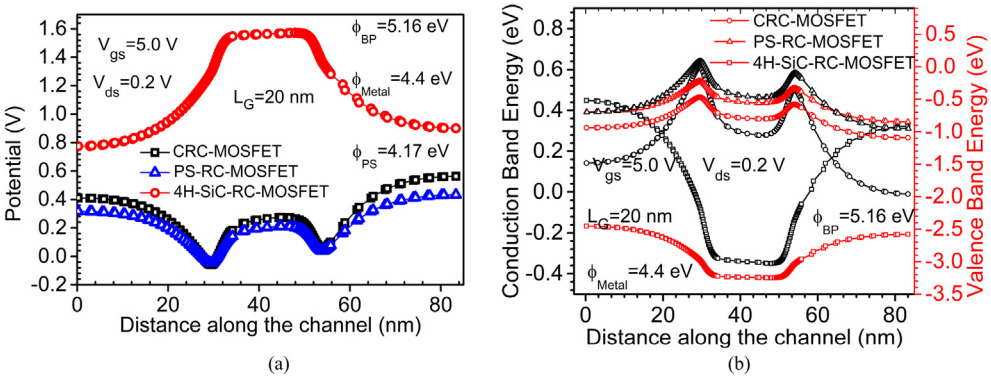


Fig. 6. (a) Surface potential, and (b) Energy band profile; along the channel for CRC, PS-RC and 4H-SiC-RC-MOSFET.

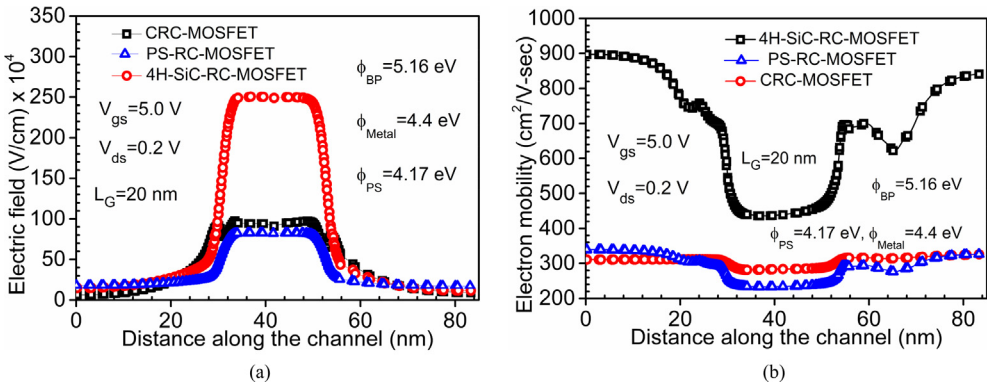


Fig. 7. (a) Electric field and (b) Electron mobility; along the channel for CRC, PS-RC and 4H-SiC-RC-MOSFET.

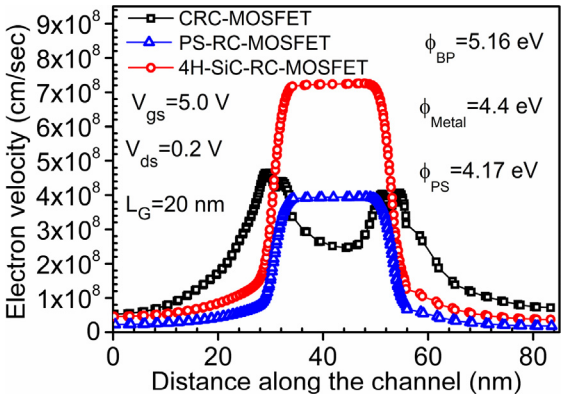


Fig. 8. Electron velocity along the channel for CRC, PS-RC and 4H-SiC-RC-MOSFET.

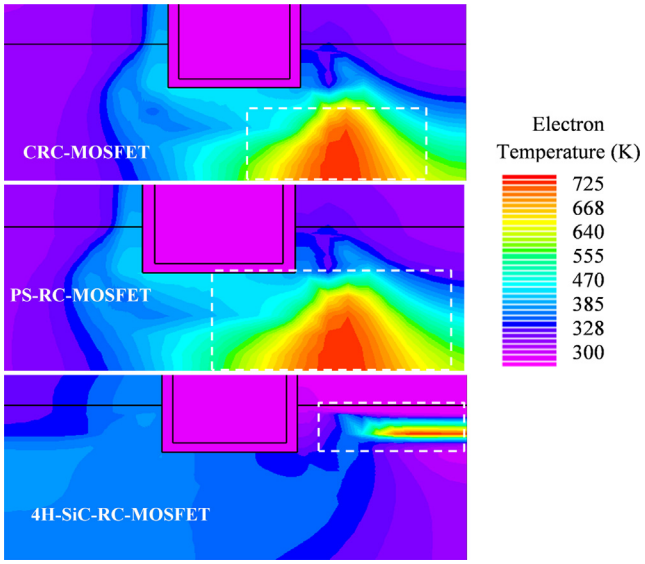


Fig. 9. Contour plots of Electron temperature for CRC, PS-RC and 4H-SiC-RC-MOSFET.

MOSFETs. It is observed that electron temperature changes from source to drain. At the source side electron temperature is slightly higher in 4H-SiC-RC (445 K) device than CRC (350 K) and PS-RC (400 K) meanwhile, it is observed that the electron temperature is very less in the drain side in 4H-SiC-RC (410 K) device while the conventional device and polysilicon gated device has higher electron temperature (520 K and 533 K) in the drain side. Due to lower electron temperature in the drain side leakage currents (such as impact ionization substrate current and a hot electron injected gate current) are eliminating (minimizing). Low electron temperature in the 4H-SiC-RC device is also due to a wide band of 4H-SiC and also due to black phosphorus gated material in the trench region.

From all the electrical characteristics of 4H-SiC-RC, PS-RC and CRC MOSFETs, it is observed that CRC device reflects improved performance (except threshold voltage) as compared to PS-RC device while both device are differ in terms of gate material only. Therefore, for further analysis CRC-MOSFET has been considered for the comparison with the 4H-SiC-RC-MOSFET.

5.2. RF performance assessment

For RF performance analysis of a MOSFET, power gains have a significant role in the designing of RF amplifiers, but meanwhile, it is observed that the primary standard for the designing of low-noise amplifiers (LNA) is stability. For an amplifier designing, oscillation should be neglected at the very high frequency. Stern stability factor (K) measures the stability of an amplifier, where the value of K usually <1 at low frequency and >1 at high frequency [23,24]. By using Eq. (1), K can be expressed as-

$$K = \frac{1 + |\Delta|^2 - |S_{11}|^2 - |S_{22}|^2}{2 \cdot |S_{12.} \cdot S_{21}|} \quad (1)$$

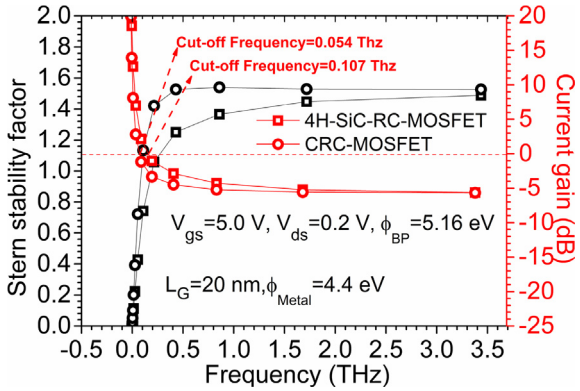


Fig. 10. Stern stability factor and cut-off frequency for CRC and 4H-SiC-RC-MOSFET.

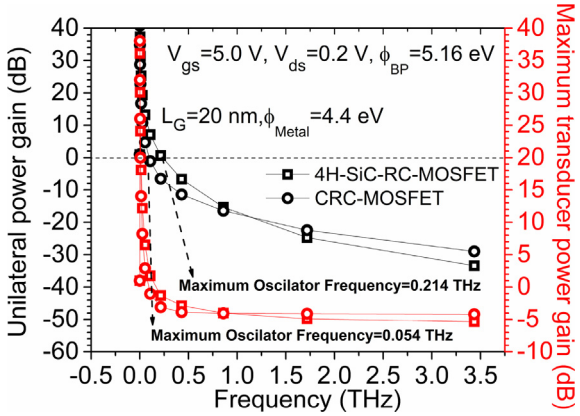


Fig. 11. Maximum oscillator frequency and Maximum transducer power gain for CRC and 4H-SiC-RC-MOSFET.

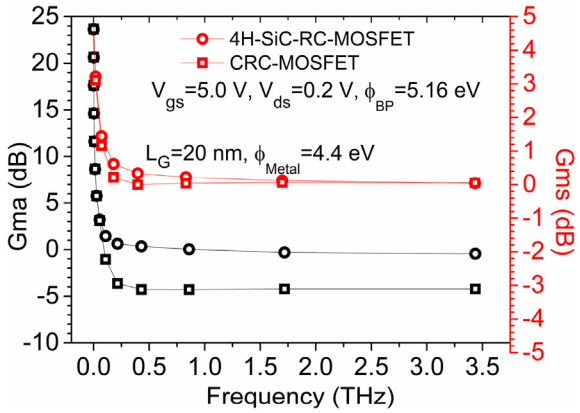


Fig. 12. Gma and Gms for CRC and 4H-SiC-RC-MOSFET.

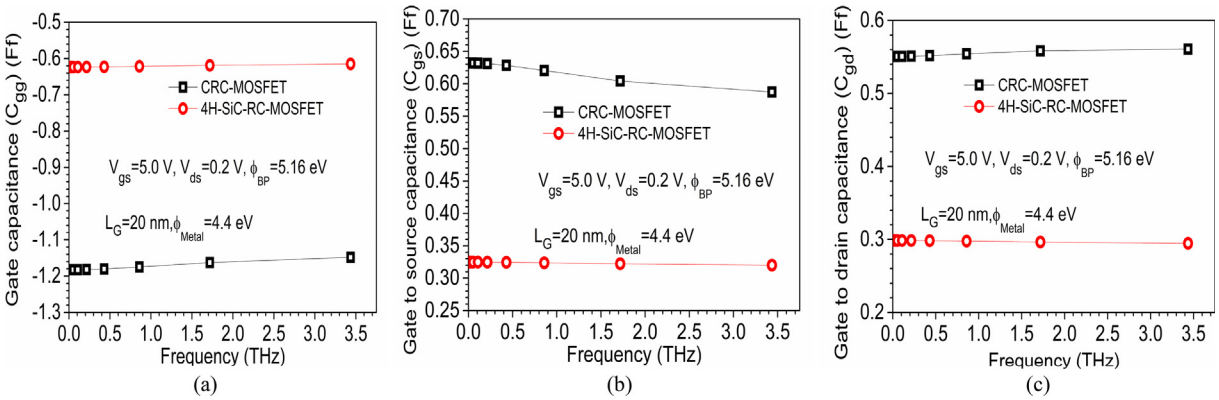


Fig. 13. (a): C_{gg}, (b) C_{gs} and (c) C_{gd} for CRC and 4H-SiC-RC-MOSFET.

where reflection coefficients are represented by S_{11} and S_{22} and transmission coefficients represented by S_{12} and S_{21}, and s-matrix Δ is defined as

$$\Delta = S_{11} \times S_{22} - S_{12} \times S_{21} \tag{2}$$

Stability factor (K) (Stern stability factor) has been calculated for 4H-SiC-RC and CRC MOSFETs as evident in Fig. 10. In the conventional device the value of $K > 1$ (at a higher frequency) while the proposed device has $K \sim 1$(slightly > 1 at a high frequency and <1 at low frequency). Thus, the 4H-SiC-RC device has candidature for RF applications.

Further, the cut-off frequency, f_T [25] has been calculated for 4H-SiC-RC and CRC MOSFETs by using Eq. (3) [26]. Basically for high-speed digital applications, f_T measures of speed and swing.

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})} \tag{3}$$

where g_m represents transconductance by and C_{gd} and C_{gs} represents stray capacitances. f_T is evaluated at unity current gain [27]. From Fig. 10, it is observed that f_T advances two times in 4H-SiC-RC-MOSFET as compared to the conventional counterpart. A higher value of f_T is achieved owing to higher I_{on} and improved g_m (as evident in Fig. 4(a) and (b)). Moreover, the evaluation of f_{MAX} is as important as f_T for higher speed of CMOS devices. In this work f_{MAX} has been evaluated by using Eq. (4) [27]:

$$f_{MAX} = \frac{f_T}{\sqrt{4R_g(g_{ds} + 2\pi f_T C_{gd})}} \tag{4}$$

f_{MAX} can also define when the unilateral power gain is unity [25]. Fig. 11 shows the higher f_{MAX} (~5 times) in 4H-SiC-RC-MOSFET owing to higher electron mobility in the channel (reflected in Fig. 7(b)) which decreases gate resistance (R_g) and hence improves f_{MAX}, as it is clearly depicted in Eq. (4) that f_{MAX} is directly related to g_m and inversely proportion to R_g [25]. Thus, the higher value of f_T and f_{MAX} reflect the capability of 4H-SiC-RC-MOSFET for RF applications.

In order to validate the candidature of proposed device for RF applications, power gains (called RF metrics) have been evaluated in terms of maximum available power gain (Gma) and maximum stable power gain (Gms) and compared with conventional counterpart. In a MOS device, Gma and Gms should as large as possible for high-frequency [28] applications. The ratio of maximum power available at load to the source defines Gma for a two-port network, and it is given by Eq. (5). Gma and Gms have improved at a higher frequency in 4H-SiC-RC-MOSFET as compared to the CRC device (as reflected in Fig. 12) due to the valuable electrical properties of silicon carbide.

$$G_{ma} = \frac{P_{load,\max}}{P_{source,\max}} \tag{5}$$

Furthermore, to enhance the device speed and to reduce the delay in logic-cell, the stray (parasitic) capacitances must be as low as possible [29]. At the input and output side, rise and fall times (mainly affects the speed of a digital switching circuit) of a digital device may increase due to the higher value of stray capacitances. The parasitic losses may cause low gain and oscillation in RF amplifiers. Thus, the investigation of these capacitances is more important for RF performance of a CMOS device. For 4H-SiC-RC and CRC MOSFETs, stray capacitances have been calculated in terms of the gate to source capacitance (C_{gs}), gate capacitance (C_{gg}), and gate to drain capacitance (C_{gd}) as depicted in Fig. 13(a), (b) and (c) respectively. It is clearly observed that the value of C_{gg}, C_{gs} and C_{gd} is very small (in femto-Farad (fF)) as compared to CRC device in the frequency range of few Hz to THz (shown in Fig. 13 (a), (b) and (c)) due to higher carrier mobility in 4H-SiC-RC-MOSFET (shown in Fig. 7(b)).

6. Conclusion

In this work, the efficacy of 4H-SiC material in a sub-20 nm device has been investigated with black phosphorus as a gate material on to the recessed gate. All the results have been investigated in terms of analog and RF FOMs at room temperature (300 K) and found that ON-current (2.3 mA) enhances and reduces OFF-current (~10^{-18} A) which improves the switching ratio to an order of 10^{16} thus, the analog performance of 4H-SiC-RC-MOSFET. Moreover, RF FOMs have also been evaluated in terms of f_T and f_{MAX}, and it has been observed that f_T (two times) and f_{MAX} (~5 times) are improved significantly. The enhanced power gains and significant reduction is observed in the proposed device which makes it a potential candidate for high swatching, analog and RF applications.

Acknowledgment

The author is thankful to Applied Physics Department, Microelectronics Research Lab, and Delhi Technological University to carry out this work.

References

[1] Chen L, Guy O, Jennings M, Igic P, Wilks S, Mawby P. Study of 4H–SiC trench MOSFET structures. Solid State Electron 2005;49(7):1081–5.
[2] Han K, Baliga B, Sung W. Split-Gate 1.2-kV 4H-SiC MOSFET: analysis and experimental validation. IEEE Electron Device Lett 2017;38(10):1437–40.
[3] Mahabadi SJ, Moghadam HA. Comprehensive study of a 4H–SiC MES–MOSFET. Physica E 2015;74:25–9.
[4] Murali R, Meindl JD. Modeling the effect of source/drain junction depth on bulk-MOSFET scaling. Solid State Electron 2007;51(6):823–7.
[5] Chen QB. Compact modeling of multi-gate MOSFETs for RF designs. 2013 IEEE international wireless symposium (IWS), 2013.

[6] Zareiee M. High performance nano device with reduced short channel effects in high temperature applications. ECS J Solid State Sci Technol 2017;6(7):M75–8.
[7] Zareiee M, Orouji AA. Superior electrical characteristics of novel nanoscale MOSFET with embedded tunnel diode. Superlattices Microstruct 2017;101:57–67.
[8] Zareiee M. A novel high performance nano-scale MOSFET by inserting Si3N4 layer in the channel. Superlattices Microstruct 2015;88:254–61.
[9] Kumar A, Tripathi MM, Chaujar R. In_2O_5Sn based transparent gate recessed channel MOSFET: RF small-signal model for microwave applications. AEU - Int J Electron Commun 2018;93:233–41.
[10] Kumar A, Tripathi M, Chaujar R. Reliability issues of In_2O_5Sn gate electrode recessed channel MOSFET: impact of interface trap charges and temperature. IEEE Trans Electron Devices 2018;65:860–6.
[11] Adhikari MS, Singh Y. A nanoscale dual-channel trench (DCT) MOSFET for analog/RF applications. Superlattices Microstruct 2015;88:567–73.
[12] Kumar A, Gupta N, Chaujar R. TCAD RF performance investigation of transparent gate recessed channel MOSFET. Microelectron J 2016;49:36–42.
[13] Lam K-T, Dong Z, Guo J. Performance limits projection of black phosphorous field-effect transistors. IEEE Electron Device Lett 2014;35(9):963–5.
[14] Das S, Zhang W, Demarteau M, Hoffmann A, Dubey M, Roelofs A. Tunable transport gap in phosphorene. Nano Lett 2014;14(10):5733–9.
[15] I. SILVACO, ATLAS User's Manual, Santa Clara, CA, Ver 5, 2011.
[16] Song Q, Zhang Y, Zhang Y, Tang X. 4H-SiC trench gate MOSFETs with field plate termination. Sci China Technol Sci 2014;57(10):2044–9.
[17] Appenzeller J, Martel R, Avouris P, Knoch J, Scholvin J, del Alamo JA, et al. Sub-40 nm SOI V-groove n-MOSFETs. IEEE Electron Device Lett 2002;23(2):100–2.
[18] Xiao-Hua M, Yue H, Bao-Gang S, Hai-Xia G, Hong-Xia R, Jin-Cheng Z, et al. Fabrication and characterization of groove-gate MOSFETs based on a self-aligned CMOS process. Chin Phys 2006;15(1):195.
[19] Zareiee M. Modifying buried layers in nano-MOSFET for achieving reliable electrical characteristics. ECS J Solid State Sci Technol 2016;5(10):M113–7.
[20] Mehrad M, Zareiee M. Improved device performance in nano scale transistor: an extended drain SOI MOSFET. ECS J Solid State Sci Technol 2016;5(7):M74–7.
[21] Arora N. MOSFET Models for VLSI Circuit Simulation Theory and Practice. New York: Springer; 1993.
[22] Zareiee M, Mehrad M. A reliable nano device with appropriate performance in high temperatures. ECS J Solid State Sci Technol 2017;6(4):M50–4.
[23] Pati SK, Koley K, Dutta A, Mohankumar N, Sarkar CK. Study of body and oxide thickness variation on analog and RF performance of underlap DG-MOSFETs. Microelectron Reliab 2014;54(6–7):1137–42.
[24] Dastjerdy E, Ghayour R, Sarvari H. Simulation and analysis of the frequency performance of a new silicon nanowire MOSFET structure. Physica E 2012;45:66–71.
[25] Sarkar A, Jana R. The influence of gate underlap on analog and RF performance of III–V heterostructure double gate MOSFET. Superlattices Microstruct 2014;73:256–67.
[26] Kumar A. Effect of trench depth and gate length shrinking assessment on the analog and linearity performance of TGRC-MOSFET. Superlattices Microstruct 2017;109:626–40.
[27] Nae B, Lazaro A, Iniguez B. High frequency and noise model of gate-all-around MOSFETs. In: Spanish conference on electron devices, 2009. CDE 2009. IEEE; 2009. p. 112–5.
[28] Minaei S, Yuce E. A simple CMOS-based inductor simulator and frequency performance improvement techniques. AEU-Int J Electron Commun 2012;66 (11):884–91.
[29] Mohapatra NR, Desai MP, Narendra SG, Rao VR. Modeling of parasitic capacitances in deep submicrometer conventional and high-K dielectric MOS transistors. IEEE Trans Electron Devices 2003;50(4):959–66.

Global Journal of Flexible Systems Management
https://doi.org/10.1007/s40171-018-0198-4

ORIGINAL RESEARCH

Analysing the Influence of Improved Situation, Capability Level of Actors and Flexible Process Workflow on Public Value of E-Governance Projects in India

Parul Jain Gupta[1] · **Pradeep Kumar Suri**[1]

Received: 2 June 2018 / Accepted: 26 September 2018
© Global Institute of Flexible Systems Management 2018

Abstract *An emerging approach for assessing performance of public services is based on public value. Many existing studies have, however, measured public value in the context of online services. In India, most of the public services are still not fully e-governed. Due to the absence of an end-to-end IT-enabled service delivery mechanism, physical presence of citizens is still required to avail public services in many cases. A citizen during his visit to a public organization faces a 'Situation', interacts with 'Actors' and follows a 'Process' or a set of processes. It is conceptualized that citizens' perception based on situation—actor—process (S–A–P)-related variables is expected to influence public value. In this article, S–A–P-related variables and their likely relationship with public value have been analysed based on a service of housing department of Delhi Development Authority. The study also aims to identify dominant actors who may be playing a critical role in the process of service delivery. It is expected that such actors may influence the service delivery time due to their dominance over the process. A dominance matrix showing ranking of different actors involved in the process of selected service has been presented by using an Interpretive Ranking Process method. The study is expected to benefit planners and implementers in terms of improved public value and delivery of public services on time.*

Keywords E-government · Interpretive Ranking Process · Public value · SAP–LAP

Introduction

Rewards of e-governance are well known to the governments all over the world. However, most developing countries are still lagging behind in terms of complete digitization of public services for making service delivery easier and transparent. For example, India's e-government development index continues to be much lower than many other countries (UN 2016) despite consistent increase in budgetary allocations and supportive policy-level changes towards effective e-governance.

During the last two decades, several initiatives have been taken to promote e-governance in India. These include a government directive issued to various departments in the year 2000 to implement 'Minimum Agenda of E-Governance'; enacting of IT act 2000 providing legal recognition to electronic transactions; setting up of the second Administrative Reform Commission in 2005; launching of National e-Governance Plan (NeGP) in May 2006 with thrust on setting up nationwide ICT infrastructure; identification of 31 mega e-governance projects to be implemented in mission mode and establishing thousands of common service centres as service delivery points near villages; constitution of Unique Identification Authority of India (UIDAI) in 2009 to provide unique identification to the residents of India and linking the same with various citizen centric government schemes; and launching of 'Digital India programme' during July 2015 with thrust on connecting rural areas with high-speed Internet networks for effective service delivery (www.meity.gov.in; Suri and Sushil 2017, pp. 3–6).

Views expressed by the authors are personal.

✉ Parul Jain Gupta
paruljain991@gmail.com

✉ Pradeep Kumar Suri
pksuri@dtu.ac.in

[1] Delhi School of Management, Delhi Technological University, New Delhi, Delhi, India

Published online: 15 October 2018

Efforts are being put in by several governments to provide benefits to citizens in terms of streamlining processes, citizens' empowerment, digitization of services, etc. through e-governance. It is well recognized that e-governance enables public organizations to build up their relationships with citizens through improved service delivery. The success rate of e-governance projects in the developing countries is, however, reported to be low (Heeks 2002; Gauld et al. 2010; Weerakkody et al. 2011). The poor success rate of e-governance projects has motivated researchers to analyse e-governance with regard to public value. Public value is a well-known concept to evaluate the public services (Moore 1995). Kelly et al. (2002) have explained public value as the value created by government via services, laws, regulations and other actions. In other words, it is a way to assess the performance of government services from citizens' perspective. Public services can be assessed by considering the creation of public value through various sources (Moore 1995; Kelly et al. 2002). The concept also facilitates decision-making as a performance measure of e-governance (Kearns 2004). Hence, public organizations should intend to create public value (Harrison et al. 2011). With the growing use of e-governance, government services are increasingly becoming citizen oriented. Therefore, it becomes necessary to analyse the execution of e-governance initiatives from citizens' viewpoint. Government organizations, therefore, are required to adopt the concept of public value while assessing e-governance performance from the viewpoint of citizens (Karunasena et al. 2011).

In the recent past, many studies have been undertaken to analyse the performance of e-governance projects. However, only a few studies have used the concept of public value to analyse e-governance. Existing studies suggest that public value of e-governance has still not been realized completely (Kearns 2004; Heeks 2008; Karunasena et al. 2011; Gupta and Suri 2017). This has motivated researchers to develop frameworks for assessing public value of e-governance.

As part of the literature review, it is realized that many studies are founded on measuring public value of online public services provided by the governments (Verdegem and Verleye 2009; Connolly et al. 2010; Bhattacharya et al. 2012; Karkin and Janssen 2014). However, in India, still in case of many government services, citizens have to visit public organizations. Examples include availing passport and driving license-related services, registration of property, registration of marriage, and conversion of property from leasehold to freehold. A citizen during his/her visit to a public organization comes across a situation characterized by queues to avail services, seating facility, availability of water for drinking, cleanliness, temperature, etc. He/she deals with employees of the public organization and observes their willingness to serve, speed of doing work, knowledge level, communication skills, availability at service counters, etc. The citizen also follows a process or a group of processes and experiences different levels of flexibility, continuity, simplicity and convenience of the processes. It is conceptualized that based on interplay of 'Situation', 'Actor' and 'Process' (S–A–P)-related variables (Sushil 2000), the citizen develops a perception about service delivery which is likely to influence public value of the public organization. In this study, it has been attempted to measure the public value of 'conversion of immovable property from leasehold to freehold' service of housing department of Delhi Development Authority (DDA). It is further attempted to analyse as to how the S–A–P-related variables influence public value.

In a service delivery process, citizens interact with various actors of a public organization. Different actors play different roles defined by the organization which are required in completion of the overall process. However, role of a few actors may be critical in terms of their relatively higher impact on the process. Hence, it is important to identify such actors for better decision-making with respect to their training, deployment, absenteeism, monitoring, appraisal, etc. to ensure smooth service delivery. Keeping this actor and process aspect into view, this study also attempts to rank all the involved actors in the process by using Interpretive Ranking Process (IRP).

The objectives of the present study are:

- to explore association between conceptualized S–A–P-related variables and public value in the context of e-governance
- to identify such dominant actors who may be playing critical role in the aspired realization of public value of an e-governance service

The article is presented in seven sections. After a brief introduction, Sect. 2 presents a review of the literature on public value. Section 3 explains methodology and theoretical framework for research. Section 4 presents a brief description of Delhi Development Authority (DDA) and one of its services, viz. 'conversion of property from leasehold to freehold'. Section 5 brings out situation–actor–process analysis in the study context which is followed by the discussion in Sect. 6. Section 7 presents learning–actions–performance synthesis. At last, Sect. 8 explains research findings and implications.

Literature Review

Public Value

Public value is referred as the measurement of public services. It can be evaluated through various sources such as

delivery of quality public services, achievement of outcomes and development of trust (Kelly et al. 2002). Based on this, Kearns (2004) proposed a framework as per which quality of public services is evaluated through level of information provision, level of e-government use, availability of choices, fairness, and cost savings. Golubeva (2007) proposed to assess public value with regard to quality of services offered by public organizations, public trust and public policy outcomes. A public value framework presented by Grimsley and Meehan (2007) has emphasized upon measuring public value by focussing on services, user satisfaction, trust, and outcomes. A theoretical framework in the context of developed countries is proposed by the European Commission (2006) consisting three drivers of public value, viz. efficiency, democracy and effectiveness. Jorgensen and Bozeman (2007) have reviewed studies in the context of countries, i.e. USA, UK and Scandinavian, and proposed seven categories of public value. These categories are transformation of interests into decisions, contribution of public sector to the society, association among public administrators and politicians, public administration's intra-organizational aspects, association among public administrators and their environment, behaviour of employees of public sector, and association among public administration and people. A study in the context of Sri Lanka has brought out a research framework to assess public value of e-governance based on three major factors, i.e. delivery of quality public services, effectiveness of public organizations, and achievement of socially desirable outcomes (Karunasena 2012).

On the basis of a review of the literature, public value is conceptualized in terms of three macrovariables, viz. 'Quality services delivery', 'Competence of public organizations', and 'Achievement of socially required outcomes'. The constituting microvariables are presented in Table 1.

'Quality services delivery' creates public value in terms of 'Quality of information', 'System functioning', 'User orientation', and 'Cost savings'. 'Quality of information' refers to the correct, updated, relevant, detailed, simple and understandable information (Golubeva 2007; DeitY 2008; Harrison et al. 2011; Suri and Sushil 2011; Bhattacharya et al. 2012; Papadomichelaki and Mentzas 2012; Osman et al. 2014). 'System functioning' comprises the easiness to fill up and submit the request form, easiness to pay for the services and processing of application without any technological issue (DeitY 2008; Bhattacharya et al. 2012; Papadomichelaki and Mentzas 2012; Satapathy 2014; Osman et al. 2014). 'User orientation' focuses on easy use for citizens having little or no formal education and facilities provided for differently abled persons, senior citizens, etc. (Golubeva 2007; DeitY 2008; Bhattacharya et al. 2012; Papadomichelaki and Mentzas 2012; Lindgren and Jansson 2013; Satapathy 2014; Osman et al. 2014). 'Cost savings' contributes to saving of time, money and efforts in terms of number of visits made by applicant, money paid to avail services and amount of efforts put in by the applicant (Kearns 2004; DeitY 2008; Belwal and Zoubi 2008; Bidyarthi and Srivastava 2011; Osman et al. 2014).

'Competence of public organizations' comprises the 'Efficiency'' 'Openness', and 'Responsiveness'. 'Efficiency' refers to fast service delivery, avoidance of duplicate tasks and increased reach to public organizations (Suri and Sushil 2006; Golubeva 2007; Deity 2008; Dash and Sangita 2011; Suri and Sushil 2011; Papadomichelaki et al. 2012; Suri 2014; Osman et al. 2014; Vela et al. 2015). 'Openness' can be measured through display of information related to public organization, for example, policy documents, expenses occurred, information related to agreements and tenders, display of other relevant information for example roles and responsibilities performed by employees, contact details of staff and number of working hours, office timings and addresses, fees (Golubeva 2007; Harrison et al. 2011; Papadomichelaki et al. 2012). 'Responsiveness' refers to response to inquiries, complaints and Right to Information (RTI) (DeitY 2008; Papadomichelaki et al. 2012; Lindgren and Jansson 2013; Satapathy 2014; Osman et al. 2014).

'Achievement of socially required outcomes' develops public value in terms of 'Equity', 'Trust', 'Self-development', 'Citizens' participation', and 'Concern for environment'. 'Equity' refers to the display of information in local language, same treatment with all applicants except people with special needs (DeitY 2008; Dash and Sangita 2011; Lindgren and Jansson 2013; Osman et al. 2014). 'Trust' is conceptualized as security and privacy of personal information (Kearns 2004; Belwal and Zoubi 2008; Dash and Sangita 2011; Papadomichelaki et al. 2012; Kalsi and Kiran 2013; Osman et al. 2014) and steps taken by public organizations to reduce the involvement of middlemen (DeitY 2008; Bidyarthi and Srivastava 2011; Bhattacharya et al. 2012; Satapathy 2014). 'Self-development' can be measured through awareness programmes conducted for knowledge promotion, training programmes conducted for non-Internet savvy citizens and access through common service centres or Kiosks (DeitY 2008; Kalsi and Kiran 2013; Navarro et al. 2014). 'Citizens' participation' is measured through ongoing updates provided on policies and processes, opportunities to take part in discussions, policies, decisions making and provisions for suggestions (Axelsson et al. 2010; Bidyarthi and Srivastava 2011; Satapathy 2014; Navarro et al. 2014; Vela et al. 2015). 'Concern for environment' is conceptualized with regard to reduction in paper printing and initiatives taken to save energy (Suri and Sushil 2011; Zhang et al. 2014; Suri 2014).

Table 1 Conceptualized variables to evaluate public value

Macrovariables		Quality service delivery (QSD)				Competence of public organizations (CPO)			Achievement of socially required outcomes (ASRO)				
Microvariables		Quality of information	System functioning	User orientation	Cost savings	Efficiency	Openness	Responsiveness	Equity	Trust	Self-development	Citizens' participation	Concern for environment
Public value studies													
1	Kearns (2004)				√					√			√
2	Suri and Sushil (2006)					√							
3	Golubeva (2007)	√		√		√	√						
4	DeitY (2008)	√	√	√	√	√		√	√	√	√		
5	Belwal and Zoubi (2008)				√					√			
6	Axelsson et al. (2010)											√	
7	Bidyarthi and Srivastava (2011)				√					√		√	
8	Harrison et al. (2011)	√					√						
9	Dash and Sangita (2011)					√			√	√			
10	Suri and Sushil (2011)	√				√							√
11	Bhattacharya et al. (2012)	√	√	√						√			
12	Papadomichelaki et al. (2012)	√	√	√		√	√	√		√			
13	Karunasena (2012)	√	√	√		√	√	√	√	√	√		√
14	Kalsi and Kiran (2013)									√	√		
15	Lindgren and Jansson (2013)			√				√	√				
16	Zhang et al. (2013)												√
17	Satapathy (2014)		√	√				√		√		√	
18	Suri (2014)					√							√
19	Navarro et al. (2014)										√	√	
20	Osman et al. (2014)	√	√	√	√	√		√	√	√			
21	Vela et al. (2015)					√						√	

Methodology

Selection of Project and Data Collection

A review of the literature was carried out to get deeper insights about public value and variables associated with situation, actors and processes. As stated above, a public service of DDA, viz. conversion of immovable property from leasehold to freehold, has been analysed for the study. Besides convenience, the selection of the service is based on the following criteria:

- The service should be citizen centric.
- It should have completed minimum 1 year.
- The service should be of state level.
- It requires the presence of applicants at public organizations for availing services.

To collect detailed information about the process of service delivery, three personal interviews were conducted during November 2016 to April 2017 with four senior officers at Delhi Development Authority. This was followed by study of several project-related documents such as office circulars, user manuals, application forms, yearly reports and other relevant information available on the website (www.dda.org.in). Meetings with personnel at different departments, i.e. IT department, housing department and Nagrik Suvidha Kendra (NSK), were also held in the month of April 2017 to get more insights about service delivery. This has helped in developing understanding about the transformation of processes and improvement in service delivery at different stages since the implementation of the service. It also facilitated to understand the applicability of conceptualized situation, actors and process-related variables (Gupta and Suri 2017) in the study context before conducting the field survey. A process flowchart has been prepared based on the knowledge gained during the visit at DDA (Exhibit 1.1). This has further helped in understanding the number of activities performed by each actor during the process (Table 5).

A survey was conducted in the month of April 2017 in Delhi city to collect responses from the beneficiaries. Citizens visiting Nagrik Suvidha Kendra for the execution of their conveyance deed were approached to fill up the questionnaire based on a five-point Likert-type scale. However, due to limited time, responses could only be collected from a few applicants initially. Hence, it was felt appropriate to seek contact details of such applicants who could not respond due to shortage of time so as to approach them later. In this manner, 72 applicants were approached out of which 26 applicants responded in person and 18 responded over phone. In all, 44 responses could be collected which is considered acceptable for this pilot study.

Based on the survey, univariate statistical analysis has been performed and is presented in Table 4. Descriptive statistics have been used to draw interpretations. Further, on the basis of understanding developed about S–A–P-related variables, L–A–P synthesis is performed (Table 6) explaining the processes, reasons of delay along with suggested measures to be taken for improving each process.

SAP–LAP Framework

As mentioned above, Situation–Actors–Process (S–A–P)–Learning–Actions–Performance (L–A–P) framework (Sushil 2000, 2009a, 2017a) has been used to conduct the case study keeping into consideration the thrust being given on transformation of working environment of DDA and significant role being played by employees in various processes aimed at serving the citizens. SAP–LAP framework is a comprehensive framework which has been widely used by researchers for case analysis in various areas of management such as resource management, supply chain management, technology management, strategy formulation, identifying planning-implementing gaps in the context of e-governance projects, etc. (Garg and Deshmukh 2010; John and Ramesh 2012; Charan 2012; Chauhan and Singh 2013; Mangla et al. 2014; Suri and Sushil 2017).

SAP–LAP framework provides generic view to deal with the problem context by considering three basic entities, viz. a 'situation' to be managed, an 'actor' or a group of actors to deal with the situation and a 'process' or a set of processes that respond to the situation and recreate it. These three entities and their relationships constitute S–A–P part of the framework in which freedom of choice exists with an actor. More freedom of choice with an actor enables him/her to keep the processes flexible and adapt to the emerging situations. On the contrary, lesser freedom of choice with the actor results into rigid processes and restricts his/her options to deal with the changing situation. Learning issues based on analysis of situation–actor–process interplay are used to suggest actions to cope up with the emerging situation and improve the overall performance (L–A–P) (Fig. 1).

A citizen who visits to a public organization also faces a situation in terms of basic facilities, hygiene factors, convenience, etc. During interaction with employees, he/she experiences their speed of work, willingness to help, knowledge level, etc. The citizen is also subjected to various processes in order to avail services. The processes encompassing government procedures generally lack flexibility (Singla and Singh 2018). It is conceptualized that situation, actor and process-related variables may be having an influence on public value of an e-governance service as perceived by the citizens. In this paper, S–A–P-related

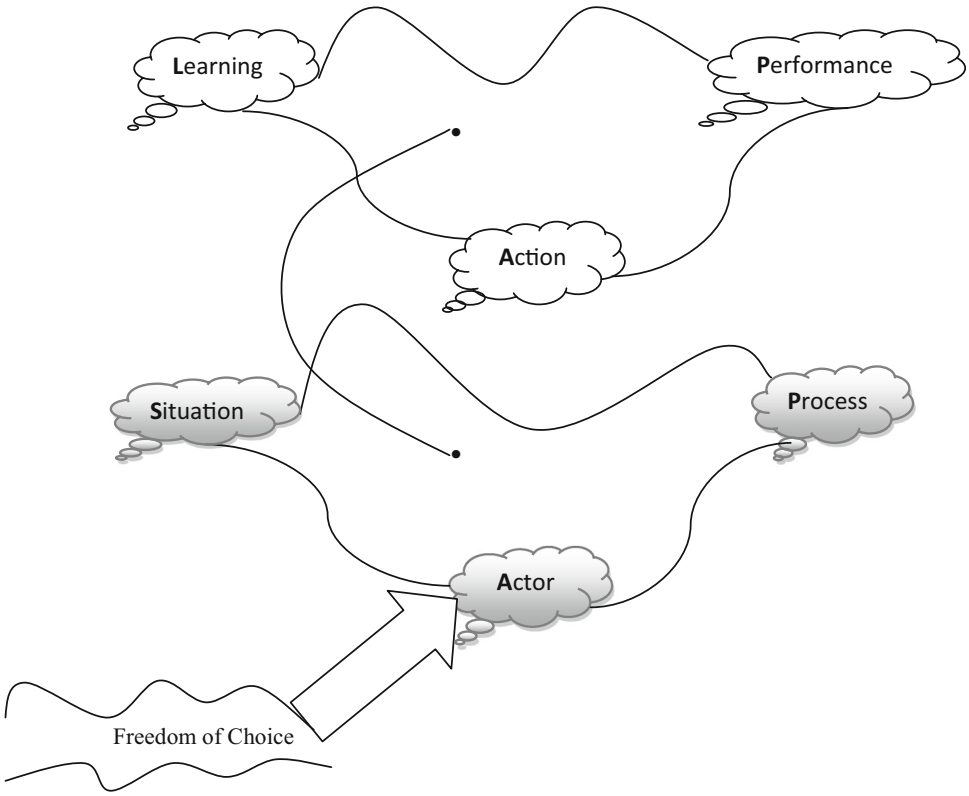


Fig. 1 SAP–LAP Framework (Source: Sushil 2000)

variables have been taken from a study conducted in the context of e-governance projects (Gupta and Suri 2017) as shown in Exhibit lll.1. Mapping of S–A–P-related microresearch variables with the literature is shown in Table 2.

A conceptual framework based on a review of the literature and SAP–LAP framework is presented below in the study context (Fig. 2).

Interpretive Ranking Process (IRP) Method

Interpretive Ranking Process (IRP) method (Sushil 2009b, 2017b) is used for decision-making by ranking available choices on the basis of paired comparisons and their interpretations. The method has been implemented in different areas such as manufacturing (Haleem et al. 2012; Sharma et al. 2016), supply chain management (Ware et al. 2014; Mangla et al. 2015), etc. Due to the paired comparisons, this method offers judgement convenience to the experts while deciding about importance of one variable over another. This is unlike other multi-criteria decision-making (MCDM) methods, for example, 'analytical hierarchy process' (Saaty 1980), where experts are required to rank n number of variables taken together due to which cognitive load on experts increases and they are also needed to devote more time. Hence, to handle large size problems consisting of many variables, an efficient IRP method (Sushil 2017b) is used.

In the present study, IRP method has been used to rank different actors involved in the process of conversion of immovable property from leasehold to freehold. The purpose to rank actors on the basis of their dominance is to identify the actors who have comparatively more influence over the processes. According to the process flowchart issued by housing department, total time taken to complete the process should be 45 days. However, as per the feedback received from citizens, mean value of variable 'Delivery of service is fast' under the microvariable 'Efficiency' is found to be only 0.25. This means that actual service time is much more than the timelines defined by DDA. In this situation, it becomes imperative to analyse the crucial role performed by various employees and their influence on total process duration. Therefore, with the help of IRP method, it has been attempted to identify actors who have more control over the process. This means that their availability, speed and skill level have implications on the overall process time. Performance of such indentified key actors can thus be closely monitored, and corrective actions can be taken in order to reduce the overall service time.

IRP methodology used in the present study is explained in brief as follows:

- In IRP, the first step is to recognize two groups of variables which include one group of those variables

Table 2 Conceptualized S–A–P-related variables and their mapping with the literature *Source*: Gupta and Suri (2017)

Macrovariables	Microvariables	Studies
Improved situation	Distance covered to reach service centre	Agus et al. (2007), DeitY (2008), Yusoff et al. (2008), Al-Borie and Damanhouri (2013) and Islam et al. (2014)
	Influence of middleman	DeitY (2008) and Satapathy (2014)
	Queue to avail services	Cavana et al. (2007), Agus et al. (2007), Chatzoglou et al. (2013) and Islam et al. (2014)
	Seating arrangements	Cavana et al. (2007), Kakouris and Meliou (2010), Ahmad and Latif (2012) and Islam et al. (2014)
	Air-conditioned environment	Ahmad and Latif (2012), Chatzoglou et al. (2013) and Islam et al. (2014)
	Cleanliness of work area and washrooms	Cavana et al. (2007), Agus et al. (2007), Yusoff et al. (2008), Kakouris and Meliou (2010), Al-Borie and Damanhouri (2013) and Islam et al. (2014)
	Drinking water	DeitY (2008) and Chatzoglou et al. (2013)
	Provision of feedback	Yusoff et al. (2008), Bharwani and Jauhari (2013), Satapathy (2014), Osman et al. (2014) and Suri (2016)
Capability level of actors	Communication skills	Parasuraman et al. (1988), Cavana et al. (2007), Yusoff et al. (2008), Ahmad and Latif (2012) and Bharwani and Jauhari (2013)
	Service orientation	Parasuraman et al. (1988), Cavana et al. (2007), Bharwani and Jauhari (2013), Al-Borie and Damanhouri (2013) and Islam et al. (2014)
	Fast execution of work	Cavana et al. (2007), Suri and Sushil (2011), Suri (2014), Osman et al. (2014) and Islam et al. (2014)
	Knowledge level about service delivery	Parasuraman et al. (1988), Cavana et al. (2007), Papadomichelaki and Mentzas, (2012), Bharwani and Jauhari (2013) and Al-Borie and Damanhouri (2013),
	Respect for ethics and values	Georgios (2004) and Salminen and Norrbacka (2010)
	Availability at service counters	Cavana et al. (2007), Kakouris and Meliou (2010), Ahmad and Latif (2012) and Bharwani and Jauhari (2013)
Flexible process workflow	Flexibility in date and time of application submission	Lovelock (1983), Parasuraman et al. (1988), Brown (1990), Kay and Russette (2000) and Bharwani and Jauhari (2013)
	Provision of uninterrupted services in case of technical fault	Kay and Russette (2000), Bharwani and Jauhari (2013) and Satapathy (2014)
	Option of both manual and online applications	Lovelock (1983), Brown (1990), DeitY (2008) and Suri (2014)
	Flexibility to apply from any service centre within the city	Lovelock (1983), Brown (1990), DeitY (2008)
	Provision to avail services without visiting service centre	Lovelock (1983), Brown (1990), Saxena (2005) and Satapathy (2014)

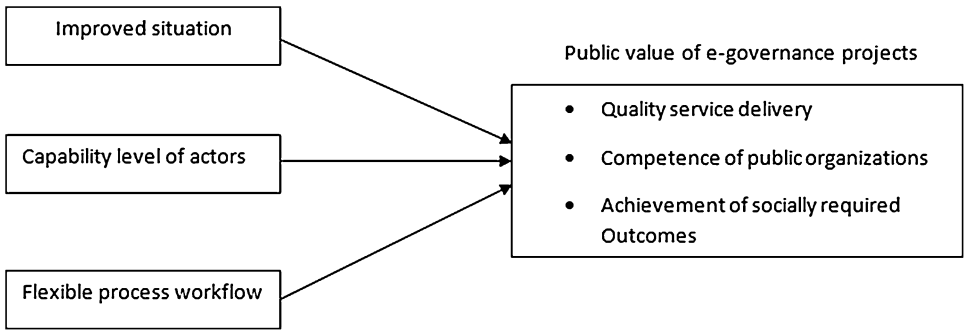


Fig. 2 Conceptual research framework based on S–A–P

that are required to be ranked and another group of reference variables. This means that one group comprises the options to be ranked and another group consists of criteria on the basis of which ranking is to be performed (Sushil 2009b). Therefore, in the study context, as we were required to rank different actors on the basis of their involvement in the overall process, ranking variables have been identified as actors (A1–

Table 3 Ranking and reference variables

Ranking variables—Actors
A1—Central Diarist
A2—Record room incharge
A3—Applicant
A4—Accounts officer of accounts branch
A5—Dealing assistant
A6—Assistant Director
A7—Deputy Director
A8—Director
Reference variables—Processes
P1—Receipt of conversion applications
P2—Transfer of conversion applications from centralized post
P3—Demand of files from record room and its receipt from record room
P4—Preliminary examination of documents and issuance of deficiency letter
P5—Receipt of reply along with documents from applicant
P6—Checking of documents and processing of file for in principal approval and reference to Accounts department
P7—Payment verification and No Objection Certificate (NOC) from accounts department
P8—Receipt of file from accounts department and processing for approval of conversion application

A8) and reference variables have been identified as processes (P1–P8). In the present study, ranking variables and reference variables as shown in Table 3 have been derived from the flowchart defined by DDA to execute the service delivery process (Table 6).

- In the second step, contextual relationship between ranking and reference variables is defined and a cross-interaction binary matrix viewing their relationship is prepared. In the study context, cross-interaction binary matrix is presented in Exhibit ll.1.
- Third step is to convert the binary matrix into interpretive matrix as presented in Exhibit ll.2. This is carried out by interpreting '1' entries of the binary matrix based on the contextual relationship of ranking and reference variables.
- Further, experts are approached based on identified paired comparisons as per interpretive matrix and interpretive logic is listed based on their judgement for each paired comparison as shown in Exhibit ll.3.
- It is then followed by preparing dominating interaction matrices of all actors with respect to each process. In the present study, the same is shown in Exhibit ll.4.
- Then an overall dominance matrix is prepared by adding all the dominance matrices for each process. In the present study, consolidated dominating matrix reflecting ranking of actors with respect to all the processes is shown in Exhibit ll.5.

An actor with the highest dominance is given first rank which reflects the importance of that actor in the whole process in terms of dependence on the actor. In the study context, actor A5, i.e. dealing assistant, is found to be dominating with rank 1. An efficient IRP method reduces the dependence on interpretive comparisons by considering implicit and transitive dominance (Exhibit ll.6). An implicit dominance is dominance of '1' over '0' in a cross-interaction binary matrix, and an implicit non-dominance is considered to be when both alternatives have '0' entries. As per an efficient IRP method, such cases are not presented to experts due to their implied nature. Also, in case of three alternatives having '1' entry each, experts are asked for two interpretations only as one interpretation is implied which is known as transitive dominance. During the study, we identified eleven interpretive dominance comparisons, for which three senior officials at DDA were approached in the month of July 2017. Interpretation based on the application of IRP method in the study context is presented in Sect. 5.2.

Case Description

About Housing Department of Delhi Development Authority

The Delhi Development Authority (DDA) was established as per the provisions of Delhi Development Act 1957 to encourage and protect the planned development of Delhi. Major functions performed by the DDA are acquisition and disposal of lands and properties, development of sports infrastructure and shopping places, launching housing schemes for residential flats and houses, allotment and payment management, conversion of leasehold immovable

property into freehold, etc. Housing department is the main branch of the DDA which was started in the year 1968 mainly to construct and provide houses to the society as per the requirements. Its roles include allotment of residential flats through housing schemes, conversion of leasehold property into freehold property, execution of conveyance deed, restricting unauthorized construction and misuse of property, etc.

Starting from 1969, 44 housing schemes were launched and about 4,20,510 flats of various categories allotted till October 2015 (www.dda.org.in). All the residential flats constructed prior to the year 1992 were allotted by the DDA on leasehold basis. The purpose of DDA to make allotment on lease basis was to retain the ownership of the property and receive recurring revenue from the allottees for supporting other developmental activities. However, after the lease period, it was found difficult by DDA to take over the property. Hence, due to growing number of litigations, in February 1992, DDA decided to convert properties from leasehold to freehold by executing conveyance deed.

Implementation of Leasehold to Freehold Scheme

As per the scheme, allotments made by DDA prior to 1992 are permitted for conversion from leasehold to freehold. However, option of acquiring property on hire purchase basis was still continued even after 1992. Due to this, property could not be allotted on freehold basis. Allotment of properties completely on freehold basis was started from 2010 when hire purchase option was excluded from the scheme. Hence, majority of the allotments made before 2010 were on leasehold basis only. During detailed discussion with senior officers at DDA, it is learnt that out of the total allotted flats, only one-third of flats have been converted from leasehold to freehold till 2017. To expedite the conversion process, directions were issued to the housing department in 2011 to complete the process of a conversion case in 45 days which used to be 90 days earlier.

Situation–Actor–Process Analysis

In this section, it has been attempted to analyse the situation, key processes and associated actors of housing department of DDA to present a holistic view of the study context.

Situation

Housing activities were taken up by the DDA in the year 1968. For several years, the functioning of housing department was characterized by manual processes and physical record maintenance. In the year 1992, government decided to computerize the records of all allotments made by the DDA under various schemes. As per discussion held with the senior officers of DDA in April 2017, only 50–60% of the data could be computerized till 2016. The department is to complete the computerization of all manual records by 2018.

Apart from the ongoing processes, a few other initiatives which were taken by the DDA towards computerization are as follows:

- In November 2005, it was decided to receive applications at a counter of housing department by giving computer generated receipts which used to be accepted in a government bank earlier. In the year 2005–2006, information kiosks were set up at the DDA offices. The website of DDA was also revamped in accordance with Right to Information Act 2005.
- In April 2007, as a major step towards e-governance, an Integrated Management Information System (iMIS) was launched which aimed at automating all the involved processes. The iMIS was expected to improve citizen interface, enable easy access to information and introduce paperless processing.
- In September 2012, a web-based interface was launched to apply online for conversion of immovable property from leasehold to freehold. The service was started to facilitate easy tracking of applications and reduce the number of visits made by applicants.

Although considerable efforts were put by introducing IT-based solutions, the housing department could not achieve the desired results. Citizens continued to face issues related to an un-friendly user interface and were required to visit the DDA office frequently. Hence, four Nagrik Suvidha Kendras (NSKs) were opened in the month of June 2014 as part of re-engineering of government processes to facilitate receipt of applications for conversion of property from leasehold to freehold.

Actors and Process interfaces

An actor–process matrix, based on the study of steps involved in the conversion of immovable property from leasehold to freehold, is presented in Exhibit 1.1. As presented in Exhibit 1.1, minimum 18 steps are required to be performed by actors. It depicts a detailed explanation of each and every step required during the process. However, as per the flowchart defined by the housing department, total eight steps are involved which are performed by mainly eight actors. As this flowchart (Table 6) is defined by DDA to monitor the days taken at each step, it has been used for the purpose of IRP analysis. Table 3 shows the list

of actors (ranking variables) and processes (reference variables) for further interpretations.

On the basis of ranking and reference variables, cross-interaction binary and interpretive matrices are prepared (Exhibit ll.1 and Exhibit ll.2) followed by interpretive logic of paired comparisons (Exhibit ll.3) and dominance matrices (Exhibit ll.4 and 5). In all, there are 117 dominances as shown in Exhibit ll.5. As per the IRP analysis, dealing assistant is found at the highest rank with 20 net dominances followed by Assistant Director and applicant with eight and three net dominances, respectively. Hence, it would be appropriate to state that dealing assistants are playing most critical role in the process of conversion of immovable property from leasehold to freehold. This needs to be kept in view while taking corrective measures to make the processes more flexible and effective for improved public value.

Further, to validate the results obtained through IRP method, count of activities performed by each actor (Table 5) are compared with dominance matrix (Exhibit II.5). Table 5 is prepared on the basis of understanding of processes and involvement of actors during visits conducted to DDA (Exhibit 1.1). It shows that the highest number of activities is performed by dealing assistant, i.e. eight. However, it reflects that Applicant is at the second position with seven activities followed by Assistant Director with six activities which are different from results received through IRP. Therefore, discussions were held with senior officers at DDA in the month of July 2018 to understand the difference between level of dominance of an applicant and an Assistant Director. During the discussion, it was informed that as per a recent development, visits performed by applicants have been reduced by executing activities fifteen to seventeen (Exhibit 1.1) on the same day. It means that now total visits required by an applicant have reduced from seven to five. Therefore, levels of dominance of top three actors obtained through IRP were found to be similar with observation made at DDA along with feedback received from the experts.

Analysis and Discussion

A univariate statistical analysis presented in Table 4 shows the values of range, mean and standard deviation of public value and variables related to S–A–P in the study context. Also, a list comprising macro- and micro-study variables with their observed mean values is presented in Exhibit lV.1. The noticeable gap between minimum and maximum observed mean values of variables related to S–A–P may be due to the complete or partial involvement of the applicants in the process. In a few cases, applicants take help of private agents to save their time due to which they do not get completely involved with the process, and thus, they do not experience S–A–P-related variables. However, applicants performing the tasks at their own get fully involved and experience all the S–A–P-related variables.

The small value of standard deviation linked with all the study variables indicates the closeness of the data to the observed mean values. The overall observed mean values of both S–A–P-related variables and public value are below the average value which reflects the likely positive relationship between them. Detailed interpretation of observed mean values of variables related to S–A–P and public value is explained below.

Interpretation of Linkage Between Public Value and S–A–P-Related Variables

The overall observed mean value of S–A–P-related variables is found to be 0.45. The macrovariable ‘Flexibility of process workflow’ has the lowest mean value which is due to the lesser mean values of the constituting variables ‘Provision to avail service without visiting to service centre’ and ‘Option of both manual and online applications’ which are found to be 0.03 and 0.27, respectively.

It has also been found that mean value of variable ‘Provision to avail service without visiting to service centre’ is lowest compared within the other constituting variables. This is expected to be so due to the requirement of

Table 4 Univariate statistical analysis of conversion of immovable property from leasehold to freehold scheme

Variable	*N*	Range	Minimum	Maximum	Mean	Std. Deviation
Public value	44	0.36	0.31	0.67	0.48	0.09
Quality service delivery	44	0.42	0.29	0.71	0.52	0.1
Competence of public organizations	44	0.64	0.17	0.81	0.47	0.15
Achievement of socially required outcomes	44	0.46	0.25	0.71	0.44	0.1
S–A–P-related variables	44	0.38	0.24	0.62	0.45	0.08
Improved situation	44	0.7	0.2	0.9	0.5	0.14
Capability level of actors	44	0.46	0.33	0.79	0.53	0.1
Flexibility of process workflow	44	0.5	0.1	0.6	0.32	0.14

Table 5 Count of activities performed by actors

Name of actors	Central Diarist	Record room incharge	Applicant	Accounts officer	Dealing assistant	Assistant director	NSK Staff	Bank employees	Computer department	Witnesses	Sub registrar
Count of activities per actor	2	1	7	5	8	6	3	2	1	2	3

multiple visits by applicants to avail the service. As per Exhibit l.1, an applicant is required to visit the service centre seven times for the execution of conveyance deed. Table 5 shows the count of activities required to be performed by all the actors involved in an ordinary process for the conversion of immovable property from leasehold into freehold.

Other variables characterized with the lower mean value associated with their respective macrovariables are 'Fast execution of work' and 'Provision of feedback mechanism'. However, variable 'Proper sitting arrangements' is observed to be having the highest mean value among all the variables constituting macrovariable 'Improved situation'. It has also been observed during analysis that employees at public organization in the study context have satisfactory knowledge about service delivery.

Mean value of the public value is also observed to be below average mean value, i.e. 0.48. Among the constituting macrovariables, 'Achievement of socially required outcomes' has the lowest mean value which may be due to the low level of participation of citizens in policy formulation and decision-making, non-sharing of regular updates on policies and procedures, the absence of taking suggestions from citizens, and lesser emphasis given on conducting awareness and training programmes for enhancing knowledge of citizens. Also, lower mean value of variable, viz. 'Lesser visits required to avail service' is due to the requirement of multiple visits by applicant as shown in Table 5. Observed mean value of the variable 'Delivery of service is fast' under the microvariable 'Efficiency' is found to be only 0.25. It may be due to the delay in service delivery as compared to the officially specified time limits. The causes of such delays and suggested measures in this respect are also explained in Sect. 7 through a synthesis of learning on the basis of S–A–P analysis.

The relatively high mean value of 'Quality service delivery' is expected due to the quality of information provided and low system interruption faced by applicants during the process. It may also be kept in view that the reason for less number of system glitches faced could be due to the preference for manual execution of most of the tasks at the service centre. The lowest mean value of the microvariable 'Cost savings' among all microvariables may be due to the more number of visits required by applicants to avail services. This is expected to be so because of lesser flexibility of the process resulting into requirement of multiple visits by applicant at the public organization. Studies have found that it has become crucial for organizations to incorporate flexibility in strategic processes for better performance (Sharma et al. 2010; Bishwas 2015; Haldar et al. 2016). A flexible process workflow enables public organizations to fulfil the requirement of customers without conducting multiple visits (Kumar et al. 2018).

The pilot study in the e-governance context suggests that it is expected that variables based on situation, actors and processes may be influencing public value of public organizations. Hence, the following conceptual research framework is proposed for further testing by conducting a comprehensive study of a few important citizen centric e-governance projects (Fig. 3).

HA1 An improved situation of service centre influences the public value of public services which are not fully e-governed

HA2 Capability level of service centre actors influences the public value of public services which are not fully e-governed

HA3 Flexible process workflow at service centre influences the public value of public services which are not fully e-governed

The respective null hypothesis for HA1, HA2 and HA3 are:

H01 An improved situation does not influence the public value of public services which are not fully e-governed

H02 Capability level of service centre actors does not influence the public value of public services which are not fully e-governed

H03 Flexible process workflow at service centre does not influence the public value of public services which are not fully e-governed

L–A–P Synthesis

Based on the study of Situation–Actors–Processes (S–A–P), Learning–Actions–Performance (L–A–P) synthesis is performed and is presented in Table 6. For this purpose, a

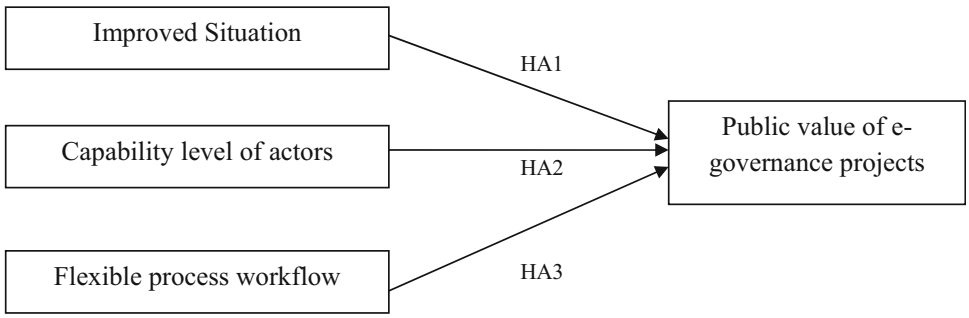


Fig. 3 Proposed research hypotheses

process chart issued by the DDA has been used which was prepared to define average time required at each stage to complete the process. As per this chart, directions were given to the housing department to complete the process of leasehold to freehold of property within 45 days from the filing of an application. However, as per the result of survey, observed mean value of the corresponding variable, viz. 'Delivery of service is fast', is only 0.25. It reflects that applicants in general are not able to get the services within specified time limits. Hence, it is crucial to synthesize learning based on S–A–P analysis which may be acted upon for improving the public value.

Research Findings and Implications

Research Findings

E-governance is increasingly being adopted by public organizations across the world. In India, a few e-governance projects have been running successfully and many are at various stages of implementation. Many of the public services are not yet fully e-governed and citizens visiting to these public organizations for seeking services such as issuance of ration card, birth certificate, election card, etc. have to go through processes which are only partially computerized. They face a 'Situation' during their visit, interact with 'Actors' and are subjected to a 'Process'. It has been conceptualized that their experience based on S–A–P-related variables may influence the perceived public value of respective public organizations. To execute the study in this context, a public service 'conversion of immovable property from leasehold to freehold' at Delhi Development Authority was analysed. A thorough understanding of 'Situation', 'Actors' and 'Process'-related variables and field survey helped to conduct univariate analysis.

Findings of the study reflect the existence of likely relationship between S–A–P-related variables and public value. Observed mean values of both public value and S–A–P-related variables in the study context were found to be below 0.50 which reflect that a public organization which lacks in terms of improved situation, efficient actors and flexible process flow as conceptualized in the study context may be characterized with lesser public value. Observed mean value of a few public value variables, viz. 'Lesser visits required to avail service' (Osman et al. 2014) and 'Delivery of service is fast' (Dash and Sangita 2011; Suri 2014), were found to be low due to requirement of multiple visits by applicant and delay in service delivery. Reasons of lower mean values of both the variables have been further analysed and are shown in Tables 5 and 6. Lesser observed mean values of a few S–A–P-related variables, viz. 'Fast execution of work' (Cavana et al. 2007; Suri and Sushil 2011; Suri 2014) and 'Provision to avail service without visiting to service centre' (Lovelock 1983; Brown 1990; Satapathy 2014), also indicate the delay in services and requirement of repetitive visits by applicants.

While exploring deeper into the process flow of 'conversion of immovable property from leasehold to freehold' public service, it has been found that total duration of the complete process officially stated by DDA is 45 days. However, lower observed mean value of respective conceptualized variables indicates overall delay in process. In order to identify actors performing crucial roles during the process, an Interpretive Ranking Process (IRP) (Sushil 2017b) has been used which helped to identify three dominating actors, viz. dealing assistant, Assistant Director, and Applicant which may play critical role in reducing the overall duration of service completion. It is expected to be so due to the highest level of involvement of these three actors at various steps in the overall process of conversion of immovable property from leasehold to freehold. The following section brings out implication of research findings.

Research Implications

The article has implications for both practitioners and researchers:

Table 6 Learning–Actions–Performance (L–A–P) synthesis

Sr. no.	Process	Branch/ officer	Average time	Process description	Reasons for delay	Measures to be taken
1	Receipt of conversion applications	Central diary	–	The moment applicant submits the request at NSK; it gets reflected in the internal portal of DDA	This process is automated and requires no manual intervention for execution. Hence, delay is not applicable here	A close watch should be kept on number of requests received on daily basis for timely transfer of applications to different branches
2	Transmission of conversion applications from central post	Concerned branch	2 days	A person from concerned branch collects physical applications from central diary and transfers applications from common pool to the concerned branch through internal portal	The designated person for this task is also responsible for retrieval of physical files from record room which is a time taking activity. In his absence, other person does retrieval, hence, transfer gets delayed	As this activity does not consume much time, it should be completed every day. By transferring files to the concerned branch, actual pendency can also be known and plan can be prepared for further action
3	Requirement of files from record room and its receipt	Concerned record room incharge	3 days	After receiving of application, it is required to retrieve physical main file from the record room	Record of main files is still not fully computerized; hence, a large number of physical files are stored in the record room at DDA central office. It requires a lot of efforts and time to retrieve physical files from the record room	All the records should be computerized for easy access and quick action to improve service delivery
4	Initial examination of documentations and issuance of deficit letter	Concerned branch	5 Days	In this process, documents submitted by applicant along with application and requirement of documents as per the main file are matched. In case of an additional requirement, a deficiency letter is sent to the applicant	As there is no computerized record of old files, every case requires repetitive efforts and time	Record of physical files should be maintained electronically along with the requirement to convert the property from leasehold to freehold. It would help to match the documents submitted by applicant and raise notice for any further requirement in lesser time
5	Receipt of reply along with documents from applicant	Applicant	7 Days	Once deficiency letter is issued, applicant is asked to submit the required documents within specified time	Sometimes documents are submitted late by the applicant. This may be due to the slowness of postal communication or lack of understanding about the required documents	Communication should be sent to the applicants electronically to avoid delay. SMS service can also be used to receive quick response from applicants
6	Verification of documents and processing of file for principal approval and referred to accounts departmen	Concerned branch	7 Days	After receiving all required documents, files are shared with accounts department for further processing	Once the documents are received, same are attached with the main file and complete file is shared with accounts section. In a few cases, actual requirement is overlooked and a new requirement may have to be raised from the applicant which causes delay in the process	Regular training programmes should be conducted on documentations and employees should be updated on the latest circulars issued by the government. This would help in bringing uniformity and avoiding duplication of work

Table 6 continued

Sr. no.	Process	Branch/ officer	Average time	Process description	Reasons for delay	Measures to be taken
7	Payment verification and No Objection Certificate (NOC) from accounts department	Concerned accounts officer of accounts branch	12 Days	NOC is issued by accounts department after verifying the receipt of payment made by applicant	Payments made by applicant with application and actual dues calculated by accounts department are compared, and in case of deficiency, applicant is asked to pay the difference. Also, an amount of more than rupees five thousand is cross-verified with the computer department to get confirmation of the payment	Dues calculated by applicant online or at NSK should be same as actual requirement. However, in case of any additional requirement, applicant should be given an option to make the payment electronically which would be useful for verification purpose also
8	Receipt of file from accounts department and processing for approval of conversion application	Dealing assistant	3 Days	After receiving of NOC from accounts department, file is processed for further approvals	Delay may be due to the rechecking of complete file before forwarding the same for further approvals	In order to save time and efforts, a checklist should be attached in each physical file for an easy reference
	Do	Assistant director	2 days	Approval of Assistant Director	It may be due to requirement of physical documents verification at each level	Approval through electronic mode along with automatic access to previous comments should be enabled for reducing process time
	Do	Deputy Director	2 days	Approval of Deputy Director	Same as above	Same as above
	Do	Director	2 days	Approval of director	Same as above	Same as above
		Total	45 Days			

Implications for Practitioners

The study reflects upon significance of analysing the e-governance initiatives in terms of their contribution to public value. The study findings are expected to benefit both planners and implementers while planning for new projects or analysing the existing ones to improve the level of public value and deliver public services within stated timelines. The planners of public services in respective organizations are expected to keep provision in the project plans for periodic assessment of e-governance initiatives in terms of constituents of public value, viz. 'Quality service delivery', 'Competence of public organizations', and 'Achievement of socially required outcomes'. In view of the relationship found between situation, actor and process-related variables and public value, practitioners need to be conscious of improving situation at service delivery locations, deploying skilled personnel for service and designing flexible processes equipped with enough options so as to avoid dependence on certain dominant actors. The findings further suggest that ranking of actors in terms of their dominance over process workflow shall be useful for implementers in order to effectively monitor and control public service delivery.

Implications for Researchers

The study has focused on analysis of an e-governance service where the beneficiaries are required to physically visit the concerned public organization. It has been attempted to capture the experience of beneficiaries as they get subjected to interplay of situation, actors and processes while availing such a partially computerized public service. The pilot study opens a window for analysing a few public services of similar nature in order to validate the proposed constructs, viz. 'Improved situation', 'Capability level of actors', 'Flexibility of process workflow' and 'Public value of e-governance'. The proposed hypotheses which have emanated from this study can be empirically tested by studying a few more projects. In Indian context, partially computerized public services which can be explored further are registration of properties, issuing of driving

license, issuing of marriage certificate, etc. The study further demonstrates use of a novel approach based on IRP method which can be applied for analysing the actor–process interfaces to identify and rank dominant actors who may be influencing public value of e-governance.

Limitations

The article is based on a pilot study in the context of one public service which is not fully e-governed. In Indian context, there are many such partially computerized public services which need to be studied to validate the proposed framework and bring out generalized findings.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix 1

Exhibit 1.1: List of activities performed by Actors

Sr.no.	Description of process	Central Diarist	Record room incharge	Applicant	Accounts Officer	Dealing Assistant	Assistant Director	NSK Staff	Bank Employees	Computer Department	Witnesses	Sub Registrar
	Count of activities per actor	2	1	7	5	8	6	3	2	1	2	3
1	Submission of application at NSK by applicant			√				√				
2	Receiving of files by Diarist from NSK Staff	√						√				
3	Diarist handover files to Dealing Assistant	√				√						
4	Record room incharge retrieve main file from the record room and handovers to Dealing Assistant		√			√						
5	Dealing Assistant sends the complete file to Accounts Officer to calculate the dues				√	√						
6	Accounts Officer calculates the dues and sends the file to Dealing Assistant				√	√						
7	Demand letter signed by Assistant Director is sent to Applicant by Dealing Assistant					√	√					
8	Applicant makes payment of dues in Branch located in DDA office and submits receipt to Assistant Director			√			√		√			
9	Dealing assistant sends the file to Accounts Officer to issue 'No dues certificate'				√	√						
10	If dues amounts more than INR 5000 the Accounts Officer sends file to Computer Department to get the confirmation of payment				√					√		

continued

Sr.no.	Description of process	Central Diarist	Record room incharge	Applicant	Accounts Officer	Dealing Assistant	Assistant Director	NSK Staff	Bank Employees	Computer Department	Witnesses	Sub Registrar
	Count of activities per actor	2	1	7	5	8	6	3	2	1	2	3
11	Accounts Officer issues 'No dues certificate' and sends file to Assistant Director				√		√					
12	Dealing Assistant sends conversion letter and three conversion deed papers signed by Assistant Director to Applicant					√	√					
13	Applicant fills up forms and pay stamp duty in the bank. Further, he submits filled up forms, payment receipt and request letter of execution of conveyance deed to Assistant Director			√			√		√			
14	Dealing Assistant sends call letter signed by Assistant Director for the execution of conveyance deed to Applicant					√	√					
15	Applicant visits NSK along with witnesses for the execution of conveyance deed			√				√			√	
16	Applicant visits Sub Registrar's Office along with Witnesses for the registration of conveyance deed			√							√	√
17	Applicant makes payment of registration fees through online mode and henceforth a date is given by Sub Registrar to the applicant to collect the registered conveyance deed			√								√
18	Applicant visits Sub Registrar's office to collect the registered conveyance deed			√								√

Appendix ll

Exhibit ll.1: Cross-interaction binary matrix

Actor									
	A1	1	1	0	0	0	0	0	0
	A2	0	0	1	0	0	0	0	0
	A3	0	0	0	1	1	0	1	0
	A4	0	0	0	0	0	1	1	1
	A5	0	1	1	1	1	1	0	0
	A6	0	0	0	1	0	1	0	1
	A7	0	0	0	0	0	0	0	1
	A8	0	0	0	0	0	0	0	1
		P1	P2	P3	P4	P5	P6	P7	P8
		Process							

Exhibit ll.2: Cross-interaction interpretive matrix

Actor								
A1	Record of online and physical applications is maintained with Central Diarist	Diarist handovers files to Dealing Assistant						
A2			Concerned record room incharge retrieve physical files from the record room					
A3				Applicant receives deficiency letter in case requirement of additional documents	Applicant submits pending documents to the Dealing Assistant		Applicant may have to pay additional amount in case of any deficit	
A4						Accounts Officer receives the file for payment verification	Accounts Officer verifies payment made by applicant and issues 'No Objection Certificate'	Accounts Officer sends the file with NOC for further approvals and processing
A5		Dealing Assistant receives files from Central Diarist	Record room incharge hands over retrieved file to the Dealing Assistant	Dealing Assistant verifies the file and sends the deficiency letter to the Applicant	Dealing Assistant receives pending documents from Applicant	Dealing Assistant again verifies the documents and sends the file to Accounts Officer		
A6				Assistant Director signs the deficiency letter		Assistant Director signs the file for further processing		Approval of Assistant Director
A7								Approval of Deputy Director
A8								Approval of Director
	P1	P2	P3	P4	P5	P6	P7	P8
	Process							

Exhibit ll.3: Interpretive logic—knowledge base—ranking of actors w.r.t. processes

Paired comparison	Interaction with process	Interpretive logic
A5 dominating A1	P2	Dealing Assistant of housing department approaches Central Diarist to receive files
A2 dominating A5	P3	Timely retrieval of file by record room incharge is important for further actions
A5 dominating A3	P4	Proper verification of documents and raising correct requirement first time is crucial
A6 dominating A5	P4	Assistant Director needs to sign the requirement before forwarding to the applicants
A3 dominating A5	P5	Timely submission of pending documents from applicant is important for further actions
A5 dominating A4	P6	Final verification by Dealing Assistant before forwarding the file to the Accounts Officer
A6 dominating A5	P6	Assistant Director needs to sign the file before forwarding to the Accounts Officer
A4 dominating A3	P7	Raising timely and correct requirement of pending dues by Accounts Officer is important
A6 dominating A4	P8	Assistant Director approves the file received from Accounts Officer
A7 dominating A6	P8	Deputy Director approves the file received from Assistant Director
A8 dominating A7	P8	Director approves the file received from Deputy Director

Exhibit ll.4: Pair-wise dominance of actors for different processes

(a) Dominating Interaction Matrix of Actors for Process P1

	A1	A2	A3	A4	A5	A6	A7	A8
A1	-	1	1	1	1	1	1	1
A2	-	-	0	0	0	0	0	0
A3	-	-	-	0	0	0	0	0
A4	-	-	-	-	0	0	0	0
A5	-	-	-	-	-	0	0	0
A6	-	-	-	-	-	-	0	0
A7	-	-	-	-	-	-	-	0
A8	-	-	-	-	-	-	-	-

(b) Dominating Interaction Matrix of Actors for Process P2

	A1	A2	A3	A4	A5	A6	A7	A8
A1	-	1	1	1	-	1	1	1
A2	-	-	0	0	-	0	0	0
A3	-	-	-	0	-	0	0	0
A4	-	-	-	-	-	0	0	0
A5	1	1	1	1	-	1	1	1
A6	-	-	-	-	-	-	0	0
A7	-	-	-	-	-	-	-	0
A8	-	-	-	-	-	-	-	-

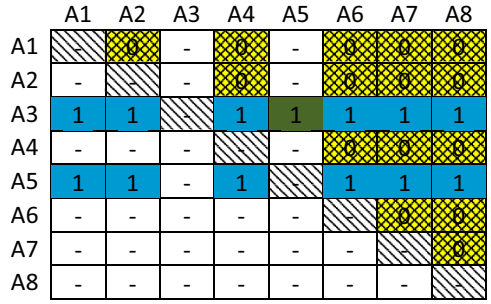
(c) Dominating Interaction Matrix of Actors for Process P3

	A1	A2	A3	A4	A5	A6	A7	A8
A1	-	-	0	0	-	0	0	0
A2	1	-	1	1	1	1	1	1
A3	-	-	-	0	-	0	0	0
A4	-	-	-	-	-	0	0	0
A5	1	-	1	1	-	1	1	1
A6	-	-	-	-	-	-	0	0
A7	-	-	-	-	-	-	-	0
A8	-	-	-	-	-	-	-	-

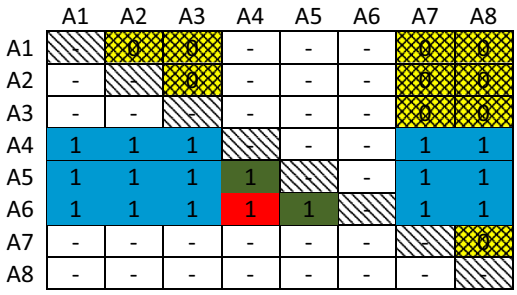
(d) Dominating Interaction Matrix of Actors for Process P4

	A1	A2	A3	A4	A5	A6	A7	A8
A1	-	0	-	0	-	-	0	0
A2	-	-	-	0	-	-	0	0
A3	1	1	-	1	-	-	1	1
A4	-	-	-	-	-	-	0	0
A5	1	1	1	1	-	-	1	1
A6	1	1	1	1	1	-	1	1
A7	-	-	-	-	-	-	-	0
A8	-	-	-	-	-	-	-	-

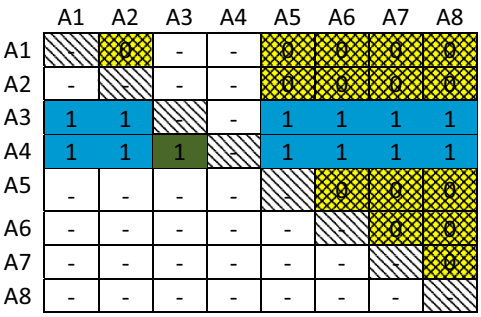
(e) Dominating Interaction Matrix of Actors for Process P5



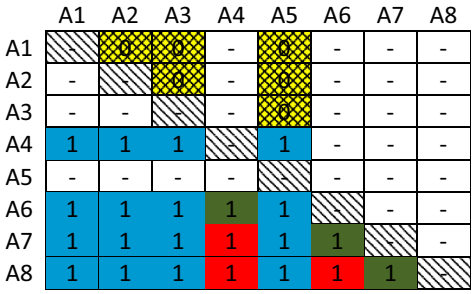
(f) Dominating Interaction Matrix of Actors for Process P6



(g) Dominating Interaction Matrix of Actors for Process P7



(h) Dominating Interaction Matrix of Actors for Process P8



1	Transitive Dominance
1	Interpretive Dominance
1	Implicit Dominance
0	Implicit non-Dominance

Exhibit ll.5: Dominance matrix—ranking of actors w.r.t. processes

	A1	A2	A3	A4	A5	A6	A7	A8	No. dominating	Net dominance	Rank dominating
A1	–	2	2	2	1	2	2	2	13	− 4	5
A2	1	–	1	1	1	1	1	1	7	− 10	7
A3	3	3	–	2	2	2	3	3	18	3	3
A4	3	3	3	–	2	1	2	2	16	1	4
A5	5	4	4	5	–	3	5	5	31	20	1
A6	3	3	3	3	3	–	2	2	19	8	2
A7	1	1	1	1	1	1	–	0	6	− 10	7
A8	1	1	1	1	1	1	1	–	7	− 8	6
No. being dominated	17	17	15	15	11	11	16	15	117		

Exhibit lll.1: Different types of dominance comparisons

Reference variables	Implicit dominance comparisons	Implicit Non-dominance comparisons	Transitive dominance comparisons	Interpretive dominance comparisons	Total comparisons	% Interpretive comparisons
P1	7	21	0	0	28	0
P2	12	15	0	1	28	9.09%
P3	12	15	0	1	28	9.09%
P4	15	10	1	2	28	18.18%
P5	12	15	0	1	28	9.09%
P6	15	10	1	2	28	18.18%
P7	12	15	0	1	28	9.09%
P8	16	6	3	3	28	27.27%
Total	101	107	5	11	224	
Percentage	45.09%	47.77%	2.23%	4.91%		

Appendix lll

Exhibit ll.6

	Description of situation—actor—process-related variables	
Macrovariables	Microvariables	Description
Improved situation	Distance covered to reach service centre	It refers to the number of service centres available for service delivery
	Influence of middleman	Presence of middleman near the service centres to approach citizens
	Queue to avail services	Time spent by citizens in the physical and electronic queues
	Seating arrangements	Availability of sufficient chairs for citizens
	Air conditioned environment	Availability and effectiveness of air conditioners at service centres
	Cleanliness of work area and washrooms	Level of cleanliness maintained at public service centres
	Drinking water	Availability of drinking water for citizens
	Provision of feedback	Existence of feedback forms at the end of the service delivery
Capability level of actors	Communication skills	Capability of employees to provide information with effective communication skills
	Service orientation	Willingness of employees to deliver services
	Fast execution of work	Speed of employees at service centre
	Knowledge level about service delivery	Clarity of processes and updated information with employees
	Respect for ethics and values	Employees providing services as per ethics and values of the public organization
	Availability at service counters	Presence of employees at their counter to deliver services
Flexible process workflow	Flexibility in date and time of application submission	Providing options to citizens' w.r.t. time and date of application submission
	Provision of uninterrupted services in case of technical fault	Arrangements of back up as per process in case of technical fault to continue with service delivery
	Option of both manual and online applications	Availability of both manual and online options for citizens' keeping their literacy level into consideration
	Flexibility to apply from any service centre within the city	Availability of choice with applicants to submit application at any service centre within the city
	Provision to avail services without visiting service centre	Provision for physically challenged applicants to submit application without visiting to service centre

Appendix lV

Exhibit lV.1: Study variables and their observed mean value

Microvariables	Question items	Mean value
Quality of information	Accurate	0.65
	Up-to-date	0.65
	Relevant	0.68
	Detailed	0.65
	Simple and understandable	0.74
System functioning	Easy to fill and submit application form	0.62
	Easy to make payment	0.62
	Easy to get the application processed without any technical error	0.67
User orientation	Easy access for people having little or no formal education	0.39
	Facilities for differently abled, senior citizens, infants etc.	0.45
Cost savings	Lesser visits required to avail service	0.19
	Excess money is paid to avail service	0.38
	Lesser efforts required to avail service	0.24
	Charges paid to Intermediaries to avail service	0.39
Efficiency	Duplicate tasks are not performed during process	0.53
	Delivery of service is fast	0.25
	Reach of service through service centres has increased	0.64
Openness	Display of information such as policies, expenses, agreements, tenders, etc.	0.34
	Display of organization charts, roles and responsibilities and contact lists of staff, etc.	0.43
	Display of working hours, lunch timings, office addresses, applicable fees, etc.	0.49
Responsiveness	Response to inquires	0.49
	Response to complaints	0.61
	Display citizen charter and RTI	0.44
Equity	Display of content in local language	0.61
	Same treatment to all applicants other than people with special needs	0.76
Trust	Security and privacy of personal information	0.70
	Initiatives to discourage role of middleman	0.37
Self-development	Awareness programmes for knowledge promotion	0.21
	Training programmes for non-Internet savvy people	0.20
	Access through common service centres/Kiosk available	0.56
Citizens' participation	Regular updates on policies and procedures	0.25
	The opportunity to participate in discussions, policy and decision-making	0.22
	Provision for suggestions	0.27
Concern for environment	Reduction in paper printing	0.60
	Energy saving (electricity, manpower, etc.)	0.57
Improved situation	Lesser distance of service centre from home	0.54
	More influence of middleman	0.51
	Long queue to avail services	0.54
	Proper sitting arrangements	0.63
	Air conditioned environment	0.48
	Clean work area and washrooms	0.51
	Availability of drinking water	0.56
	Provision of feedback mechanism	0.27
Capability level of actors	Communication skills	0.53
	Service orientation	0.53

continued

Microvariables	Question items	Mean value
	Fast execution of work	0.28
	Knowledge level about service delivery	0.66
	Respect for ethics and values	0.58
	Punctuality and presence at service counters	0.56
Flexible process workflow	Flexibility in date and time of application	0.44
	Provision of uninterrupted service in case of any technical fault	0.34
	Option of both manual and online applications	0.27
	Flexibility to apply from any service centre within Delhi/NCR	0.52
	Provision to avail service without visiting to service centre	0.03

References

Agus, A., Barker, S., & Kandampully, J. (2007). An exploratory study of service quality in the Malaysian public service sector. *International Journal of Quality & Reliability Management, 24*(2), 177–190.

Ahmad, S. A., & Latif, I. A. (2012). The effectiveness of public advisory services: What are the important factors? *International Journal of Business and Management, 7*(23), 19–30.

Al-Borie, M. H., & Damanhouri, S. M. A. (2013). Patients' satisfaction of service quality in Saudi hospitals: A SERVQUAL analysis. *International Journal of Health Care Quality Assurance, 26*(1), 20–30.

Axelsson, K., Melin, U., & Lindgren, I. (2010). Exploring the importance of citizen participation & involvement in e-government projects; Practice, incentives, and organization. *Transforming Government: People, Process and Policy, 4*(4), 299–321.

Belwal, R., & Zoubi, A. K. (2008). Public centric e-governance in Jordan, a field study of people's perception of e-governance awareness, corruption, and trust. *Journal of Information, Communication & Ethics in Society, 6*(4), 317–333.

Bharwani, S., & Jauhari, V. (2013). An exploratory study of competencies required to co-create memorable customer experiences in the hospitality industry. *International Journal of Contemporary Hospitality Management, 25*(6), 823–843.

Bhattacharya, D., Gulla, U., & Gupta, M. P. (2012). E-service quality model for Indian government portals: citizens' perspective. *Journal of Enterprise Information Management, 25*(3), 246–271.

Bidyarthi, H. M. J., & Srivastava, A. K. (2011). Citizens' perspectives of e- governance. *E-governance in practice*. http://www.csi-sigegov.org/egovernance_pdf/9_69-76.pdf. Accessed 18 Sep 2017.

Bishwas, S. K. (2015). Achieving organization vitality through innovation and flexibility: An empirical study. *Global Journal of Flexible Systems Management, 16*(2), 145–156.

Brown, G. L. (1990). Convenience in services marketing. *Journal of Services Marketing, 4*(1), 53–59.

Cavana, R. Y., Corbett, L. M., & Glenda Lo, Y. L. (2007). Developing zones of tolerance for managing passenger rail service quality. *International Journal of Quality & Reliability Management, 24*(1), 7–31.

Charan, P. (2012). Supply chain performance issues in an automobile company: A SAP-LAP analysis. *Measuring Business Excellence, 16*(1), 67–86.

Chatzoglou, P., Chatzoudes, D., Vraimaki, E., & Diamantidis, A. (2013). Service quality in the public sector: The case of the Citizen's Service Centers (CSCs) of Greece. *International Journal of Productivity and Performance Management, 62*(6), 583–605.

Chauhan, G., & Singh, T. P. (2013). Resource flexibility for lean manufacturing: SAP-LAP analysis of a case study. *International Journal of Lean Six Sigma, 4*(4), 370–388.

Connolly, R., Bannister, F., & Kearney, A. (2010). Government website service quality: A study of the Irish revenue online service. *European Journal of Information Systems, 19,* 649–667.

Dash, B. C., & Sangita, S. N. (2011). Governance reforms in power sector: initiatives and outcomes in Orissa. http://www.isec.ac.in/WP%20262%20-%20S%20N%20Sangita.pdf. Accessed 12 Dec 2017.

DeitY. (2008). Impact assessment of e-government projects. http://www.iimahd.ernet.in/egov/documents/impact-assessment-of-egovernance-projects.pdf. Accessed 13 Mar 2017.

European Commission. (2006). Measurement framework final version: E-government economics project. http://www.epractice.eu/files/media/media1299.pdf. Accessed 20 Jan 2018.

Garg, A., & Deshmukh, S. G. (2010). Engineering support issues for flexibility in maintenance. *Asia Pacific Journal of Marketing and Logistics, 22*(2), 247–270.

Gauld, R., Goldfinch, S., & Horsburgh, S. (2010). Do they want? Do they use it? The "Demand –Side" of e-Government in Australia and New Zealand. *Government Information Quarterly, 27*(2), 177–186.

Georgios, Z. (2004). Ethics versus corruption in globalization. *Journal of Management Development, 23*(7), 631–647.

Golubeva, A. A. (2007). Evaluation of regional government portal on the basis of public value concept. Case study from Russian Federation. In *ACM international conference proceeding series*. https://doi.org/10.1145/1328057.1328139.

Grimsley, M., & Meehan, A. (2007). E-government information systems: Evaluation-led design for public value and client trust. *European Journal of Information Systems, 16,* 134–148.

Gupta, P. J., & Suri, P. K. (2017). Measuring public value of e-governance projects in India: Citizens' Perspective. *Transforming Government: People, Process and Policy, 11*(2), 236–261.

Haldar, A., Rao, S. V. D. N., & Momaya, K. S. (2016). Can flexibility in corporate governance enhance international competitiveness? Evidence from knowledge-based industries in India. *Global Journal of Flexible Systems Management, 17*(4), 389–402.

Haleem, A., Sushil, Qadri, M. A., & Kumar, S. (2012). Analysis of critical success factors of world-class manufacturing practices: An application of interpretive structural modeling and interpretive ranking process. *Production Planning & Control, 23*(10–11), 722–734.

Harrison, T. M., Guerrero, S., Burke, G. B., Cook, M., Cresswell, A., Helbig, N., Hrdinová, J., & Pardo, T. (2011). Open government and e-government: Democratic challenges from a public value perspective. In *Proceedings of the 12th annual international conference on digital government research*. https://doi.org/10.3233/ip-2012-0269.

Heeks, R. (2002). Failure, success and improvisation of information systems projects in developing countries. http://unpan1.un.org/intradoc/groups/public/documents/NISPAcee/UNPAN015601.pdf. Accessed 12 Feb 2018.

Heeks, R. (2008). Benchmarking e-government: Improving the national and international measurement evaluation and comparison of e-government. *Evaluating Information Systems*. https://doi.org/10.4324/9780080570105.

Islam, R., Chowdhury, M. S., Sarker, M. S., & Ahmed, S. (2014). Measuring customer's satisfaction on bus transportation. *American Journal of Economics and Business Administration, 6*(1), 34–41.

John, L., & Ramesh, A. (2012). Humanitarian supply chain management in India: A SAP-LAP framework. *Journal of Advances in Management Research, 9*(2), 217–235.

Jorgensen, T. B., & Bozeman, B. (2007). Public values an inventory. *Administration & Society, 39*(3), 354–381.

Kakouris, A. P., & Meliou, E. (2010). New public management: promote the public sector modernization through service quality. Current experiences and future challenges. *Public Organization Review, 11*(4), 351–369.

Kalsi, N. S., & Kiran, R. (2013). E-governance success factors. *International Journal of Public Sector Management, 26*(4), 320–336.

Karkin, N., & Janssen, M. (2014). Evaluating websites from a public value perspective: A review of Turkish local government websites. *International Journal of Information Management, 34*(3), 351–363.

Karunasena, K., Deng, H., & Singh, M. (2011). Measuring the public value of e-government: A case study from Sri Lanka. *Transforming Government: People, Process and Policy, 5*(1), 81–99.

Karunasena, K. (2012). An investigation of the public value of e-government in Sri Lanka. https://researchbank.rmit.edu.au/eserv/rmit:160100/Thanthri_Waththage.pdf. Accessed 14 Nov 2017.

Kay, C., & Russette, J. (2000). Hospitality-management competencies: Identifying manager's essential skills. *Cornell Hotel and Restaurant Administration Quarterly, 41*(2), 52–63.

Kearns, I. (2004). Public value and e-government. http://www.ippr.org/uploadedFiles/projects/Kearns_PublicValueandeGovenrment_ippr.pdf. Accessed 18 Jan 2016.

Kelly, G., Mulgan, G., & Muers, S. (2002). Creating public value: An analytical framework for public service reform. http://www.allamreform.hu/letoltheto/kozfeladatok/kulfoldi/public_value2.pdf. Accessed 5 February 2015.

Kumar, P., Bera, S., Dutta, T., & Chakraborty, S. (2018). Auxiliary flexibility in healthcare delivery system: An integrative framework and implications. *Global Journal of Flexible Systems Management, 19*(2), 173–186.

Lindgren, I., & Jansson, G. (2013). Electronic services in the public sector: A conceptual framework. *Government Information Quarterly, 30*(2), 163–172.

Lovelock, C. H. (1983). Classifying services to gain strategic marketing insights. *Journal of Marketing, 47*(9), 9–20.

Mangla, S. K., Kumar, P., & Barua, M. K. (2015). Flexible decision modeling for evaluating the risks in green supply chain using fuzzy AHP and IRP methodologies. *Global Journal of Flexible Systems Management, 16*(1), 19–35.

Moore, M. H. (1995). *Creating public value: Strategic management in government*. London: Harvard University Press.

Navarro, J. G. C., Perez, A. G., & Cegarraa, J. L. M. (2014). Technology knowledge and governance: Empowering citizen engagement and participation. *Government Information Quarterly, 31*(4), 660–668.

Osman, I. H., Anouze, A. L., Irani, Z., Ayoubi, B. A., Lee, H., Balc, A., et al. (2014). COBRA framework to evaluate e-government services: A citizen-centric perspective. *Government Information Quarterly, 31*(2), 243–256.

Papadomichelaki, X., & Mentzas, G. (2012). e-GovQual: A multiple-item scale for assessing e-government service quality. *Government Information Quarterly, 29*(1), 98–109.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: Multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(1), 12–40.

Saaty, T. L. (1980). *The analytical hierarchy process: Planning, priority setting, resource allocation*. New York: Mcgraw-Hill.

Salminen, A., & Norrbacka, R. I. (2010). Trust, good governance and unethical actions in Finnish public administration. *International Journal of Public Sector Management, 23*(7), 647–668.

Satapathy, S. (2014). An analysis for service quality enhancement in electricity utility sector of India by SEM. *Benchmarking: An International Journal, 21*(6), 964–986.

Saxena, K. B. C. (2005). Towards excellence in e-governance. *International Journal of Public Sector Management, 18*(6), 498–513.

Sharma, K., Sushil, & Jain, K. P. (2010). Revisiting flexibility in organizations: Exploring its impact on performance. *Global Journal of Flexible Systems Management, 11*(3), 51–68.

Sharma, V., Dixit, A. R., & Qadri, M. A. (2016). Modeling lean implementation for manufacturing sector. *Journal of Modeling in Management, 11*(2), 405–426.

Singla, M., & Singh, S. (2018). Impact of institutional set-up on the responsiveness to change in a firm's governance structure: A comparative study of public and private sector enterprises in India. *Global Journal of Flexible Systems Management, 19*(2), 159–172.

Suri, P. K., & Sushil, (2006). E-governance through strategic alliances-a case of agricultural marketing information system in India. *IIMB Management Review, 18*(4), 389–401.

Suri, P. K., & Sushil, (2011). Multi-perspective analysis of e-governance performance: A study of select agriculture related projects in India. *International Journal of Electronic Governance, 4*(3), 259–272.

Suri, P. K. (2014). Flexibility of processes and e-governance performance. *Transforming Government: People, Process and Policy, 8*(2), 230–250.

Suri, P. K. (2016). Towards Linkage between Strategy Formulation and E-governance Performance. In Sushil, K. Bhal, & S. Singh (Eds.), *Managing flexibility: Developing a framework of flexibility maturity model* (pp. 43–59). New Delhi: Springer.

Suri, P. K., & Sushil, (2017). *Strategic planning and implementation of e-governance*. Singapore: Springer Science and Business Media.

Sushil, (2000). SAP-LAP models of inquiry. *Management Decision, 38*(5/6), 347–353.

Sushil, (2009a). SAP-LAP linkages: A generic interpretive framework for analyzing managerial contexts. *Global Journal of Flexible Systems Management, 10*(2), 11–20.

Sushil, (2009b). Interpretive ranking process. *Global Journal of Flexible Systems Management 10*(4), 1–10.

Sushil, (2017a). Theory building using SAP-LAP linkages: An application in the context of disaster management. *Annals of Operations Research*. https://doi.org/10.1007/s10479-017-2425-3.

Sushil, (2017b). Efficient interpretive ranking process incorporating implicit and transitive dominance relationships. *Annals of*

Operations Research. https://doi.org/10.1007/s10479-017-2608-y.
United Nations (2016). United Nations E-Government Survey. http://workspace.unpan.org/sites/Internet/Documents/UNPAN96407.pdf. Accessed 11 Jan 2018.
Vela, S. P., Martinez, I. G., Duhamel, F., Luna, D. E., & Luna-Reyes, L. F. (2015). Inter-organizational collaboration and value creation in digital government projects. https://doi.org/10.1145/2757401.2757403
Verdegem, P., & Verleye, G. (2009). User-centred e-government in practice: A comprehensive model for measuring user satisfaction. *Government Information Quarterly, 26*(3), 487–497.
Ware, N. R., Singh, S. P., & Banwet, D. K. (2014). Modeling flexible supplier selection framework. *Global Journal of Flexible Systems Management, 15*(3), 261–274.
Weerakkody, V., Janssen, M., & Dwivedi, Y. K. (2011). Transformational change and business process re-engineering (BPR): Lessons from the British and Dutch public sector. *Government Information Quarterly, 28*(1), 320–328.
Yusoff, W. Z. W., Ismail, M., & Newell, G. (2008). FM-SERVQUAL: A new approach of service quality measurement framework in local authorities. *Journal of Corporate Real Estate, 10*(2), 130–144.
Zhang, Y., Wang, Z., & Zhou, G. (2014). Determinants of employee electricity saving: The role of social benefits, personal benefits and organizational electricity saving climate. *Journal of Cleaner Production, 66,* 280–287.

Websites

www.meity.gov.in. Accessed 15 February 2017.
www.dda.org.in. Accessed 18 February 2017.

Key Questions for Reflection

1. Which are the Situation–Actor–Process-related variables which may be influencing public value of an e-governance project related to your organization?
2. How Interpretive Ranking Process (IRP) method can be used for improving Actor–Process interfaces of an e-governance project related to your organization?

Parul Jain Gupta is a Research Scholar at the Delhi School of Management, Delhi Technological University (DTU). Her area of research is e-governance. She has worked as a Customer Service Manager in a government project in India from Jan 2012 to Sep 2015. She is working as an Analyst since Oct 2015.

Pradeep Kumar Suri is serving as a Professor at the Delhi School of Management (DSM), Delhi Technological University (DTU). His areas of interest for teaching, research and consultancy include: E-Governance, IT Strategy and Management, Project Management, and Business Analytics. He has served as Head, Delhi School of Management, during 2013–2016. He is originally from the 1986 batch of the Indian Statistical Service. Prior to joining the DTU in 2012, he served as the Senior Technical Director at the National Informatics Centre (NIC), Government of India. He joined NIC during 1993, where he served in different capacities during the life cycle of several ICT-based projects in the agriculture and allied sectors. He has served in consulting roles for the UN bodies, viz. World Tourism Organization (UNWTO) and Food and Agricultural Organization (UNFAO). He has also co-authored a book titled '*Strategic Planning and Implementation of E-governance'* published by Springer.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/328576628

Application of Linear Quadratic Methods to Stabilize Cart Inverted Pendulum Systems

Conference Paper · October 2018

CITATIONS
0

READ
1

3 authors, including:

Sudarshan K Valluru
Delhi College of Engineering (Now Delhi Technological University) Govt.of NCT Delhi
27 PUBLICATIONS **36** CITATIONS

SEE PROFILE

Mayank Singh
Delhi Technological University
2 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

meta algorithms application to nonlinear dynamical systems View project

meta algorithms applications to nonlinear dynamical systems View project

All content following this page was uploaded by Sudarshan K Valluru on 29 October 2018.

The user has requested enhancement of the downloaded file.

Application of Linear Quadratic Methods to Stabilize Cart Inverted Pendulum Systems

Sudarshan K Valluru
Incubation Center for Control,Dynamical Systems Electrical Engineering,Delhi Technological University
Delhi-110042, India
sudarshan_valluru@dce.ac.in

Madhusudan Singh
Incubation Center for Control,Dynamical Systems Electrical Engineering,Delhi Technological University
Delhi-110042, India
madhusudan@dce.ac.in

Mayank Singh
Incubation Center for Control,Dynamical Systems Electrical Engineering,Delhi Technological University
Delhi-110042, India
mayank202244@gmail.com

***Abstract*—This paper presents the application of linear-quadratic methods of optimal control for a class of reference tracking problem. Specifically, linear quadratic regulator and linear-quadratic Gaussian methods are outlined to synthesize a desired closed loop system. Selection criteria of weights for both state weighting matrix, Q, and control weighting matrix, R, is addressed for the LQR design to satisfy the robustness measures for a closed loop system. The design is extended to employ linear quadratic estimator for the LQG design. An unstable linearized model of the cart inverted pendulum is considered to validate the proposed methodology.**

Keywords— LQR controller, LQG controller, Cart Inverted Pendulum System(CIPS).

I. INTRODUCTION

This paper utilizes the theory of deterministic linear regulator to design a desired closed loop system satisfying a stringent set of conditions. The cart inverted pendulum system often considered a benchmark system[1], [2] for controller design is considered to test the proposed design strategies. A simple tracking problem for the system is encountered where the pendulum angle with the vertical tracks the cost position to balance the pendulum in the upright position. The inherent instability of the system due to its under-actuated nature presents conceptual as well as practical difficulties for the control system design and real-time implementation. This paper presents linear quadratic regulator and linear quadratic Gaussian design methods to stabilize this unstable SIMO system.

The LQR controller obtains a closed loop system with optimal performance[3]. This is done by minimizing a cost function comprising the state vector and the control input which establishes a direct relationship with the desired performance indices. This process results in the optimal state feedback gain, K. The value of this gain, K, is closely related to the values/weights selected for the state weighting matrix, Q, and control weighting matrix, R. Weights of Q and R, are selected so as to design the closed loop system to ensure the robustness properties for both LQR and LQG to small parametric changes, i.e., reasonable changes in the parameters of open-loop transfer functions which can be accommodated in the closed loop design with ensured stability.

Towards this goal, this paper is organized under the following section headings. Section II presents the mathematical description of the system for both open loop and closed loop for real-time implementation. Controllability and observability measures are also discussed in this section. In Section III, design methodologies are elucidated with the determination of parameters for experimental implementation. Section IV presents the obtained results for both design methods, followed by conclusion in Section V.

II. MATHEMATICAL DESCRIPTION OF CART INVERTED PENDULUM SYSTEM

A. Open Loop Analysis

The cart inverted pendulum system is sown in Fig.1, which is considered for closed-loop design. The system parameters are specified in Table- I.

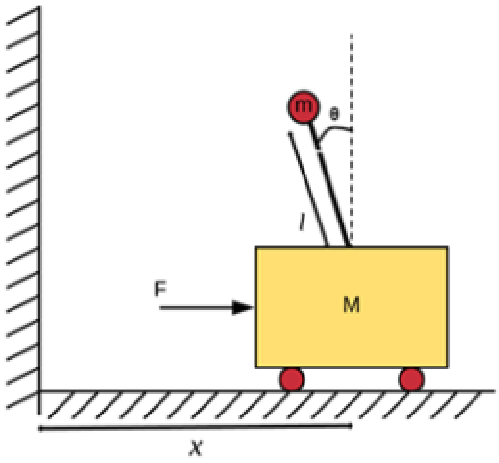


Fig.1 Cart Inverted Pendulum System

TABLE I. C I P S PARAMETERS[4]

Parameters	Value
F - Force applied to cart ±24N	±24N
x - Cart position from the reference ±0.5 m	±0.5m
θ - Pendulum angle w.r.t vertical axis 0.1 rad	0.1 rad
g - Gravity 9.81m/s2	9.81m/s^2
l - Length of the pole 0.38m	0.38m
M - Cart mass 2.4 Kg	2.4kg
m - Pole mass 0.23 Kg	0.23kg
I - Moment of inertia of pole 0.099Kg/m2	0.099kg/m2
b - Cart friction coefficient 0.05Ns/m	0.05Ns/m
d - Pendulum damping coefficient 0.05 s /rad	0.05s/rad

The standard non-linear differential equations can be designed using Newtonian mechanics or Lagrange's

978-1-5386-6625-8/18/$31.00 ©2018 IEEE

equations. For state vector as $x = [x(t) \quad \dot{x}(t) \quad \theta(t) \quad \dot{\theta}(t)]^T$, and output vector as $y = [x \quad \theta]^T$, which is given as a non-linear differential equation in (1).

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} x_2 \\ \frac{(mlcosx_2)(dx_4 - mglsinx_2) + (I + ml^2)(mglsinx_2)(x_4)^2 + bx_2}{\Delta(\theta)} \\ x_4 \\ \frac{(M+m)(dx_4 - mglsinx_3) + (mlcosx_3)(mglsinx_3(x_4)^2) + bx_2}{\Delta(\theta)} \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{I + ml^2}{\Delta(\theta)} \\ 0 \\ -\frac{mlcosx_3}{\Delta(\theta)} \end{bmatrix} F \tag{1}$$

where $\Delta(\theta) = (ml)^2 - (I + ml^2)(M + m)cosx_3$. Also, for the experimental setup force F(s) is related to control voltage U(s) which is given in equation (2).

$$\frac{F(S)}{U(S)} = 18 \tag{2}$$

The system can be linearized at θ=0, cos θ≈1, sin θ≈0, and $(\dot{\theta})^2 = 0$, the system can be rewritten as equations (3) and (4)

$$\dot{x} = Ax + Bu \tag{3}$$

$$y = Cx + Du \tag{4}$$

where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, $D \in \mathbb{R}^{p \times m}$ are dynamic matrix, input matrix, output matrix, feed through the matrix, respectively, which are obtained as

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -0.019 & 0.22 & -0.001 \\ 0 & 0 & 0 & 1 \\ 0 & -0.013 & 6.6 & -0.04 \end{bmatrix} \tag{5}$$

$$B = \begin{bmatrix} 0 \\ 6.99 \\ 0 \\ 4.625 \end{bmatrix} \tag{6}$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{7}$$

$$D = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{8}$$

Cart Position transfer function $T_x(s)$, and pendulum angle transfer function $T_\theta(s)$ are obtained[5] from the state-space representation is given in (9) and (10)

$$\frac{X(s)}{U(s)} = T_x(s) = \frac{6.99s^2 + 0.246s - 45.38}{s^4 + 0.05811s^3 - 6.63s^2 - 0.1261s} \tag{9}$$

$$\frac{\theta(s)}{U(s)} = T_\theta(s) = \frac{4.626s^2}{s^4 + 0.05311s^3 - 6.63s^2 - 0.1261s} \tag{10}$$

B. Controlability and Observability

A constant coefficient linear system is completely controllable if the matrix M_c is

$$M_c = [B \quad AB \quad A^2B \ldots \ldots A^{n-1}B] \tag{11}$$

The M_c has rank *n*, where $x(t) \in \mathbb{R}^n$. A constant coefficient is linearly observable if M_o has rank *n*.

$$M_o = [C' \quad A'C' \quad (A^2)'C' \ldots \ldots (A^{n-1})'C'] \tag{12}$$

For the cart inverted pendulum system, the controllability and observability matrices are given in equations (13) and (14).

$$M_c = \begin{bmatrix} 0 & 6.99 & -0.14 & 1.02 \\ 6.99 & -0.14 & 1.02 & -0.11 \\ 0 & 4.62 & -0.268 & 30.68 \\ 4.62 & -0.268 & 30.68 & -2.98 \end{bmatrix} \tag{13}$$

$$M_o = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -0.019 & -0.013 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.22 & 6.6 & -0.013 & -0.26 \\ 0 & 0 & 0 & 1 & 0 & -0.03 & 0.22 & 6.6 \end{bmatrix}^T \tag{14}$$

Here the rank of $(M_c) = (M_o) = 4 = n$, which proves the system is fully observable and controllable.

C. Closed Loop Analysis of CIPS

The Fig. 2 describes the closed-loop architecture comprising the two PID controller where gain values are computed from proposed methodologies.

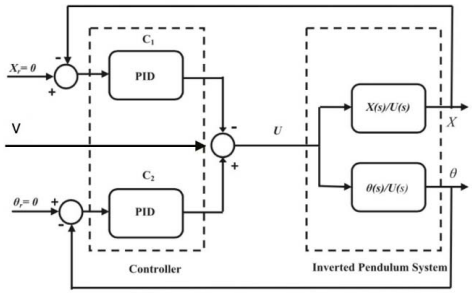


Fig. 2 Closed loop architecture for CIPS with Controllers

The closed-loop system can be summarised as follows in (15), (16), and (17).

$$C_1(s) = K_{p1} + \frac{K_{i1}}{s} + K_{d1}s \tag{15}$$

$$C_2(s) = K_{p2} + \frac{K_{i2}}{s} + K_{d2}s \tag{16}$$

$$\begin{bmatrix} T_{x_CL} \\ T_{\theta_CL} \end{bmatrix} = \frac{1}{1 + T_xC_1 + T_\theta C_2} \begin{bmatrix} T_x \\ T_\theta \end{bmatrix} \tag{17}$$

Where $T_{x\text{-}cl}$ and $T_{\theta\text{-}cl}$ are the closed loop cart position and pendulum angle transfer function respectively, and C_1 and C_2 are the controller parameters. The net characteristic equation from (17) is obtained as equation (18)

$$\begin{aligned} &s^5 + (6.99K_{d1} + 4.626K_{d2} + 0.053)s^4 + \\ &(6.99K_{p1} + 0.246K_{d1} + 4.626K_{p2} - 6.63)s^3 \\ &+(0.246K_{p_1} + 6.99K_{i1} - 45.38K_{d1} + 4.626K_{i2} - 0.1261)s^2 \\ &+(-45.38K_{p1} + 0.246K_{i1})s + (-45.38)K_{i1} = 0 \end{aligned} \tag{18}$$

The characteristic equation (18) be similar to a fifth order characteristic equation (19). Which is compactly rewritten as equation (20).

$$s^5 + p_1 s^4 + p_2 s^3 + p_3 s^2 + p_4 s + p_5 = 0 \quad (19)$$

$$\begin{bmatrix} 0 & 0 & 6.99 & 0 & 0 & 4.626 \\ 6.99 & 0 & 0.246 & 4.626 & 0 & 0 \\ 0.246 & 6.99 & -45.38 & 0 & 4.626 & 0 \\ -45.38 & 0.246 & 0 & 0 & 0 & 0 \\ 0 & -45.38 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} K_{p1} \\ K_{i1} \\ K_{d1} \\ K_{p2} \\ K_{i2} \\ K_{d2} \end{bmatrix} = \begin{bmatrix} p_1 - 0.058 \\ p_2 + 6.63 \\ p_3 + 0.1261 \\ p_4 \\ p_5 \end{bmatrix} \quad (20)$$

The equation (20) is compared with equation (21)

$$SR = p \quad (21)$$

The least norm solution for equation (21) can be calculated as equation (22)

$$R = (S^T S)^{-1}(S^T P) \quad (22)$$

The desired characteristic equation (19) can be obtained from the proposed control strategies.

III. Design of Controllers

A. Design of LQRController

The linear quadratic regulation (LQR) method assumes full state feedback for linear controller design[6]. For a continuous time, the linear dynamical system as represented by equations (3)and (4), the optimization around certain constraints are posed on the minimization of a cost index defined in equation (23).

$$J = \int_0^\infty [x^T Q x + u^T R u]\, dt \quad (23)$$

Where Q is the stable weighting matrix, and R is the control weighting matrix. Both Q and R, are symmetric, where the former is positive semi-definite, and the latter is positive definite. The steady state solution of the regulator problem in the linear time-invariant, state feedback control law is given in equation (24).

$$u(t) = -Kx(t) \quad (24)$$

where $K \in \mathbb{R}^{m \times n}$ is the control gain matrix, which is calculated by equation (25)

$$K = R^{-1} B^T P \quad (25)$$

Moreover, P is symmetric positive semi-definite, which is calculated the solution by using algebraic Riccati equation (26).

$$PA + A^T P + Q - PBR^{-1}B^T P = 0 \quad (26)$$

The selection of weights for Q and R matrices and then their evolution of cost function represents the subjective part of the problem. These weights are selected to ensure a good degree of robustness, chacharacterized LQ optimal design by a gain margin, GM > 2, and a phase margin, PM > 60 degrees. The following state control weighting matrices in (27) and (28) were selected.

$$Q = \begin{bmatrix} 50 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (27)$$

$$R = [1] \quad (28)$$

Based on these matrices and using the solution of algebraic Riccati equation, the control gain matrix, K is calculated by equation (29)

$$K = [-3.16 \quad -3.73 \quad 22.91 \quad 8.48] \quad (29)$$

To obtain a fifth order characteristics equation, a fifth pole was added at 4-10 times the distance of the farthest pole. The closed loop characteristics equation is obtained from LQR design in equation (30).

$$s^5 + 40.4s^4 + 576s^3 + 3733s^2 + 7755s + 6177 = 0 \quad (30)$$

B. Design of LQG Controller

When the LQR controller is combined with linear quadratic estimator (Kalman filter), it results into the linear quadratic Gaussian controller. To overcome the assumption of full state feedback, a full state observer is designed to estimate all states from available state measurements. A stochastic model given in equations (31) and (32) are considered to account for system uncertainties, i. e., process noise, and sensor noise.

$$\dot{x} = Ax + Bu + Fd \quad (31)$$

$$y = Cx + n \quad (32)$$

Where d is the process disturbance vector and x is the sensor noise vector. Both noises are assumed to be zero mean Gaussian white noises with $E[d(t)] = E[n(t)] = 0$. The disturbance and noise covariances are given by V_d and V_n respectively, and given in equations (33) and (34).

$$E[d(t)d(\tau)^T] = V_d \delta(t - \tau) \quad (33)$$

$$E[n(t)n(\tau)^T] = V_n \delta(t - \tau) \quad (34)$$

The state equation for the Kalman filter is given by equation (35).

$$\dot{x}_o = Ax_o + Bu + L(y - Cx_o) \quad (35)$$

Where x_0 is the estimated state vector and L is the optimal filter gain. The objective of Kalman filter is to minimize the covariance of the estimation error, given by (36).

$$R_e(t) = E[e(t)e(t)^T] \quad (36)$$

Where e(t) is the estimation error given by x(t)-$x_0(t)$ and $R_e(t)$ is the covariance of the estimation error which has to

be minimized. The error dynamics equation is obtained in (37)

$$\dot{e_o} = (A - LC)e_o + Fd - Ln \quad (37)$$

The optimal Kalman gain filter is given by (38)

$$L = YC^T V_n \quad (38)$$

Where Y is the solution to the Riccati equation (39)

$$YA^T + AY - YC^T V_n^{-1} CY + V_d = 0 \quad (39)$$

The state space representation of the system using the LQG compensation is given by equations (40) and (41)

$$\dot{x_o} = (A - BK - LC)x_o + Ly \quad (40)$$

$$u = -Kx_o \quad (41)$$

Where K is the LQR gain. To find the Kalman filter gain matrices (42) and (43) are used for V_d and V_n respectively.

$$V_d = \begin{bmatrix} 25 & 0 & 0 & 0 \\ 0 & 25 & 0 & 0 \\ 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 25 \end{bmatrix} \quad (42)$$

$$V_n = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix} \quad (43)$$

The Kalman filter gain matrix is obtained in equation (44)

$$L = \begin{bmatrix} 4.8328 & 0.0288 \\ 3.865 & 0.2647 \\ 0.0288 & 6.6196 \\ 0.0651 & 14.0976 \end{bmatrix} \quad (44)$$

The closed loop poles were obtained for the matrix $[A - BK - LC]$ and given by G in equation (45)

$$G = \begin{bmatrix} -10.36 + i9.36 \\ -10.36 - i9.36 \\ -3.46 \\ -0.41 \end{bmatrix} \quad (45)$$

From the LQG design, the closed loop characteristic equation is obtained in equation (46)

$$s^5 + 33s^4 + 344s^3 + 1700s^2 + 3600s + 3000 = 0 \quad (46)$$

IV. Experimental Results and Comparative Analysis

A. Experimental time response for CIPS with LQR

The experimental responses position, angle and control voltage of CIPS with LQR controller are shown in Fig.3, Fig.4 and Fig.5 respectively. The cart and pendulum responses are noisy even though stabilized.

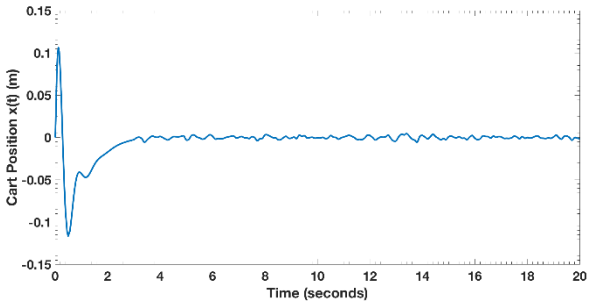


Fig. 3 Experimental time response for cart position x(t)

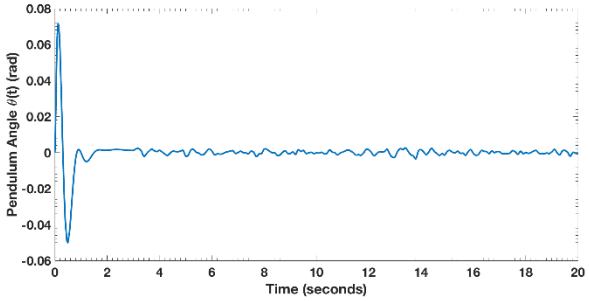


Fig. 4 Experimental time response for pendulum angle θ(t)

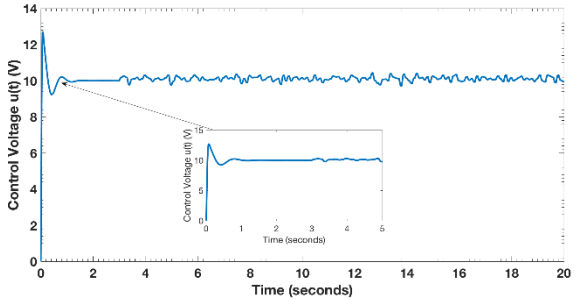


Fig. 5 Experimental time response for control voltage u(t)

B. Experimental time response for CIPS with LQG

The experimental responses position, angle and control voltage of CIPS with LQR controller are shown in Fig.6, Fig.7 and Fig.8 respectively. It is observed that from implementing LQG controller with cart inverted pendulum system responses are lesser noisy as compared to LQR controller.

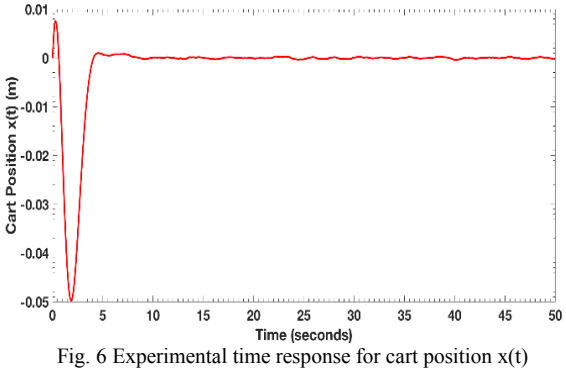


Fig. 6 Experimental time response for cart position x(t)

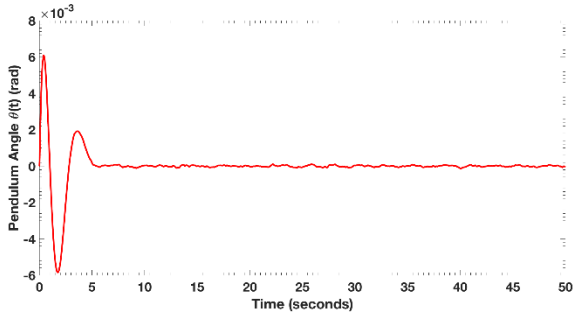


Fig. 7 Experimental time response for pendulum angle θ(t)

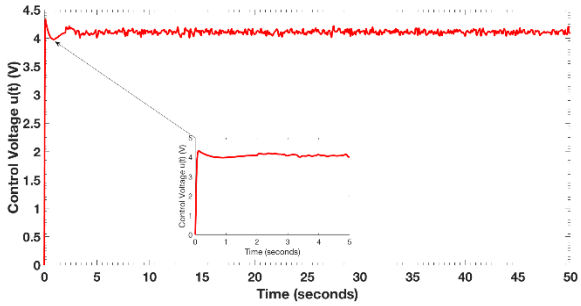


Fig. 8 Experimental time response for control voltage u(t)

C. *Comparison of PID Gain Values*

The characteristic equation obtained from LQR and LQG design in (30) and (46) are utilized to determine the PID gain values using equations (21) and (22). Table-II enlists the obtained PID gain values.

TABLE II. PID GAIN VALUES FOR LQR AND LQG CONTROLLER

Controller Parameters	LQR	LQG
K_{p1}	-172	-80
K_{i1}	-136	-66
K_{d1}	-416	-728
K_{p2}	407	158
K_{i2}	606	611
K_{d2}	637	503

D. *Comparison by performance indices*

Table-III presents the comparison between proposed methodologies based on standard performance indices.

TABLE III. COMPARISON OF PERFORMANCE INDICES

Performance Indices	LQR		LQG	
	x(t)	θ(t)	x(t)	θ(t)
Settling Time (sec)	3.12	2.81	5.23	5.18
Gain Margin (dB)	50.4	61.2	78	82
Phase Margin (deg)	60.1	62.3	76.4	79.8

From the obtained time responses, the closed loop system designed by the LQG method has a higher settling time w.r.t. the closed loop system designed by the LQR method. However, the LQG method leads to a system with higher tolerance to disturbances as greater PM and GM values are obtained for the respective closed loop system. Also, the experimental implementation by LQG method has lower PID gain values than the LQR counterpart.

V. CONCLUSION

In this paper, LQR and LQG methods are used to design a robust closed loop system validated a cart inverted pendulum system. Proposed methodologies were used to predetermine the PID gain values for experimental implementation. Experimental time responses were presented which indicates a more robust closed loop system designed via the LQG method.

ACKNOWLEDGMENT

The authors would like to thank Govt. of India and Govt. NCT Delhi for providing funding through Delhi Technological University under TEQIP-II to procure Digital Pendulum System Experimental Setup.

REFERENCES

[1] E.J.Davison, "Benchmark Problems for the Control System Design: Report of the IFAC Theory Committe," Schlossplatz 12,A-2361 Laxenburg, Austria, 1990.

[2] Jonathan P. How, "Benchmarks," *IEEE Control Systems Magazine*, pp. 6–7, 2015.

[3] A. Ghosh, T. R. Krishnan, and B. Subudhi, "Robust Proportional-Integral-Derivative Compensation of an Inverted Cart-Pendulum System: An Experimental Study," *IET Control Theory Appl. Theory Appl.*, vol. 6, no. 8, pp. 1145–1152, 2012.

[4] FeedbackInstruments, "Digital Pendulum Control Experiments Manual," 2016.

[5] S. K. Valluru and M. Singh, "Stabilization of Nonlinear Inverted Pendulum System Using MOGA and APSO Tuned Nonlinear PID Controller," *Cogent Eng.*, vol. 4, no. 1, pp. 1–15, 2017.

[6] Sudarshan K. Valluru Madhusudan Singh S.Singh, "Prototype Design and Analysis of Controllers for One Dimensional Ball and Beam System," in *1st IEEE International Conference on Power Electronics. Intelligent Control and Energy Systems*, 2016, pp. 1–6.

Design of Multi-Loop L-PID and NL-PID Controllers: An Experimental Validation

Sudarshan K. Valluru, Madhusudan Singh, Arnav Goel, Manpreet Kaur, Daksh Dobhal, Kumar Kartikeya, Aditya Verma and Anshul Gupta
Incubation Center for Control, Dynamical Systems and Computation
Department of Electrical Engineering, Delhi Technological University
Delhi-110042, India

Abstract—**The three-term PID controllers are significantly used in most of the benchmarked systems as controllers. However, the benchmarked Single Input Multiple Output (SIMO) systems may require more than one PID controllers in the control loops, the way to design PID controllers is not an easy task. Although, the linear PID (L-PID) controllers are not yielded a satisfactory control results for a benchmarked system due to the presence of inherent nonlinearities. In order to improve performance, the linear PID controllers are modified by nonlinear characteristics. In this paper, the linear and nonlinear PID (NL-PID) controllers are implemented for the stabilization and control of Gantry Crane System as a test problem. The experimental results prove that the multi-loop NL-PID controllers exhibit more robustness to outer large with fast external disturbances. The L-PID and NL-PID controllers are compared to find out the best controller, which gives optimum responses and can also handle external disturbances.**

Keywords— Multiloop L-PID; Multiloop NL-PID; Gantry Crane Control

I. INTRODUCTION

The Linear PID is one of the conventional controllers used in Control System industry and is widely accepted for providing generic and efficient solutions to real-world control problems[1], [2]. Few reasons for its popularity include simplicity in simulation, being inexpensive and efficacious in performance characteristics. Zeigler and Nichols pioneered significant work on the tuning of PID controller. Ziegler-Nichols type rules to much extent[3] are now widely used in PIDs. The design of an L-PID controller for a particular problem is a tedious task. This is because when the system under consideration has nonlinearity or the system parameters vary over a wide range, the performance of linear PID deteriorates considerably.

When used practically, L-PID controllers[4] quite often exhibit unexpected behavior including undesired oscillations and haphazard cart movement. This happens because the reference inputs are usually not continuous or smooth as they are subjected to the disturbance or noise signal, while the output of the system is required to be continuous and smooth. Since the smooth continuous output is taken as the direct objective, the inertia influence of the system is neglected which causes abrupt oscillations when used practically. Also, the reference signals are usually non-differentiable signals which makes it problematic to achieve the differential signal of the error. Therefore, If not properly implemented it can cause the system to collapse causing damage to the apparatus. Hence, care needs to be taken while designing the controller.

In recent years, there has been an upsurge in finding alternatives for Linear PID. The need for substitutes has arisen because the transient performance and parameters of Linear PID are usually degraded. Non-Linear PID has emerged to be an effective alternative to L-PID. In this context, the NL-PID designed uses a nonlinear sigmoid function. The overall response is more fast and robust than that of L-PID.

The paper is divided into five sections. Section I contains the introduction; Section II illustrates the design of Multi PID controllers. The benchmarked gantry crane system is explained in Section III. Experimentation validation is presented in Section IV, and the paper is concluded with the discussion presented in Section V.

II. DESIGN OF MULTI-LOOP PID CONTROLLERS

For effective minimization of swing angle and cart tracking of the input signal, two PID controllers are required, one each for the angle and position error. First, the PID controller for angle error was tuned, then keeping the earlier obtained values fixed; the position controller was tuned. Experimentation was carried out using the following two approaches:

A. Multi-Loop L-PID Controller

The L-PID takes in error, e(t) which is the difference between the desired output and real-time output, and processes it according to the equation (1) and its schematic structure is shown in Fig.1.

$$u(t) = K_p e(t) + K_i \int_{t_1}^{t_2} e(t)dt + K_d \frac{de(t)}{dt} \qquad (1)$$

Where K_p , K_i , and K_d are the proportional, integral and derivative gains respectively. This gives the control signal, $u(t)$. Each term in the above equation has its own significance. The proportional term leads to an increase in the speed of the system response, but it can be increased only up to a certain limit varying from system to system. A further increase will lead to oscillation of the variable being processed. The major steady error reducing term in the equation is the Integral term. The Derivative term results in a decrease of the output variable if the error is increasing and minimization of `oscillations.

978-1-5386-6625-8/18/$31.00 ©2018 IEEE

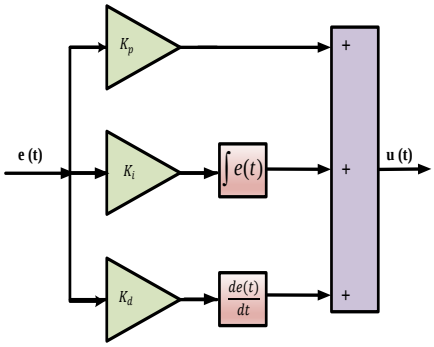


Fig. 2 Schematic structure of L-PID Controller[5]

B. *Multi-Loop NL-PID Controller*

In order to alleviate the problems associated with the L-PID, a novel method involving nonlinearities is proposed. The P, I & D error terms are processed through a non-linear function[6] to get the control signal, $u(t)$ is given in equation (2) and its schematic diagram is shown in Fig.2.

$$u(t) = e_p(t) + e_i(t) + e_d(t) \quad (2)$$

where $e_p = f(K_p, e)$, $e_i = f(K_i, e)$, $e_d = f(K_d, e)$

Here, f is a non-linear function, which is defined as equation (3)

$$f\left(K_\alpha, e\right) = \frac{1}{1+\exp(-K_\alpha e)} \quad (3)$$

This function maps high values to a value of 1 and a low value to 0. This behavior leads control error to be constrained in a bounded region and so doesn't permits voltage input to rise to high value. Further, in both the controllers a velocity filter is added in series with a derivative block in order to decrease oscillations and smooth out movement of the pendulum. The filter follows the equation (4) in the frequency domain.

$$d(s) = \frac{\exp(4)}{s^2 + 70.7s + \exp(4)} \quad (4)$$

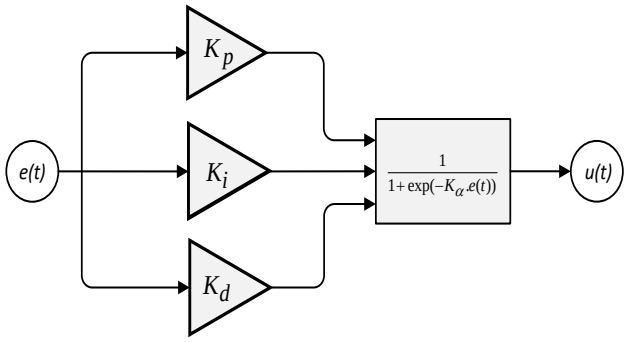


Fig. 2 Schematic structure of NL-PID Controller

III. Benchmarked Gantry Crane System

Despite its simple structure, the crane system[7], just like the inverted pendulum model has many complexities owing to its non-linear nature. Due to these problems, it is regarded as one of the ten benchmarks of control systems and its Phenomenological model is shown in Fig.3. Assuming the payload to be a lumped-mass distribution rather than a distributed-mass system, we can model the gantry crane system as a simple pendulum. The Fig.2. Illustrates the approximated structure of a gantry crane system with the assumptions of (1) payload has a point mass distribution (2) cable is massless (3) cable is not flexible, and its length unvarying and(4) damping coefficient of the pendulum is neglected.

Let M, m, l, b and I be the mass of the trolley, mass of the pendulum bob, length of the cable, the coefficient of friction of cart and the MOI of the bob respectively. Also, x and θ be the position of trolley along the belt and angular position of the pendulum.

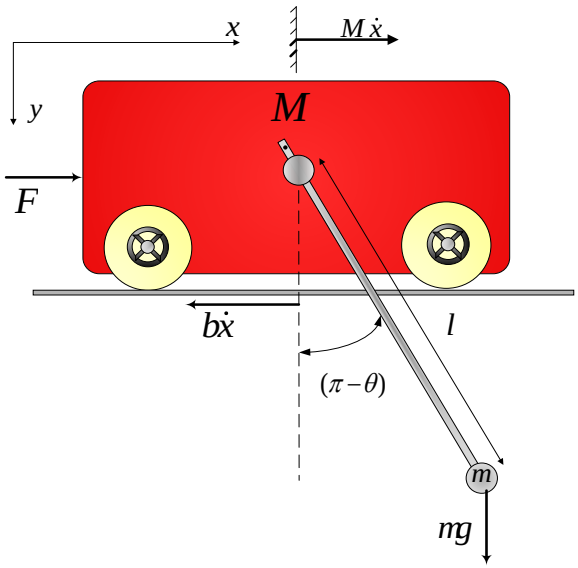


Fig.3 Phenomenological model of Gantry Crane System

The crane system is an under-actuated system as the degrees of freedom for this system is more than the actuators, i.e., we have trolley position, and vertical pendulum angle controlled by changing the voltage (V) applied to the DC motor. The force F (Tension force) on the trolley mass M is a linear function of the voltage applied to the servo DC motor coupled to the socket wheel.

The crane system is an under-actuated system as the degrees of freedom for this system is more than the actuators, i.e., we have trolley position, and vertical pendulum angle controlled by changing the voltage (V) applied to the DC motor. The force F (Tension force) on the trolley mass M is a linear function of the voltage applied to the servo DC motor coupled to the socket wheel.

The two equations obtained for the system are:

$$F = (M+m).\ddot{x} + b\dot{x} - ml\sin\theta.\dot{\theta}^2 + ml\cos\theta.\ddot{\theta} \quad (5)$$

$$0 = (\mathrm{I} + \mathrm{ml}^2).\ddot{\theta} - \mathrm{mgl}\sin\theta + \mathrm{ml}\cos\theta.\ddot{x} \quad (6)$$

The system parameters as specified in equation (5) and (6) are given in Table –I.

TABLE I. GANTRY CRANE SYSTEM PARAMETERS

S No.	System Parameters	Symbol	Value
1.	Mass of the Cart	M	2.4 kg
2.	Mass of the Pole	m	0.23 kg
3.	Friction Coefficient of the Cart	b	0.05 Ns/m
4.	Length of the Pole	l	0.32 m
5.	Moment of Inertia of the pole	I	0.099 kgm^2
6.	Gravity constant	g	9.81 ms^{-2}
7.	Range of operating voltage of D.C. Servo Motor	V	± 24 Volts

For $\theta \to 0$, $\sin\theta \approx -\theta$, $\cos\theta \approx -1$, $\dot{\theta}^2 \approx 0$

For the lucid understanding of carrying out the stability analysis of the dynamics of the Gantry Crane System, the system equations (5) and (6) linearized and given as equation (7) and (8).

$$F = (M+m).\ddot{x} + b\dot{x} - ml.\ddot{\theta} \quad (7)$$

$$0 = (I+ml^2).\ddot{\theta} + mgl\theta - ml.\ddot{x} \quad (8)$$

Alternatively, the above two equations can also be derived using Lagrange equation.

$$\frac{d}{dt}\left(\frac{d\mathcal{L}}{d\dot{q}}\right) - \frac{d\mathcal{L}}{dq} = 0 \quad (9)$$

Where q is the degrees of freedom of the system (x, θ) and $\mathcal{L}$ is a difference between total and potential energy.

IV. Experimental Validation and Comparative Analysis

The controller maintained a controlled voltage within the rated limits of $\pm$ 24V for DC servo motor. Real-time experimentation was done on the Feedback Instruments made Digital Control Pendulum model. The experimental diagram of the gantry crane system is shown in Fig.4

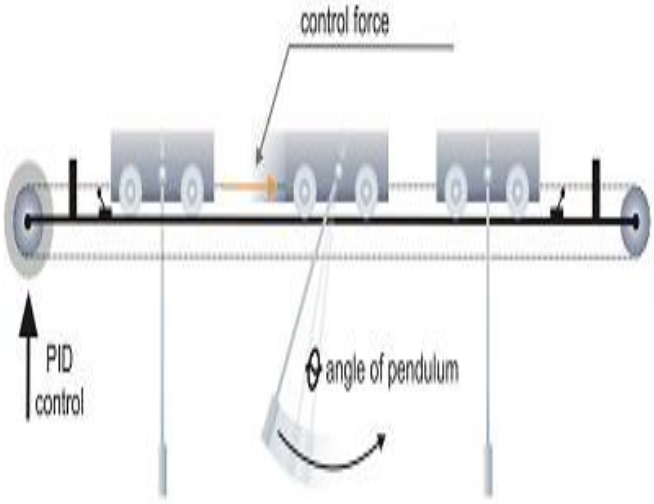


Fig. 4 Block Diagram for Gantry Crane System[8]

The gain values for angle and cart control were tuned manually by the trial-and-error method. The runtime for the experiments was taken as 100secs. The best gain values obtained are given in Table-II.

TABLE II. Controllers Gain Parameters

S. No.	Gain	L-PID		NL-PID	
		Cart	Pendulum	Cart	Pendulum
1.	K_p	4.30	8.00	5.00	6.00
2.	K_i	0.50	0.15	0.15	0.10
3.	K_d	6.00	0.30	5.00	0.30

A. Multi-Loop L- PID Controlled Gantry Crane System

The responses of the L-PID controller for the stabilization of the cart and pendulum of the gantry crane system is shown in Fig.5. The gantry crane system is operated for 100 sec and for every 20 sec interval some external disturbances are applied manually to the pendulum.

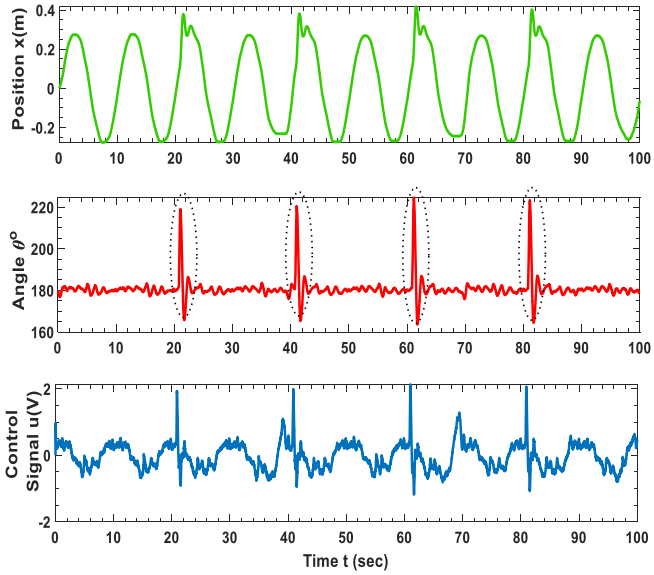


Fig.5 Responses of multi-loop L-PID Controller

B. Multi-Loop NL- PID Controlled Gantry Crane System

The responses of NL-PID controller for the stabilization of the cart and pendulum of the gantry crane system is shown in Fig.5. The gantry crane system is operated for 100 sec and for every 20 sec interval some external disturbances are applied manually to the pendulum.

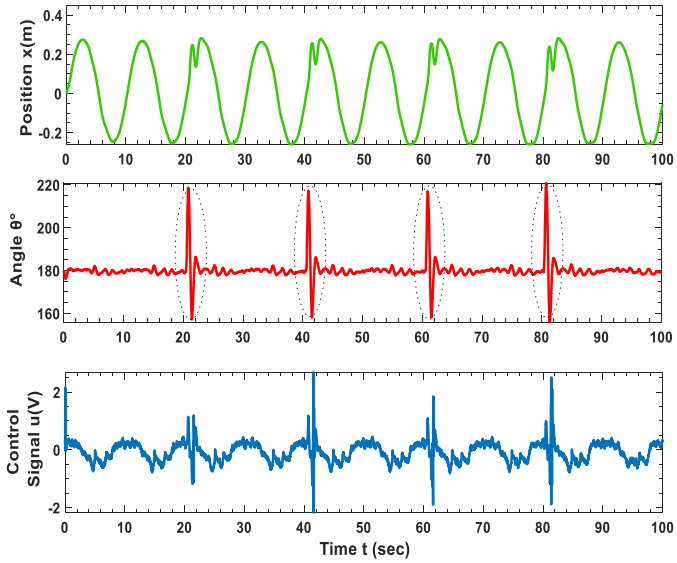


Fig.6 Responses of multi-loop NL-PID Controller

C. Comparative Analysis of L-PID and NL-PID controllers

The L-PID and NL-PID responses to stabilization of the pendulum angle against the external noise/ disturbance is given in Table-III and Table-IV respectively. From the tables, it is observed that the NL-PID stabilizes (θ about π) in the shorter period than the L-PID controller.

TABLE III. PENDULUM ANGLE STABILIZATION RESPONSE FOR THE L-PID CONTROLLER

Sl.	Disturbance initiating time (sec) For L-PID (a)	Instant of Stabilization (sec) For L-PID (b)	Time taken for stabilization(sec) by L-PID (c =b-a)
1	20.04	23.13	3.09
2	39.99	43.74	3.75
3	60.4	63.24	2.84
4	79.89	83.17	3.28

TABLE IV. PENDULUM ANGLE STABILIZATION RESPONSE FOR THE NL-PID CONTROLLER

Sl.	Disturbance initiating time (sec) For NL-PID (a)	Instant of Stabilization (sec) For NL-PID (b)	Time taken for stabilization(sec) by NL-PID (c =b-a)
1	20.5	22.69	2.19
2	40.5	42.76	2.26
3	60.51	62.74	2.23
4	80.37	82.63	2.26

V. RESULTS AND DISCUSSION

In this paper, multi-loop L-PID and multi-loop NL-PID schemes are designed and validated on a benchmarked gantry crane system. It is observed that the multi-loop NL-PID controller quickly brings the pendulum to the desired position under the external disturbances as compared to the multi-loop L-PID controller.

ACKNOWLEDGMENT

The authors would like to thank Govt. of India and Govt. NCT Delhi for providing funding through Delhi Technological University under TEQIP-II to procure Digital Pendulum System Experimental Setup.

REFERENCES

[1] K. J. Astrom and P. R. Kumar, "Control: A perspective," *Automatica*, vol. 50, no. 1, pp. 3–43, 2014.

[2] Sudarshan K. Valluru Madhusudan Singh Bharat Bhushan, "Comparative Analysis of PID, NARMA L-2 and PSO Tuned PID Controllers for Nonlinear Dynamical System," *J. Autom. Syst. Eng.*, vol. 9, no. 2, pp. 94–108, 2015.

[3] Astrom.K.J. and Hagglund.T, "The Future of PID Control," *Control Eng. Pract.*, vol. 9, pp. 1163–1175, 2001.

[4] S. K. Valluru and M. Singh, "Stabilization of Nonlinear Inverted Pendulum System Using MOGA and APSO Tuned Nonlinear PID Controller," *Cogent Eng.*, vol. 4, no. 1, pp. 1–15, 2017.

[5] S. K. Valluru and M. Singh, "Investigation of NARMA L-2 and Artificial Bee Colony Tuned PID Controllers for Bench Scaled Nonlinear Dynamical System," *Int. J. Control Theory Appl.*, vol. 10, no. 6, pp. 363–374, 2017.

[6] G. Zaidner, S. Korotkin, E. Shteimberg, A. Ellenbogen, M. Arad, Y. Cohen, and B. Sheva, "Non Linear PID and its application in Process Control," in *IEEE 26-th Convention of Electrical and Electronics Engineers in Israel*, 2010, pp. 574–577.

[7] L. Ramli, Z. Mohamed, A. M. Abdullahi, H. I. Jaafar, and I. M. Lazim, "Control strategies for crane systems: A comprehensive review," *Mech. Syst. Signal Process.*, vol. 95, pp. 1–23, 2017.

[8] FeedbackInstruments, "Digital Pendulum Control Experiments Manual," 2016.

View publication stats

Neural Computing and Applications
https://doi.org/10.1007/s00521-018-3796-3

ORIGINAL ARTICLE

Binary whale optimization algorithm and its application to unit commitment problem

Vijay Kumar[1] · Dinesh Kumar[2]

Received: 7 February 2017 / Accepted: 3 October 2018
© The Natural Computing Applications Forum 2018

Abstract
Whale optimization algorithm is a novel metaheuristic algorithm that imitates the social behavior of humpback whales. In this algorithm, the bubble-net hunting strategy of humpback whales is exploited. However, this algorithm, in its present form, is appropriate for continuous problems. To make it applicable to discrete problems, a binary version of this algorithm is being proposed in this paper. In the proposed approach, the solutions are binarized and sigmoidal transfer function is utilized to update the position of whales. The performance of the proposed algorithm is evaluated on 29 benchmark functions. Furthermore, unpaired *t* test is carried out to illustrate its statistical significance. The experimental results depict that the proposed algorithm outperforms others in respect of benchmark test functions. The proposed approach is applied on electrical engineering problem, a real-life application, named as "unit commitment." The proposed approach uses the priority list to handle spinning reserve constraints and search mechanism to handle minimum up/down time constraints. It is tested on standard IEEE systems consisting of 4, 10, 20, 40, 80, and 100 units and on IEEE 118-bus system and Taiwan 38-bus system as well. Experimental results reveal that the proposed approach is superior to other algorithms in terms of lower production cost.

Keywords Whale optimization · Binary whale optimization · Metaheuristics · Benchmark functions · Economic load dispatch · Unit commitment

1 Introduction

In the recent years, it has been observed that more attention is being paid to metaheuristic algorithms which are inspired from the natural phenomena. The well-known nature-inspired metaheuristic algorithms are genetic algorithm (GA) [1], particle swarm optimization (PSO) [2], ant colony optimization [3], bat algorithm (BA) [4], gravitational search algorithm (GSA) [5], etc. These approaches are able to solve real-life computational problems. However, these are developed to solve some specific problems not all types of problems [6]. Hence, novel algorithms are developed every year to solve complex real-life computational problems.

Whale optimization algorithm (WOA) [7] is a recently developed metaheuristic algorithm, which mimics the hunting behavior of whales. It simulates bubble-net attacking mechanism of whales. The WOA was applied to solve many engineering design problems successfully [7]. The original version of WOA was developed for continuous search space. However, several real-life optimization problems require binary search spaces such as feature selection [8], data mining [9], cell formation [10], and unit commitment [11]. The binary algorithms are required to solve these problems. This fact motivates us to develop binary version of WOA. In this paper, binary version of WOA (BWOA) is used to solve the unit commitment problem.

The unit commitment problem (UCP) is a well-known optimization problem in electrical engineering domain.

✉ Vijay Kumar
vijaykumarchahar@gmail.com

Dinesh Kumar
dinesh_chutani@yahoo.com

[1] Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala, Punjab, India

[2] Electronics and Communication Engineering Department, Delhi Technological University, Delhi, India

Published online: 16 October 2018

UCP deals with determining the on/off states of power generating units to minimize the operating cost for a given time [12]. The main goal of UCP is to find the optimized schedule to on/off the power generating units to meet the electric power demand while satisfying constraints such as spinning reserve, load balance, minimum up/down time, etc. [13]. Over the last few decades, a number of techniques have been developed to solve UCP.

The techniques used for UCP are broadly classified into three categories. These are classical, metaheuristics and hybrid techniques [13]. The first one has classical optimization techniques such as priority list (PL) [14], Lagrangian relaxation (LR) [15], mixed integer linear programming (MILP) [16], branch and bound approaches (BB) [17], dynamic programming (DP) [18]. The main advantages of these techniques are fast convergence and the simplest form of representation. However, these techniques suffer from poor solution quality (in case of PL), computational time increasing exponentially with dimensionality of system (in case of BB), and problems with dimensionality of system (in case of DP) [13]. These problems have directed to use metaheuristic techniques.

The metaheuristic techniques are successfully used to solve the UCP. These are genetic algorithm (GA) [19], enhanced simulated annealing (SA) [20], binary particle swarm optimization (BPSO) [21], evolutionary programming (EP) [22], binary grey wolf algorithm (BGWA) [13], imperialistic competition algorithm (ICA) [23], etc. The hybrid techniques are developed by integrating both classical and metaheuristic techniques to solve the UCP. The well-known hybrid techniques are Lagrangian relaxation particle swarm optimization (LRPSO) [24], hybrid genetic-ICA (HGICA) [25], hybrid harmony search random search (HHSRSA) [26], local convergence-averse BPSO (LCA-PSO) [27], and hybrid particle swarm optimization grey wolf (PSO-GWO) [28]. A large number of nature-inspired metaheuristic techniques for solving UCP are reported in the literature [29].

The main contribution of this paper is to develop a binary version of whale optimization algorithm (BWOA) for finding optimal regions of the search space. In the original version of WOA, bubble-net mechanism changes the position of search points in continuous search space. On the contrary, in BWOA, the outcome of bubble-net mechanism is converted into a probability value for each component of binary vector that dictates whether that component will take 0 or 1 value. For solving unit commitment problem, the proposed approach uses priority list to handle spinning reserve constraints and search mechanism to handle minimum up/down time constraints efficiently. The proposed approach is applied on 29 standard benchmark test functions. It is compared with seven recently developed metaheuristic techniques. It is also applied on the unit commitment problem with varying number of units from 10 to 100 and complex power systems.

The rest of this paper is organized as follows. Section 2 presents the basic concepts of whale optimization algorithm. The proposed binary version of whale optimization algorithm is described in Sect. 3. In Sect. 4, the proposed approach is investigated on a set of benchmark test functions and compared with existing metaheuristic techniques. Section 5 describes a binary version of whale optimization algorithm for unit commitment problem. The experimental results are discussed in Sect. 4. Section 6 outlines the conclusions.

2 Whale optimization algorithm

Mirjalili and Lewis proposed a whale optimization algorithm (WOA), which is inspired from bubble-net hunting strategy of humpback whales. The humpback whales have a preference to hunt small fishes near the surface. Therefore, they swim near the prey by generating distinctive bubbles along a circle. WOA mimics two main phases. The first phase is exploitation, i.e., encircling a prey and spiral bubble-net attacking method. The second phase is exploration, i.e., searching the prey. The mathematical modeling of whale optimization algorithm (WOA) is given below [7].

2.1 Encircling prey

Humpback whales encircle the prey (i.e., small fishes). They update their position toward the best search agent during the course of iteration. The following equations are used to model their behavior.

$$\text{Dist} = |C \times Y_B(t_{\text{cur}}) - Y(t_{\text{cur}})| \tag{1}$$

$$Y(t_{\text{cur}} + 1) = Y_B(t_{\text{cur}}) - A \times \text{Dist} \tag{2}$$

where Dist represents the distance between the position of best solution (Y_B) and current solution (Y). t_{cur} is the current iteration. $Y_B(t_{\text{cur}})$ represents the position of best solution. A and C are the coefficient vectors that are computed as follows [7]:

$$A = 2 \times au \times rand_1 - au \tag{3}$$

$$C = 2 \times rand_2 \tag{4}$$

The value of au is linearly decreased from 2 to 0 during the course of iterations. The $rand_1$ and $rand_2$ are random variables. The value of these variables lies in the range $[0, 1]$.

2.2 Spiral bubble-net attacking method

The two approaches are used to model the bubble-net behavior of humpback whales. These are shrinking and encircling, and spiral updating mechanism. The shrinking and encircling mechanism is achieved by decreasing the value of *au* linearly making A to have values in the range $[-1, 1]$.

The spiral equation between the positions of humpback whale and prey to mimic the helix-shaped movement of humpback whales is given below:

$$Y(t_{\text{cur}}+1) = \text{Dist}' \times e^{bl} \times cos(2\pi l) + Y(t_{\text{cur}}) \tag{5}$$

Here,

$$Dist' = Y_B(t_{\text{cur}}) - Y(t_{\text{cur}}) \tag{6}$$

where b is a constant parameter for defining the shape of logarithmic spiral, l is a random number that lies in range of $[0, 1]$.

To simulate these approaches simultaneously, a suitable assumption is made, that is, to update the position of whales, and it is equally likely to choose either the shrinking and encircling or spiral path. The mathematical modeling of these mechanisms is given below:

$$Y(t_{\text{cur}}+1) = \begin{cases} Y(t_{\text{cur}}) - A \times \text{Dist} & \text{if } Rand < 0.5 \\ \text{Dist}' \times e^{bl} \times \cos(2\pi l) + Y(t_{\text{cur}}) & \text{if } Rand \geq 0.5 \end{cases} \tag{7}$$

where *Rand* is a random number in the range $[0, 1]$.

2.3 Prey search

The prey search is done through the variation of A. For the exploration process, the value of A should be > 1. The mathematical modeling for prey search is given below:

$$\text{Dist}'' = |C \times Y_{rand} - Y| \tag{8}$$

$$Y(t_{\text{cur}}+1) = Y_{rand} - A \times \text{Dist}'' \tag{9}$$

where Y_{rand} is a random whale chosen from the population. The pseudocode of WOA is depicted in Fig. 1 [7].

3 Binary whale optimization algorithm

3.1 Motivation

There are many real-life engineering problems that require discrete binary search space. The well-known problems are feature selection [8], dimensionality reduction [8], data mining [9], and unit commitment [11]. For these problems, the solutions should be encoded in binary vectors. Besides this, the search space should also be considered as binary space. The binary search space can be considered as a hypercube in which the search agents (whales) may go to nearer and farther corners of the hypercube by flipping different number of bits [30].

Hussien et al. [31] and Eid [32] proposed two binary versions of WOA. They proposed change in only one phase, i.e., prey search thus leaving the other two phases, namely shrinking and encircling phases, untouched. The bubble-net behavior of whales also needed modification. These facts motivated us to develop a novel binary version of WOA that incorporates all the modifications mentioned as above. These phases are to be adapted in the discrete search space thus making us move one step further so far as the position updating mechanism of whales in proposed BWOA is concerned.

3.2 Mathematically foundation

For designing the binary version of WOA, it is an imperative to make changes in some basic concepts of original WOA. Due to the binary search space, each dimension can take either 0 or 1 value. Therefore, the movement of search agents (whales) through this dimension may change from 0 to 1 or vice versa.

The foremost difference between original and binary version of WOA is that of position updating mechanism. In BWOA, the toggling between the values 0 and 1 indicates position updating. Furthermore, the value of current bit is changed with a probability that is computed in accordance with helix-shaped movement of whale.

To achieve this, an appropriate transfer function is required to map the helix-shaped movement values into probability values of position updating. The transfer function forces whales to travel in a binary space. Based on the above-mentioned concept, an appropriate probability function can be formulated as:

$$Cstep = \frac{1}{1 + e^{-10(A \times \text{Dist} - 0.5)}} \tag{10}$$

where *Cstep* is the step size that can be computed using sigmoidal function. Dist is the distance between the position of prey and humpback whale.

3.3 Binary whale optimization algorithm (BWOA)

Based on the concept mentioned above in Sect. 3.2, the three major changes in original whale optimization algorithm are proposed.

First, the shrinking and encircling prey phase is modified. The position of whale is modified according to the equation mentioned below:

Fig. 1 Pseudocode of whale optimization algorithm (WOA) [7]

Whale Optimization Algorithm (WOA)

Input:
Number of whales in the pack, n;
Control Coefficient, au;
Maximum number of iterations, MAX_ITER;

Output:
Global best whale position, Y_{best}
Best fitness value, $fit(Y_{best})$

begin
Generate initial population of n whales $Y_i \ (i = 1, 2, ..., n)$
Set iteration counter $t_{cur} = 0$
Compute the fitness of each whale
Identify the best whale based on fitness, i.e., Y_{best}
while $(t_{cur} < MAX_ITER)$
for *each whale* ***do***
Compute control coefficients A and C by Eqs. 3 and 4
if $(Rand < 0.5)$
if $(|A| < 1)$
Update the position of the current whale by Eq.2
else if $(|A| \geq 1)$
Select a random whale, Y_{rand}
Update the position of the current whale by Eq. 9
end if
else if $(Rand \geq 0.5)$
Update the position of the current whale by Eq. 5
end if
end for
Compute the fitness of all whales
Update the value of Y_{best} *based on fitness*
Increment the current iteration t_{cur} *by 1.*
end while
return the best solution, Y_{best}
end

$$Y(t_{\text{cur}}+1) = \begin{cases} \text{complement}(Y(t_{\text{cur}})), & \text{if } rand < C_{\text{step}} \\ Y(t_{\text{cur}}), & \text{otherwise} \end{cases} \tag{11}$$

where *Cstep* is computed as is given by Eq. (10). *A* and Dist are computed using Eqs. (1) and (3), respectively.

The second modification is done in the bubble-net behavior of whales that makes use of *Cstep′* computed as follows:

$$Cstep' = \frac{1}{1 + e^{-10(A \times \text{Dist}' - 0.5)}} \tag{12}$$

where, *A* and *Dist′* are computed using Eqs. (3) and (6), respectively.

The position of helix-shaped movement of humpback whales is updated according to Eq. (5). Thereafter, the modification in position updating process is done as follows:

$$Y(t_{\text{cur}}+1) = \begin{cases} \text{complement}(Y(t_{\text{cur}})), & \text{if } rand < Cstep' \\ Y(t_{\text{cur}}), & \text{otherwise} \end{cases} \tag{13}$$

The third modification is done in the searching of prey. The mathematical formulation of *Cstep″* is given below:

$$Cstep'' = \frac{1}{1 + e^{-10(A \times Dist'' - 0.5)}} \tag{14}$$

where, A and $Dist''$ are computed using Eqs. (3) and (8), respectively. Hence, the position of whale is updated according to Eq. (15).

$$Y(t_{\text{cur}} + 1) = \begin{cases} \text{complement}(Y(t_{\text{cur}})), & \text{if } rand < Cstep'' \\ Y(t_{\text{cur}}), & \text{otherwise} \end{cases} \tag{15}$$

The general steps of Binary version of WOA are depicted in Fig. 2. The changes are highlighted (in red color).

3.4 Computational analysis

In this section, the complexity analysis of BWOA is presented. The time and space complexities of BWOA are described (Fig. 3).

3.4.1 Time complexity

The time complexity of intermediate states of the proposed BWOA is given below.

1. Initialization of BWGO requires $O(n \times D)$ time where n denotes the number of whales and D is the dimension of the benchmark test function.

Fig. 2 Pseudocode of the binary whale optimization algorithm (BWOA)

Binary Whale Optimization Algorithm (BWOA)

Input:
Number of whales in the pack, n;
Control Coefficient, au;
Maximum number of iterations, MAX_ITER;

Output:
Global best binary whale position, Y_{best}
Best fitness value, $fit(Y_{best})$

begin
Generate initial population of n whales $Y_i \ (i = 1, 2, ..., n)$
Set iteration counter $t_{cur} = 0$
Compute the fitness of each whale
Identify the best whale based on fitness, i.e., Y_{best}
while $(t_{cur} < MAX_ITER)$
for *each whale* ***do***
Compute control coefficients A, and C by Eqs. 3 and 4
if $(Rand < 0.5)$
if $(|A| < 1)$
Update the position of the current whale by Eq.11
else if $(|A| \geq 1)$
Select a random whale, Y_{rand}
Update the position of the current whale by Eq. 15
end if
else if $(Rand \geq 0.5)$
Update the position of the current whale by Eq. 13
end if
end for
Compute the fitness of all whales
Update the value of Y_{best} *based on fitness*
Increment the current iteration t_{cur} *by 1.*
end while
return the best solution, Y_{best}
end

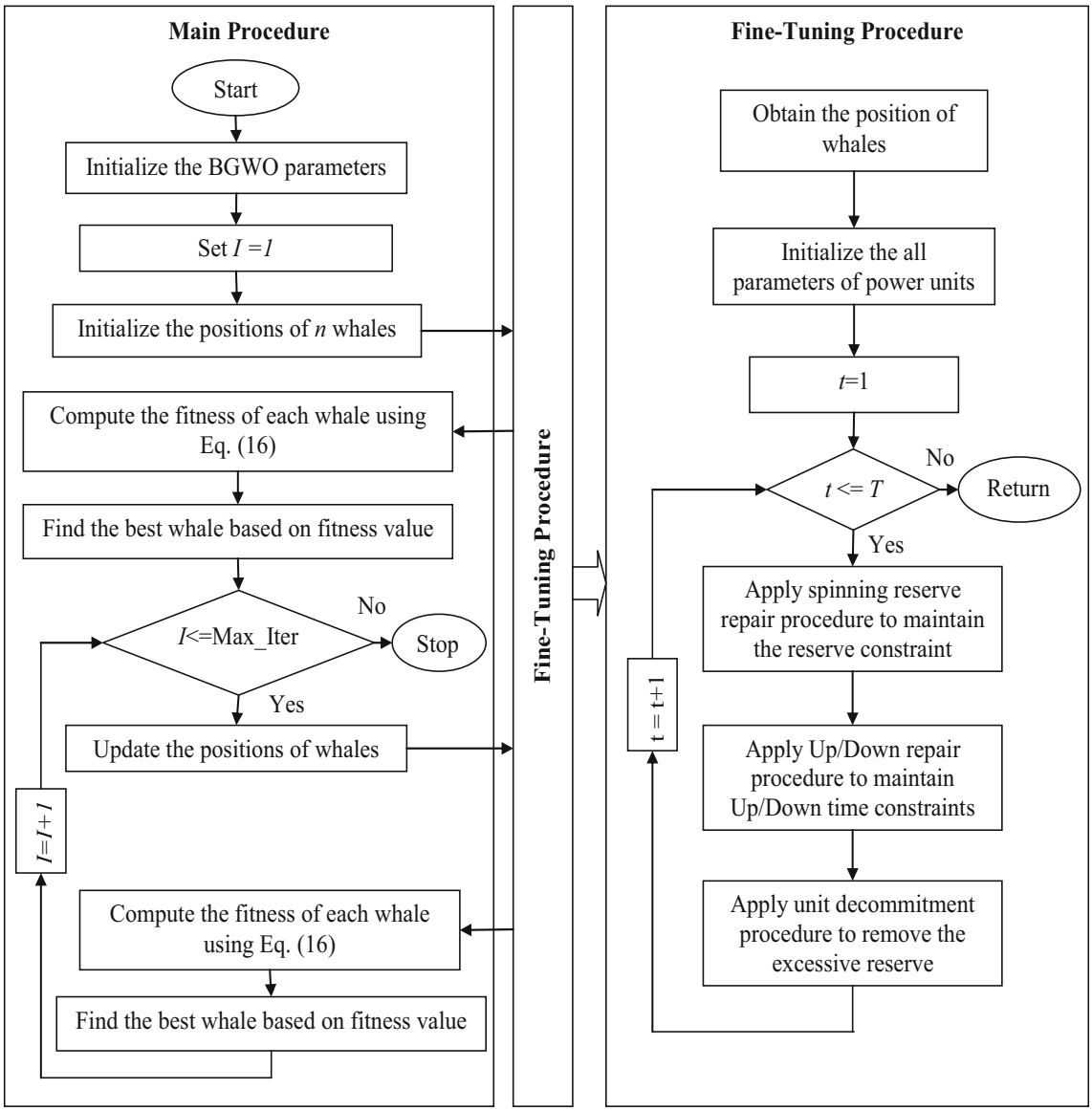


Fig. 3 Flowchart for unit commitment problem using binary whale optimization algorithm

2. Computation of fitness function requires $O(n \times D)$ time.
3. Computation of control parameters and position update steps of BWOA requires $O(n \times D)$ each.

Therefore, summing up the complexities of the above steps and the total time complexity of BWOA becomes $O(n \times D)$ per iteration/generation. The total time complexity of BWOA for maximum number of iterations is $O(n \times D \times \text{MaxIteration})$. Here, MaxIteration denotes the maximum number of iterations/generations.

3.4.2 Space complexity

The space requirement of BWOA is due to its initialization process. Thus, the total space complexity of BWOA is $O(n \times D)$.

4 Experimental results

To demonstrate the performance of proposed binary whale optimization algorithm, a set of 29 well-known benchmark test functions [33, 34] are utilized, and the results are reported along with seven well-known metaheuristic algorithms. The simulations are carried out in MATLAB

R2016b environment operating on Windows 7 and 2.2 GHz processor with 8 GB RAM.

4.1 Benchmark test functions used

The benchmark test functions are divided into four categories. These are unimodal, multimodal having many local minima, multimodal having no local minima, and composite functions, which are described in Tables 22, 23, 24, and 25 in Appendix, respectively [33, 34]. In these tables, D indicates dimension of the search space. The first category includes seven test functions $(F_1 - F_7)$. The second category comprises six test functions, i.e., $(F_8 - F_{13})$. The third category consists of 10 test functions, i.e., $(F_{14} - F_{23})$. The fourth category comprises six composite test functions, i.e., $(F_{24} - F_{29})$. These functions are minimization problems.

4.2 Algorithms for comparisons and parameter setting

To assess the performance of proposed algorithm, it is compared with seven recently developed metaheuristic techniques such as genetic algorithm (GA) [1], Grey wolf algorithm (GWA) [35], binary particle swarm optimization (BPSO) [36], binary gravitational search algorithm (BGSA) [30], binary bat algorithm (BBA) [37], binary dragonfly algorithm (BDA) [38], and binary grey wolf algorithm (BGWA) [39]. The parameter settings of above-mentioned algorithms used in this paper are mentioned in Table 1. All the parameters are set according to trial and error on small simulations and are commonly reported in literature.

In addition to these parameters, the maximum number of iterations and population size for the above-mentioned algorithms are set to 500 and 30, respectively. Due to stochastic nature of these algorithms, the results are averaged over 30 independent runs under 30 different random seeds. The mean best-of-run solution and standard deviation of best solution in the last iteration are reported in tables.

Table 1 Parameter setting for algorithms used

Algorithm	Parameters	Values
GA	Selection procedure	Roulette wheel
	Crossover (probability)	Single-point (0.90)
	Mutation (probability)	Uniform (0.005)
GWA	Control parameter (au)	Decrease from 2 to 0
BPSO	Constant parameters $(c_1,\ c_2)$	2, 2
	Inertia weight	Decrease from 0.9 to 0.4
	Maximum velocity	6
BGSA	Gravitational constant (G_0)	100
BBA	Frequency range	[0, 2]
	Loudness	0.25
	Pulse emission rate	0.5
	Constant parameter (γ)	0.9
BDA	Cohesion weight	0.7
	Separation weight	0.1
	Alignment weight	0.1
	Inertia weight	Decrease from 0.9 to 0.2
	Enemy factor	1
BGWA	Control parameter (au)	Decrease from 2 to 0
BWOA	au	Decrease from 2 to 0
	b	1

4.3 Results and discussion

The performance of proposed approach is compared with above-mentioned algorithms on benchmark test functions mentioned in Sect. 4.1.

4.3.1 Estimation of exploitation ability

The unimodal benchmark test functions $(F_1 - F_7)$ have only one global optimum. These functions are used to test the exploitation capability of examined metaheuristic algorithms. Table 2 shows the results for unimodal test functions. The table demonstrates that BWOA offers better results than the other metaheuristic algorithms on all unimodal test functions in terms of mean and standard deviation. The results also reveal that the BWOA is able to find global optimum in unimodal test functions. Hence, the proposed algorithm can provide good exploitation.

4.3.2 Estimation of exploration ability

The multimodal benchmark test functions $(F_8 - F_{23})$ have many local optima whose number increases exponentially with increase in problem size. The purpose of these functions is to test the exploration capability of examined metaheuristic algorithms. The results for multimodal test functions $(F_8 - F_{13})$ and fixed dimension multimodal test functions $(F_{14} - F_{23})$ are reported in Tables 3 and 4, respectively. These tables reveal that the proposed BWOA gives better results than the other metaheuristic algorithms on most of multimodal test functions in terms of mean and standard deviation. The results reveal that the BWOA is able to locate global optimum in multimodal test functions. Hence, the proposed algorithm can provide very good exploration capability.

Table 2 Mean and standard deviation of the best-of-run solution on unimodal test functions

Fun	Mean best-of-run solution (SD)							
	GA	GWA	BPSO	BGSA	BBA	BDA	BGWA	BWOA
F_1	1.01E+01	5.29E+00	5.29E+00	2.05E+03	1.85E+00	2.82E−01	1.38E−01	**0.00E+00**
	(2.49E+01)	(9.34E−02)	(2.76E+00)	(4.14E+01)	(2.49E+00)	(4.17E−01)	(1.50E−01)	**(0.00E+00)**
F_2	2.69E−01	1.56E−01	2.22E−01	1.32E+00	9.65E−02	5.89E−02	3.42E−02	**0.00E+00**
	(2.37E−01)	(1.94E−02)	(9.38E−02)	(6.73E−01)	(6.46E−02)	(6.93E−02)	(4.69E−02)	**(0.00E+00)**
F_3	5.55E+02	7.01E+01	2.24E+01	5.09E+02	7.81E+00	1.42E+01	5.10E+00	**0.00E+00**
	(2.51E+02)	(2.19E+01)	(1.02E+03)	(2.66E+02)	(9.79E+00)	(2.26E+01)	(1.84E+00)	**(0.00E+00)**
F_4	1.59E+00	1.31E+00	2.60E+00	7.99E+00	1.15E+00	2.48E−01	1.27E−01	**0.00E+00**
	(1.21E+00)	(3.17E−01)	(8.38E−01)	(3.45E+00)	(6.14E−01)	(2.31E−01)	(3.55E−02)	**(0.00E+00)**
F_5	3.69E+02	9.75E+01	1.48E+02	2.62E+03	25.07E+00	23.55E+00	17.68E+00	**0.00E+00**
	(3.42E+02)	(6.01E+01)	(1.37E+02)	(1.73E+03)	(2.84E+01)	(3.46E+01)	(2.19E+01)	**(0.00E+00)**
F_6	6.98E+00	8.16E−01	8.49E+00	1.26E+02	2.69E+00	9.53E−02	4.45E−02	**0.00E+00**
	(7.01E+00)	(1.26E−02)	(6.14E+00)	(8.80E+01)	(2.74E+00)	(1.29E−01)	(1.06E−01)	**(0.00E+00)**
F_7	4.71E−02	2.21E−02	1.55E−02	2.28E−02	6.00E−03	1.22E−02	3.16E−03	**9.32E−04**
	(4.35E−02)	(1.03E−03)	(7.47E−03)	(2.68E−02)	(4.40E−03)	(1.46E−02)	(2.72E−03)	**(8.73E−04)**

Bold values indicate the best results

Table 3 Mean and standard deviation of the best-of-run solution on multimodal test functions

Fun	Mean best-of-run solution (SD)							
	GA	GWA	BPSO	BGSA	BBA	BDA	BGWA	BWOA
F_8	− 9.29E+02	− 6.12E+02	**− 9.88E+02**	− 8.28E+02	− 9.85E+02	− 9.24E+02	− 9.27E+02	− 9.21E+02
	(2.79E+01)	(1.15E+01)	**(1.42E+01)**	(4.26E+01)	(2.76E+01)	(6.57E+01)	(4.15E+01)	(1.00E−31)
F_9	2.18E+00	1.59E+00	4.97E+00	5.99E+00	1.58E+00	1.80E+00	1.40E+00	**0.00E+00**
	(1.83E+00)	(1.33E+00)	(1.59E+00)	(2.96E+00)	(1.33E+00)	(1.05E+00)	(1.21E+00)	**(0.00E+00)**
F_{10}	1.39E+00	1.89E+00	2.72E+00	2.94E+00	1.15E+00	3.88E−01	2.72E−01	**8.88E−16**
	(1.34E+00)	(5.06E−01)	(4.72E−01)	(1.48E+00)	(7.28E−01)	(5.71E−01)	(3.00E−01)	**(0.00E+00)**
F_{11}	7.06E−01	3.21E−01	3.87E−01	6.47E−01	2.46E−01	1.93E−01	6.59E−02	**0.00E+00**
	(3.22E−01)	(1.78E−02)	(1.32E−01)	(2.28E−01)	(8.39E−02)	(1.13E−01)	(1.76E−02)	**(0.00E+00)**
F_{12}	1.91E−01	2.76E−01	6.21E−01	1.26E+01	2.71E−01	1.49E−01	1.82E−01	**1.12E−01**
	(2.44E−01)	(1.14E−01)	(3.88E−01)	(8.77E+00)	(3.28E−01)	(4.51E−01)	(2.97E−01)	**(0.00E+00)**
F_{13}	1.93E−01	1.67E−01	4.44E−01	9.22E−01	1.29E−01	3.52E−02	2.64E−02	**1.35E−32**
	(2.54E−01)	(3.10E−02)	(2.11E−01)	(2.45E−01)	(7.36E−02)	(5.65E−02)	(6.79E−02)	**(5.56E−48)**

Bold values indicate the best results

4.3.3 Ability to escape from local minima

The composite benchmark test functions $(F_{24} - F_{29})$ are very difficult to optimize as proper balance between exploration and exploitation is required. The optimization results of these functions are reported in Table 5. As seen from this table, the proposed BWOA outperforms other competitive algorithms on all composite functions in terms of mean and standard deviation. The results reveal that the proposed approach provides better balance between exploration and exploitation. The proposed approach uses the concept of $|A| > 1$ to explore the search space and $|A| < 1$ is responsible for exploitation.

4.3.4 Test for statistical significance

Besides the basic statistical analysis (i.e., mean and standard deviation), the unpaired t test is used to test whether the proposed approach differs or not from the rest of the competitor algorithms in a statistically significant way. The sample size for this test is taken as 30. Table 6 illustrates the statistical results based on the optimization result

Table 4 Mean and standard deviation of the best-of-run solution on fixed multimodal test functions

Fun	Mean best-of-run solution (SD)							
	GA	GWA	BPSO	BGSA	BBA	BDA	BGWA	BWOA
F_{14}	4.30E+00	4.04E+00	9.98E−01	1.00E+00	1.20E+00	9.80E−01	9.27E−01	**8.70E−01**
	(2.48E+01)	(4.25E+00)	(6.37E−07)	(8.88E−05)	(6.70E−05)	(7.54E−06)	(5.65E−08)	**(2.11E−09)**
F_{15}	1.29E−02	3.37E−03	9.00E−04	2.10E−03	1.48E−03	2.04E−03	1.34E−03	**4.36E−04**
	(8.29E−05)	(6.25E−04)	(2.65E−08)	(2.99E−06)	(2.71E−06)	(1.58E−06)	(1.42E−06)	**(1.92E−08)**
F_{16}	− 8.93E−01	− 1.03E+00	− 1.02E+00	− 1.02E+00	− 7.68E−01	− 9.26E−01	− 1.05E+00	**− 1.03E+00**
	(1.70E−02)	(5.88E−03)	(9.95E−05)	(4.90E−05)	(2.04E−02)	(1.94E−02)	(3.94E−05)	**(5.01E−05)**
F_{17}	4.39E−01	3.97E−01	3.99E−01	3.99E−01	2.80E−01	2.36E−01	2.39E−01	**2.28E−01**
	(1.12E−02)	(1.51E−04)	(5.23E−07)	(2.99E−10)	(1.47E−02)	(1.27E−03)	(1.62E−05)	**(1.11E−09)**
F_{18}	8.30E+00	3.00E+00	3.00E+00	3.03E+00	2.95E+00	2.91E+00	2.92E+00	**2.90E+00**
	(1.25E+02)	(2.29E−03)	(2.95E−05)	(2.00E−03)	(1.20E−02)	(1.00E−03)	(1.21E−04)	**(1.00E−06)**
F_{19}	− 3.41E+00	− 3.32E+00	− 3.85E+00	− 3.85E+00	− 3.56E+00	− 3.30E+00	− 3.85E+00	**− 3.86E+00**
	(5.76E−02)	(2.30E−02)	(2.98E−06)	(2.94E−07)	(1.19E−02)	(2.24E−03)	(2.87E−04)	**(8.89E−09)**
F_{20}	− 1.75E+00	− 3.28E+00	− 3.11E+00	− 3.30E+00	− 1.35E+00	− 3.05E+00	− 3.16E+00	**− 3.31E+00**
	(2.68E−01)	(6.05E−03)	(2.30E−03)	(1.40E−03)	(3.70E−02)	(2.69E−02)	(1.95E−02)	**(1.06E−05)**
F_{21}	− 1.08E+00	− 6.86E+00	− 6.35E+00	− 3.56E+00	− 3.23E+00	− 3.99E+00	− 4.55E+00	**− 7.89E+00**
	(8.14E−01)	(3.01E+00)	(7.70E+00)	(1.06E+00)	(2.11E+00)	(1.10E+00)	(2.09E+00)	**(2.28E−01)**
F_{22}	− 9.69E−01	− 8.45E+00	− 7.50E+00	− 5.16E+00	− 4.11E+00	− 5.65E+00	− 6.82E+00	**− 9.89E+00**
	(1.20E−01)	(3.08E+00)	(5.63E+00)	(8.81E+00)	(1.80E+00)	(2.75E+00)	(2.31E+00)	**(6.35E−01)**
F_{23}	− 1.44E+00	− 6.57E+00	− 5.75E+00	− 3.57E+00	− 3.18E+00	− 4.02E+00	− 4.73E+00	**− 8.28E+00**
	(7.90E−01)	(2.14E+00)	(2.60E+00)	(1.29E+00)	(2.12E+00)	(1.90E+00)	(2.43E+00)	**(3.60E+00)**

Bold values indicate the best results

Table 5 Mean and standard deviation of the best-of-run solution on composite test functions

Fun	Mean best-of-run solution (SD)							
	GA	GWA	BPSO	BGSA	BBA	BDA	BGWA	BWOA
F_{24}	1.93E+02	1.00E+02	1.94E+02	2.52E+02	9.32E+01	9.27E+01	9.10E+01	**7.09E+01**
	(1.21E+02)	(8.16E+01)	(6.00E+01)	(4.36E+01)	(6.43E+01)	(4.20E+01)	(4.09E+01)	**(3.98E+01)**
F_{25}	2.05E+02	1.55E+02	1.46E+02	2.54E+02	1.56E+02	1.28E+02	1.31E+02	**1.12E+02**
	(1.60E+02)	(2.31E+01)	(2.90E+01)	(3.62E+01)	(3.18E+01)	(2.05E+01)	(2.79E+01)	**(1.58E+01)**
F_{26}	3.84E+02	1.72E+02	4.45E+02	2.13E+02	1.49E+02	1.03E+02	1.19E+02	**9.62E+01**
	(1.18E+02)	(3.27E+01)	(4.93E+01)	(6.36E+01)	(3.87E+01)	(2.76E+01)	(1.02E+01)	**(9.03E+00)**
F_{27}	5.88E+02	3.14E+02	4.79E+02	2.55E+02	1.46E+02	1.16E+02	1.22E+02	**9.98E+01**
	(1.02E+02)	(2.06E+01)	(3.01E+01)	(5.01E+01)	(2.29E+01)	(1.98E+01)	(1.56E+01)	**(1.03E+01)**
F_{28}	2.46E+02	1.83E+02	1.72E+02	2.45E+02	1.66E+02	1.46E+02	1.37E+02	**1.12E+02**
	(1.83E+02)	(1.01E+01)	(6.42E+01)	(5.28E+01)	(4.98E+01)	(3.55E+01)	(2.93E+01)	**(2.44E+01)**
F_{29}	9.14E+02	8.61E+02	6.91E+02	2.32E+02	1.52E+02	1.34E+02	1.28E+02	**9.94E+01**
	(1.23E+01)	(1.25E+02)	(1.49E+02)	(2.76E+01)	(3.36E+01)	(2.77E+01)	(2.09E+01)	**(1.88E+01)**

Bold values indicate the best results

obtained from benchmark test functions. According to the p values in Table 6, BWOA is significantly different from other competitive algorithms for all unimodal benchmark functions $(F_1 - F_7)$. The p values reported in Table 6 show that BWOA do not only have a significant improvement in F_{12} and F_{21}, but this attains significant improvement in all the remaining multimodal benchmark test functions $(F_8 - F_{11},\ F_{13} - F_{20},\ and\ F_{22} - F_{23})$ also. BWOA is significantly better in five out of six composite functions based on the p values reported in Table 6. Therefore,

Table 6 Unpaired t test among BWOA and second-best algorithm

	SE	t	95% CI	Two-tailed p	Significance
F_1	0.076	3.7040	− 0.43771042 to − 0.12628958	0.0009	Extremely significant
F_2	0.013	4.6552	− 0.08477706 to − 0.03302294	< 0.0001	Extremely significant
F_3	1.787	4.3695	− 11.38787195 to − 4.23212805	< 0.0001	Extremely Significant
F_4	0.042	5.8803	− 0.33242170 to − 0.16357830	< 0.0001	Extremely significant
F_5	6.317	3.7280	− 36.19498157 to − 10.90501843	0.0004	Extremely significant
F_6	0.024	4.0464	− 0.14244458 to − 0.04815542	0.0002	Extremely significant
F_7	0.001	6.1881	0.00342862 to 0.00670738	< 0.0001	Extremely significant
F_8	3.042	22.2100	− 73.65441381 to − 61.47558619	< 0.0001	Extremely significant
F_9	0.243	6.5068	− 2.06606432 to − 1.09393568	< 0.0001	Extremely significant
F_{10}	0.104	3.7218	− 0.59667874 to − 0.17932126	0.0004	Extremely significant
F_{11}	0.021	9.3549	− 0.23429719 to − 0.15170281	< 0.0001	Extremely significant
F_{12}	0.082	0.4494	− 0.12782331 to 0.20182331	0.0004	Not significant
F_{13}	0.010	3.4124	0.01455140 to 0.05584860	0.0012	Statistically significant
F_{14}	0.000	7906.47	− 0.11000276 to − 0.10999724	< 0.0001	Extremely significant
F_{15}	0.000	2109.99	− 0.00104499 to − 0.00104301	< 0.0001	Extremely significant
F_{16}	0.000	7253.21	− 9.28002561 to − 9.27997439	< 0.0001	Extremely significant
F_{17}	0.000	34.5022	0.00753586 to 0.00846414	< 0.0001	Extremely significant
F_{18}	0.000	54.7722	0.00963454 to 0.01036546	< 0.0001	Extremely significant
F_{19}	0.000	1862.07	0.00999989 to 0.01000011	< 0.0001	Extremely significant
F_{20}	0.000	39.1219	− 0.01051166 to − 0.00948834	< 0.0001	Extremely significant
F_{21}	1.406	1.0950	− 1.27528997 to 4.35528997	0.2781	Not significant
F_{22}	1.034	2.3105	0.31940356 to 4.46059644	0.0244	Statistically significant
F_{23}	0.811	3.1205	− 4.15291419 to − 0.90708581	0.0028	Statistically significant
F_{24}	10.564	2.0636	0.65355238 to 42.94644762	0.0435	Statistically significant
F_{25}	4.725	3.3859	6.54102997 to 25.45897003	0.0013	Statistically significant
F_{26}	5.302	1.2826	− 3.81288318 to 17.41288318	0.2047	Not significant
F_{27}	4.075	3.9756	8.04331950 to 24.35668050	0.0002	Extremely significant
F_{28}	7.865	4.3231	18.25707400 to 49.74292600	0.0001	Extremely significant
F_{29}	6.112	5.6609	22.36532040 to 46.83467960	< 0.0001	Extremely significant

BWOA is statistically significant over other competitive algorithms in most of benchmark functions.

5 Application to unit commitment problem

In this section, the performance of BWOA is evaluated by solving a real-life problem in electrical engineering. The unit commitment problem (UCP) is one of the most important problems in electrical engineering. Hence, the proposed BWOA is being applied to UCP.

5.1 Mathematical formulation of UCP

5.1.1 Objective function

The foremost objective of UCP is to find the optimal schedule for operating the available generation units such that the total generation cost should be minimized. The total power generation cost consists of fuel cost, start-up and shutdown costs. The fuel cost comprises fuel price, heat rate of generating units, turn-on, turn-off, and initial status of units. The total fuel cost (T_{fc}) over time t is given below [40, 41]:

$$T_{\mathrm{fc}} = \sum_{t=1}^{T}\sum_{i=1}^{N}\left[f_{\mathrm{cost}}\left(P_i^t\right) + SUC_i^t\left(1-u_i^{t-1}\right)\right]u_i^t \tag{16}$$

where N is the number of generating units; T is total scheduling time; $f_{\mathrm{cost}}\left(P_i^t\right)$ is the fuel cost of i^{th} unit at time t; u_i^t is the on/off status of ith unit at time t; P_i^t is power generation of ith unit at time t; SUC_i^t is the start-up cost of ith unit at time t.

The mathematical representation of fuel cost function is given below [42, 43]:

$$f_{\text{cost}}\left(P_i^t\right) = \alpha_i + \beta_i P_i^t + \gamma_i \left(P_i^t\right)^2 \tag{17}$$

where α_i, β_i and γ_i are the fuel cost coefficients of ith unit. The start-up cost can be defined as [28]:

$$\text{SUC}_i^t = \begin{cases} \text{HSC}_i^t, & \forall \text{MDT}_i^t \le \text{MDT}_{i,on}^t \le \left(\text{MDT}_i^t + \text{CSH}_i^t\right) \\ \text{CSC}_i^t, & \forall \text{MDT}_{i,on}^t > \left(\text{MDT}_i^t + \text{CSH}_i^t\right) \end{cases} (i \in N;\ \ t \in T) \tag{18}$$

where CSC_i^t and HSC_i^t are cold and hot start-up cost of ith unit at time t, respectively. CSH_i^t and MDT_i^t are cold start hour and minimum down time of ith unit at time t, respectively. $\text{MDT}_{i,\text{on}}^t$ is the continuously on time of ith unit up to time t. The constraints are associated with UCP that are stated below.

5.1.2 Power balance constraint

The total generated power must be equal to power demand in each hour. The constraint is stated below [43]:

$$\sum_{i=1}^{N} P_i^t u_i^t - P_D^t = 0, \quad t = 1, 2, \ldots, T \tag{19}$$

where P_D^t is the power demand at time t.

5.1.3 Spinning reserve constraint

The important feature of reliability is a condition of overload capacity of generation unit. The overload capacity of generation is known as spinning reserve constraint. The spinning reserve constraint is defined as follows:

$$\sum_{i=1}^{N} P_{i(\max)} u_i^t \ge P_D^t + P_R^t \tag{20}$$

where P_R^t is the spinning reserve at time t. $P_{i(\max)}$ is the maximum power limit of unit i.

5.1.4 Minimum up and down time constraints

Minimum up/down time constraints means that a generation unit should be on or off for a minimum number of hours before commitment (i.e., off to on state) or decommitment (i.e., on to off state), respectively. These constraints are defined as follow:

$$u_i^t = \begin{cases} \text{off} \rightarrow \text{on}, & \text{if } T_{i,\text{off}}^t \ge \text{MDT}_i \\ \text{on} \rightarrow \text{off}, & \text{if } T_{i,\text{on}}^t \ge \text{MUT}_i \end{cases} \tag{21}$$

where $T_{i,\text{on}}^t$ represents continuously on time of unit i up to time t, MUT_i is minimum up time of unit i, $T_{i,off}^t$ represents continuously off time of unit i up to time t and MDT_i is unit i minimum down time.

5.1.5 Ramping constraints

Due to the unit's mechanical characteristics, the output power of each unit is restricted by their corresponding ramp rate constraints. The ramp-up and ramp-down constraints are given below [43]:

$$P_i^t - P_i^{t-1} \le \text{RU}_i \quad \text{as generation increases} \tag{22}$$

$$P_i^{t-1} - P_i^t \le \text{RD}_i \quad \text{as generation decreases} \tag{23}$$

where RD_i and RU_i are the ramp-down and ramp-up limits of unit i, respectively.

5.1.6 Unit generation limit constraint

The power generation of each unit must be within their specified limits, which are given below:

$$u_i^t \cdot \Gamma_{i(min)}^t \le P_i^t \le u_i^t \cdot \Gamma_{i(max)}^t \tag{24}$$

where $\Gamma_{i(min)}^t$ and $\Gamma_{i(max)}^t$ are minimum and maximum generating power limit of unit i at time t respectively. These are computed as follows:

$$\Gamma_{i(min)}^t = \begin{cases} \max\left(P_{i(min)}, P_i^{t-1} - RD_i\right), & \text{if } u_i^{t-1} = u_i^t = 1 \text{ and } \psi = 1 \\ P_{i(min)}, & \text{if } u_i^{t-1} = 0 \text{ and } u_i^t = 1 \\ 0, & \text{if } u_i^{t-1} = 0 \text{ or } 1 \text{ and } u_i^t = 0 \end{cases} \tag{25}$$

$$\Gamma_{i(max)}^t = \begin{cases} \min\left(P_{i(max)}, P_i^{t-1} + RU_i\right), & \text{if } u_i^{t-1} = u_i^t = 1 \text{ and } \psi = 1 \\ P_{i(max)}, & \text{if } u_i^{t-1} = 0 \text{ and } u_i^t = 1 \\ 0, & \text{if } u_i^{t-1} = 0 \text{ or } 1 \text{ and } u_i^t = 0 \end{cases} \tag{26}$$

where ψ represents the indicator for ramp rate constraints. $P_{i(\min)}$ and $P_{i(\max)}$ are the minimum and maximum power limit of unit i respectively. The ramp rate constraints are to be considered when the value of ψ is 1; otherwise, these constraints are not to be considered.

5.2 Binary whale optimization algorithm for unit commitment

Figure 3 shows the flowchart of unit commitment problem using binary whale optimization algorithm. The basic steps of BWOA for unit commitment are mentioned below:

Step 1. Initialization of population

The status of each unit on/off is distinguished as a chromosome. The status of available unit at each hour constitutes a sub-chromosome. Therefore, an agent consists of T sub-chromosomes. An agent denotes unit commitment schedule over time horizon T. It is represented in terms of integer matrix U having dimensions $N \times T$

$$U = \begin{bmatrix} u_1^1 & u_1^2 & u_1^3 & \cdots & u_1^T \\ u_2^1 & u_2^2 & u_2^3 & \cdots & u_2^T \\ u_3^1 & u_3^2 & u_3^3 & \cdots & u_3^T \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ u_N^1 & u_N^2 & u_N^3 & \cdots & u_N^T \end{bmatrix} \quad (27)$$

where u_i^t is the on/off status of unit i at a time t.

A set of agents are formed randomly in the initialization process. For NP agents, the candidate solution of each agent $U_k(k = 1, 2, \ldots, NP)$ is randomly initialized. The element u_i^t of each agent U_k is generated either 0 or 1 using random function.

Step 2. Evaluation of priority list for unit commitment

The priority list is based on fuel cost obtained from the average fuel cost of each unit operating at its maximum output power. The average full load cost (fc) of a unit is defined as the cost per unit of power when power unit is at its full capacity. The fuel cost of unit i is obtained from Eq. 17, fc_i can be stated as:

$$\mathrm{fc}_i = \frac{\alpha_i}{P_{i(\max)}} + \beta_i + \gamma_i \cdot P_{i(\max)} \quad (28)$$

The priority list of power units will be formulated based on the order of fc. The power unit having lowest fc will be considered as highest priority.

Step 3. Modification of agents to assure spinning reserve constraints

The primary unit commitment may not satisfy the spinning reserve constraint, which is given by Eq. (20). The spinning reserve violations in primary unit commitment are revamped using following procedure.

Step 3.1 *Set* $t = 1$

Step 3.2 *Compute* fc_i *for all the uncommitted units* (u_i) *at hour t. Prepare a sorted list say,* $LU(\mathrm{fc}_i)$ *based on the value of fuel cost*

Step 3.3 *Compute excessive spinning reserve at each hour* (R^t)

$$R^t = \sum_{i=1}^{N} P_{i(\max)}\, u_i^t - P_D^t - P_R^t. \quad (29)$$

Step 3.4 *If* $R^t \geq 0$ *then goto Step 3.6.*

Step 3.5 *Generate a random number* $\zeta \in [0, 1]$. *If* $\zeta < P_R$, *then commit an uncommitted unit in* $LU(\mathrm{fc}_i)$ *that having lowest* fc_i *and return to Step 3.3; Otherwise, randomly commit an uncommitted unit in* $LU(\mathrm{fc}_i)$ *and return to Step 3.3*

Step 3.6 *If* $t < T$ *then* $t = t + 1$ *and return to Step 3.2. Otherwise, Stop.*

Step 4. Modification of agents to assure minimum up/down time constraints The unit schedule obtained from previous step, may be defy the minimum up and down time constraints. The on and off times of units should be determined for violation of up/down time constraints. The continuously on/off times of unit i up to hour t is computed as follows:

$$T_{i,\text{on}}^{t} = \begin{cases} T_{i,\text{on}}^{t-1} + 1 & \text{if } u_i^t = 1 \\ 0 & \text{if } u_i^t = 0 \end{cases}, \quad T_{i,\text{off}}^{t} = \begin{cases} T_{i,\text{off}}^{t-1} + 1 & \text{if } u_i^t = 0 \\ 0 & \text{if } u_i^t = 1 \end{cases} \tag{30}$$

The procedure for repair violations of minimum up and down times constraints is mentioned below.

Step 4.1 *Compute on and off times of all units for whole schedule time horizon using Eq.* (30)
Step 4.2 *Set* $t = 1$.
Step 4.3 *Set* $i = 1$.
Step 4.4 *If* $u_i^t = 0$ and $u_i^{t-1} = 1$ and $T_{i,\text{on}}^{t-1} < \text{MUT}_i$, *then set* $u_i^t = 1$.
Step 4.5 *If* $u_i^t = 0$ and $u_i^{t-1} = 1$ and $t + \text{MDT}_i - 1 \le T$ and $T_{i,\text{off}}^{t+\text{MDT}_i-1} < \text{MDT}_i$, *then set* $u_i^t = 1$.
Step 4.6 *If* $u_i^t = 0$ and $u_i^{t-1} = 1$ and $t + \text{MDT}_i - 1 > T$ and $\sum_{l=t}^{T} u_i^l > 0$, *then set* $u_i^t = 1$.
Step 4.7 *Update the on and off times for unit i using Eq.* (30).
Step 4.8 *If* $i < N$ *then* $i = i + 1$ *and goto Step 4.4.*
Step 4.9 *If* $t < T$ *then* $t = t + 1$ *and goto Step 4.3. Otherwise, Stop.*

*Step 5. Decommitment of excess units*The previous procedure for repairing minimum up and down time constraints may lead to excessive spinning reserves. The procedure for decommitment of excessive units is described as below

Step 5.1 *Set* $t = 1$.
Step 5.2 *Compute* fc_i *for each committed unit at hour t. Prepare a sorted list* $LU(\text{fc}_i)$ *based on* fc_i. *Suppose the first unit in* $LU(\text{fc}_i)$ *be* CU^t
Step 5.3 *Compute excessive spinning reserve at each hour*

$$R^t = \sum_{i=1}^{N} P_{i(\max)}\, u_i^t - P_D^t - P_R^t \tag{31}$$

Step 5.4 *If* $R^t \ge P_{\max}(CU^t)$ *then goto Step 5.6. Here* $P_{\max}$ *denotes maximum generating power.*
Step 5.5 *If decommitting* CU^t *does not violate their minimum up/down time constraint, then decommit* CU^t *and update on/off status of all units.*
Step 5.6 *Delete* CU^t *from* $LU(fc_i)$
Step 5.7 *If* $LU(fc_i)$ *is not empty. Let the first unit in* $LU(fc_i)$ *be* CU^t *and goto Step 5.3.*
Step 5.8 *If* $t < T$ *then* $t = t + 1$ *and goto Step 5.2. Otherwise, Stop.*

*Step 6. Evaluation of fitness function of all agents*The fitness function for each agent is computed using Eq. (16). The agent are sorted according to the fitness values of agents*Step 7. Update the position of agents*The position of each agent is updated using the BWOA which is mentioned in Sect. 2.2*Step 8. Termination Criterion*If maximum number of iteration is reached, then optimal UCP schedule and solutions are obtained. Otherwise, the iteration counter is incremented and go to Step 3

5.3 Performance evaluation

In order to demonstrate the effectiveness of BWOA for solving UCP, it is tested on three test systems such as small-scale, medium-scale and large-scale power systems. For these test systems, the spinning reserve requirement is considered to 10% of the load demand. The load demand and characteristics of generating power systems are taken from [12, 44] and described in Appendix 1. The load demand pattern of generating power systems is shown in Fig. 4. The population size and maximum number of iterations for BWOA are set to 50 and 500, respectively. The other parameters remain same as mentioned in Sect. 4.2. Due to stochastic nature of BWOA, 30 test trails are constructed for each test set. Each trail starts with different initial populations. The best, average, and worst costs are reported in tables. The optimality gap for small (10 units) and medium-scale power system (20 units) is also

mentioned as the optimal solution values for these instances are obtained thorough CPLEX software [45].

5.3.1 Test system 1: small-scale power system

For small-scale power system, two different cases are considered. The 4-unit and 10-unit test systems are taken into consideration for this power system.

Case-I: 4-unit test system

The first test system consists of 4-unit test system. The scheduling periods are 8 h for 4-unit test system. The load demand and characteristics of generating unit test are shown in Tables 26 and 27, respectively. Table 7 shows optimal commitment and generation schedule obtained from the proposed approach. The total operational and start-up costs obtained from proposed approach are \$74644.07 and \$0, respectively. The results were reported from literature when same problem was solved using improved Lagrangian relaxation (ILR) [15], A.SMP [44], B.SMP [44], Lagrangian relaxation and PSO (LRPSO) [41], binary differential evolution (BDE) [40], genetic algorithm (GA) [19], binary fireworks algorithm (BFWA) [42], binary grey wolf algorithm (BGWA) [13], and hybrid genetic imperialist competitive algorithm (HGICA) [25]. The performance comparison between BWOA and the above-mentioned metaheuristic techniques is depicted in Table 8. The experimental results reveal that the proposed BWOA outperforms the other techniques in terms of total cost. The superiority of BWOA is evident that it satisfies all the constraints except the ramp rate constraints that are not considered in this case. Among the existing techniques, HGICA is the second-best algorithm.

Case-II: 10-unit test system

The second test system consists of 10-generating unit systems. The scheduling periods are 24 h. The load demand data and characteristics of generating systems are shown in Tables 28 and 29 respectively. The optimal commitment and generation schedules obtained from BWOA are depicted in Tables 9 and 10 respectively. From these tables, it is observed that the total operational and start-up costs, obtained from BWOA, are \$559846.02 and \$4090, respectively. The performance of BWOA is compared with eleven well-known techniques and comparison results are reported in Table 11. These are local convergence-averse binary particle swarm optimization (LCA-PSO) [27], evolutionary programming (EP) [22], Lagrangian relaxation (LR) [15], enhanced simulated annealing (ESA) [20], binary PSO (BPSO) [21], improved binary PSO (IBPSO) [11], BDE [40], GA [19], BFWA [42], BGWA [13], and HGICA [25]. The results obtained from

Table 8 Comparison of results obtained for 4-unit system without ramp rate constraints

Method	Best cost (\$)	Average cost (\$)	Worst cost (\$)
ILR [15]	75,231	NA	NA
A.SMP [44]	74,812	74,877	75,166
B. SMP [44]	74,812	74,877	75,166
LRPSO [24]	74,808	NA	NA
BDE [40]	74,676	NA	NA
GA [19]	74,675	NA	NA
BFWA [42]	74,723	74,723	74,723
BGWA [13]	74,666	74,666	74,666
HGICA [25]	74,653	74,653	74,653
BWOA	**74,644.07**	**74,644.07**	**74,644.07**

Bold values indicate the best results

Table 7 Optimal commitment schedule and unit output power of 4-unit system using BWOA

Commitment schedule					Output power				Run cost (\$/h)	Start-up cost (\$/h)
Hours	PU1	PU2	PU3	PU4	PU1	PU2	PU3	PU4		
1	1	1	0	0	300	150	0	0	9145.36	0
2	1	1	1	0	300	205	25	0	10,892.24	0
3	1	1	1	1	300	250	30	20	12,570.54	0
4	1	1	1	0	300	215	25	0	11,079.38	0
5	1	0	1	1	300	0	80	0	8523.53	0
6	1	0	1	0	255	0	25	0	6103.14	0
7	1	0	1	0	265	0	25	0	6279.82	0
8	1	1	0	0	300	200	0	0	10,050.06	0
	Total operational cost = 74,644.07\$				Total start-up cost = 0\$				Total cost = 74,644.07\$	

Table 9 Optimal commitment schedule of 10-unit power system using BWOA

Hour	PU1	PU2	PU3	PU4	PU5	PU6	PU7	PU8	PU9	PU10
1	1	1	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0
3	1	1	0	0	1	0	0	0	0	0
4	1	1	0	0	1	0	0	0	0	0
5	1	1	0	1	1	0	0	0	0	0
6	1	1	1	1	1	0	0	0	0	0
7	1	1	1	1	1	0	0	0	0	0
8	1	1	1	1	1	0	0	0	0	0
9	1	1	1	1	1	1	1	0	0	0
10	1	1	1	1	1	1	1	1	0	0
11	1	1	1	1	1	1	1	1	1	0
12	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	0	0
14	1	1	1	1	1	1	1	0	0	0
15	1	1	1	1	1	0	0	0	0	0
16	1	1	1	1	1	0	0	0	0	0
17	1	1	1	1	1	0	0	0	0	0
18	1	1	1	1	1	1	0	0	0	0
19	1	1	1	1	1	1	0	0	0	0
20	1	1	1	1	1	1	1	1	0	0
21	1	1	1	1	1	1	1	0	0	0
22	1	1	1	0	1	0	1	0	0	0
23	1	1	0	0	0	1	0	0	0	0
24	1	1	0	0	0	0	0	0	0	0

Table 11 reveal that the BWOA gives comparable results with HGICA. The minimum generation cost obtained from BWOA and HGICA are $563,936.02 and $563,935.31, respectively. Table 11 also shows the optimality gap for this test instance as the optimal solution value is $563,938, which is obtained from CPLEX. Table 11 shows that the proposed approach is able to find optimal solution. The proposed approach performs better than other techniques in terms of generation cost.

5.3.2 Test system 2: medium-scale power system

Medium-scale power systems are tested for two different cases. The 20-unit and 40-unit test systems are taken into consideration for this power system. The scheduling periods are 24 h for these power systems. The ramp rate constraints are not considered for this test system. The proposed approach is compared with eleven well-established techniques such as local convergence-averse binary particle swarm optimization (LCA-PSO) [27], evolutionary programming (EP) [22], Lagrangian relaxation (LR) [15], enhanced simulated annealing (ESA) [20], binary differential evolution (BDE) [40], binary PSO (BPSO) [21], improved binary PSO (IBPSO) [11], genetic algorithm (GA) [19], binary fireworks algorithm (BFWA) [42], binary grey wolf algorithm (BGWA) [13], and hybrid genetic imperialist competitive algorithm (HGICA) [25].

Case-I: 20-unit test system

The first system consists of 20-unit test system. For 20-unit test system, the characteristics of 10-unit test system are replicated and load demand is taken as two times of 10-unit test system's load. Tables 12 and 13 show the optimal commitment and generation schedule obtained from BWOA. The operational and start-up costs obtained from BWOA are $1,115,120 and $8440, respectively.

The performance of proposed BWOA is compared with other techniques reported in the literature, which are reported in Table 14. In Table 14, the "NA" means the values are not available in the literature. It is observed from this table that the best cost obtained from BWOA is $1,123,560. The exploration and exploitation capability of BWOA provide quality solution for UCP. After the proposed BWOA, HGICA provides better results than the other techniques. The best cost obtained from HGICA is $1,124,565, which is $1005 higher than the proposed approach. According to the results reported in Table 14, HGICA is the second-best algorithm among the above-

Table 10 Optimal unit output power for 10-unit system using BWOA

Hour	PU1	PU2	PU3	PU4	PU5	PU6	PU7	PU8	PU9	PU10	Run cost	Start-up cost
1	455	245	0	0	0	0	0	0	0	0	13,683.13	0
2	455	295	0	0	0	0	0	0	0	0	14,554.50	0
3	455	370	0	0	25	0	0	0	0	0	16,809.45	900
4	455	455	0	0	40	0	0	0	0	0	18,597.67	0
5	455	390	0	130	25	0	0	0	0	0	20,020.02	560
6	455	360	130	130	25	0	0	0	0	0	22,387.04	1100
7	455	410	130	130	25	0	0	0	0	0	23,261.98	0
8	455	455	130	130	30	0	0	0	0	0	24,150.34	0
9	455	455	130	130	85	20	25	0	0	0	27,251.05	860
10	455	455	130	130	162	33	25	10	0	0	30,057.55	60
11	455	455	130	130	162	73	25	10	10	0	31,916.06	60
12	455	455	130	130	162	80	25	43	10	10	33,890.16	60
13	455	455	130	130	162	33	25	10	0	0	30,057.55	0
14	455	455	130	130	85	20	25	0	0	0	27,251.05	0
15	455	455	130	130	30	0	0	0	0	0	24,150.34	0
16	455	310	130	130	25	0	0	0	0	0	21,513.66	0
17	455	260	130	130	25	0	0	0	0	0	20,641.82	0
18	455	340	130	130	25	20	0	0	0	0	23,037.12	0
19	455	440	130	130	25	20	0	0	0	0	23,897.57	0
20	455	455	130	130	162	33	25	10	0	0	30,057.60	490
21	455	455	130	130	85	20	25	0	0	0	27,251.05	0
22	455	455	130	0	35	0	25	0	0	0	22,336.53	0
23	455	425	0	0	0	20	0	0	0	0	17,645.36	0
24	455	345	0	0	0	0	0	0	0	0	15,427.42	0
	Total operational cost = 559,846.02$						Total start-up cost = 4090$				Total cost = 563,936.02 $	

Table 11 Comparison of results obtained for 10-unit system without ramp rate constraints

Method	Best cost ($) (optimality gap)	Average cost ($) (optimality gap)	Worst cost ($) (optimality gap)
LCA-PSO [27]	570,006 (1.0760)	NA	NA
EP [22]	564,551 (0.1087)	565,352 (0.2507)	566,231 (− 0.4066)
LR [15]	566,107 (0.3846)	566,493 (0.4530)	566,817 (− 0.5105)
ESA [20]	565,828 (0.3351)	565,988 (0.3635)	566,260 (− 0.4117)
BDE [40]	563,997 (0.0104)	563,997 (0.0104)	563,997 (− 0.0104)
BPSO [21]	563,977 (0.0069)	563,977 (0.0069)	563,977 (− 0.0069)
IBPSO [11]	563,977 (0.0069)	564,155 (0.0384)	565,312 (− 0.2436)
GA [19]	563,977 (0.0069)	564,275 (0.0597)	566,606 (− 0.4731)
BFWA [42]	563,977 (0.0069)	564,018 (0.0141)	564,855 (− 0.1626)
BGWA [13]	563,976 (0.0067)	564,378 (0.0780)	565,518 (− 0.2802)
HGICA [25]	563,935.31 (− 0.0004)	563,937 (− 0.0002)	563,938 (0.0000)
BWOA	563,936.02 (− 0.0003)	563,938.89 (0.0001)	563,939.01 (− 0.0001)

Table 12 Optimal commitment schedule of 20-unit power system using BWOA

	PU1	PU2	PU3	PU4	PU5	PU6	PU7	PU8	PU9	PU10	PU11	PU12	PU13	PU14	PU15	PU16	PU17	PU18	PU19	PU20
1	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
3	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
4	1	1	0	0	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0
5	1	1	1	0	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0
6	1	1	1	1	1	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0
7	1	1	1	1	1	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0
8	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0
9	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	1	0	0	0
10	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0	1	0
14	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	0
15	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0
16	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0
17	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0
18	1	1	1	1	1	0	1	0	0	0	1	1	1	1	1	1	0	0	0	0
19	1	1	1	1	1	0	1	0	0	0	1	1	1	1	1	1	0	0	0	0
20	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0
21	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	1	0	0	0
22	1	1	1	0	1	1	0	0	0	0	1	1	0	0	1	0	1	0	0	0
23	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
24	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0

Table 13 Unit output power for 20-unit system using BWOA

	PU1	PU2	PU3	PU4	PU5	PU6	PU7	PU8	PU9	PU10	PU11	PU12	PU13	PU14	PU15	PU16	PU17	PU18	PU19	PU20	Run Cost	Start-up Cost
1	455	245	0	0	0	0	0	0	0	0	455	245	0	0	0	0	0	0	0	0	28,653.69	0
2	455	295	0	0	0	0	0	0	0	0	455	295	0	0	0	0	0	0	0	0	32,528.58	0
3	455	382	0	0	25	0	0	0	0	0	455	382	0	0	0	0	0	0	0	0	37,322.27	900
4	455	455	0	0	40	0	0	0	0	0	455	455	0	0	40	0	0	0	0	0	41,733.66	900
5	455	455	130	0	25	0	0	0	0	0	455	455	0	0	25	0	0	0	0	0	41,492.05	560
6	455	425	130	130	25	0	0	0	0	0	455	425	130	0	25	0	0	0	0	0	44,242.54	2220
7	455	455	130	130	45	0	0	0	0	0	455	455	130	0	45	0	0	0	0	0	47,235.77	0
8	455	455	130	130	30	0	0	0	0	0	455	455	130	130	30	0	0	0	0	0	50,227.23	1100
9	455	455	130	130	95	20	25	0	0	0	455	455	130	130	95	0	25	0	0	0	52,963.06	1200
10	455	455	130	130	162	33	25	10	10	0	455	455	130	130	162	33	25	0	0	0	58,143.56	640
11	455	455	130	130	162	73	25	10	10	10	455	455	130	130	162	73	25	10	0	0	59,835.27	120
12	455	455	130	130	162	80	25	43	10	10	455	455	130	130	162	80	25	43	10	10	61,780.33	120
13	455	455	130	130	162	33	25	10	0	0	455	455	130	130	162	33	25	0	10	0	58,161.73	0
14	455	455	130	130	98	20	25	0	0	0	455	455	130	130	98	20	0	0	0	0	52,417.18	0
15	455	455	130	130	30	0	0	0	0	0	455	455	130	130	30	0	0	0	0	0	47,525.06	0
16	455	310	130	130	25	0	0	0	0	0	455	310	130	130	25	0	0	0	0	0	41,182.77	0
17	455	260	130	130	25	0	0	0	0	0	455	260	130	130	25	0	0	0	0	0	49,532.78	0
18	455	338	130	130	25	0	25	0	0	0	455	338	130	130	25	20	0	0	0	0	42,839.79	0
19	455	438	130	130	25	0	25	0	0	0	455	438	130	130	25	20	0	0	0	0	48,711.38	0
20	455	455	130	130	162	33	25	10	10	0	455	455	130	130	162	33	25	0	0	0	56,712.77	680
21	455	455	130	130	95	20	25	0	0	0	455	455	130	130	95	0	25	0	0	0	52,091.11	0
22	455	455	130	0	103	20	0	0	0	0	455	455	0	0	103	0	25	0	0	0	43,718.41	0
23	455	433	0	0	25	0	0	0	0	0	455	433	0	0	0	0	0	0	0	0	33,447.13	0
24	455	345	0	0	0	0	0	0	0	0	455	345	0	0	0	0	0	0	0	0	32,621.89	0
Total operational cost = $1,115,120										Total start-up cost = $8440										Total cost = 1,123,560$		

Table 14 Comparison of results obtained for 20-unit system without ramp rate constraints

Method	Best cost ($) (optimality gap)	Average cost ($) (optimality gap)	Worst cost ($) (optimality gap)
LCA-PSO [27]	1,139,005 (1.3984)	NA	NA
EP [22]	1,126,494 (0.2846)	1,127,257 (0.3525)	1,129,793 (0.5782)
LR [15]	1,128,362 (0.4509)	1,128,395 (0.4538)	1,128,444 (0.4582)
ESA [20]	1,126,254 (0.2632)	1,127,955 (0.4147)	1,129,112 (0.5176)
BDE [40]	1,126,998 (0.3295)	1,127,374 (0.3629)	1,127,927 (0.4122)
BPSO [21]	1,128,192 (0.4358)	1,128,213 (0.4376)	1,128,403 (0.4546)
IBPSO [11]	1,196,029 (6.4748)	NA	NA
GA [19]	1,126,243 (0.2622)	1,128,790 (0.4890)	1,132,059 (0.7800)
BFWA [42]	1,124,658 (0.1211)	1,124,941 (0.1464)	1,125,087 (0.1594)
BGWA [13]	1,125,546 (0.2002)	1,126,126 (0.2518)	1,127,393 (0.3646)
HGICA [25]	1,124,565 (0.1129)	1,124,933 (0.1456)	1,125,147 (0.1647)
BWOA	**1,123,560 (0.0234)**	**1,123,978 (0.0600)**	**1,124,186 (0.0791)**

Bold values indicate the best results

mentioned techniques. The solutions (i.e., best, average, and worst cost) obtained from BFWA are comparable with HGICA. The best cost obtained from BFWA is $1,124,658, which is slightly higher than HGICA. The optimality gap is also reported in Table 14. The optimal solution value obtained from CPLEX is $1,123,297. The solution obtained from the proposed approach is within 0.07% of optimality. The results reveal that the BWOA provides better cost as compared with other reported techniques.

Case-II: 40-unit test system

The second test system consists of 40-unit system. For 40-unit test system, the characteristics of 10-unit test system were replicated twice and load demand was multiplied by 4. The optimal commitment and generation schedules over 24-hour time interval are mentioned in Tables 15 and 16 respectively. The operational and start-up costs obtained from proposed BWOA are $2,205,546 and $18380, respectively.

Table 17 shows the performance comparison of proposed BWOA with other well-established techniques reported in the literature. The "NA" mentioned in this table means that the values are not available in the literature. It is observed from table that the best cost obtained from BWOA is $2,223,926, which is far better than the other techniques. The average and worst cost obtained from BWOA are $2,228,753 and $2,231,547, respectively. These values are much better than the best value of other existing techniques. Due to better trade-off between exploration and exploitation, the BWOA provide better quality solution for unit commitment problem. The best cost obtained with HGICA algorithm for 40-unit system is $2,239,186, which is $15260 more than that of proposed approach. It can be seen from Table 17 that HGICA is the second-best algorithm among the above-mentioned techniques. The solutions obtained from BPSO are comparable with IBPSO. The best cost obtained from BPSO is $2,243,210, which is $4024 more than that of HGICA. The results reveal that the BWOA provides better cost as compared with other reported techniques.

5.3.3 Test system 3: large-scale power system

Large-scale power systems are tested for two different cases. The 80-unit and 100-unit test systems are taken into consideration for this power system. The scheduling periods are 24 h for these power systems. The ramp rate constraints are not considered for this test system. The proposed approach is compared with well-known techniques such as local convergence-averse binary particle swarm optimization (LCA-PSO) [27], evolutionary programming (EP) [22], Lagrangian relaxation (LR) [15], enhanced simulated annealing (ESA) [20], binary differential evolution (BDE) [40], binary PSO (BPSO) [21], improved binary PSO (IBPSO) [11], genetic algorithm (GA) [19], binary fireworks algorithm (BFWA) [42], binary grey wolf algorithm (BGWA) [13], and hybrid genetic imperialist competitive algorithm (HGICA) [25].

Case-I and II: 80 and 100-unit test system

For 80- and 100-unit test systems, the proposed BWOA is compared with other techniques which are reported in Table 18. The "NA" mentioned in this table indicates that the values are not available in the literature. For 80-unit system, the best cost obtained from BWOA is $4,478,412. The best cost obtained from HGICA is $4,485,936, which is $7524 more than that of proposed approach. HGICA is the second-best algorithm among the above-mentioned techniques. The solutions obtained from IBPSO are comparable with BDE. For 100-unit system, the best cost

Table 15 Commitment schedule of 40-unit power system using BWOA

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
PU1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PU2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PU3	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1
PU4	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1
PU5	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
PU6	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	0	0	0	0
PU7	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	0	0	0	0
PU8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
PU9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
PU10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PU11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PU12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PU13	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
PU14	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
PU15	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
PU16	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1	0	0	0
PU17	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0	0
PU18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PU19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
PU20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
PU21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PU22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PU23	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1
PU24	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
PU25	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
PU26	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1	0	0	0
PU27	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1	0	0	0	0
PU28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
PU29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
PU30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PU31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
PU32	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
PU33	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1
PU34	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	0
PU35	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
PU36	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	0	0	0
PU37	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1	0	0	0	0
PU38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PU39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PU40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

obtained from BWOA is \$5,599,281. The best cost obtained from HGICA is \$5,604,022, which is \$4741 more than that of proposed approach. The results reveal that the BWOA provides better cost as compared with other reported techniques for both 80- and 100-unit systems. It is also observed that HGICA is the second-best algorithm for these test systems.

5.4 Effect of ramp rate constraints

To devise the power system more realistic, the ramp rate constraints are integrated in the formulation of UCP. By including ramp rate constraints, the minimum and maximum generating power of each unit can be computed thorough Eqs. (25) and (26), respectively. According to

Table 16 Unit output power of 40-unit power system using BWOA

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
PU1	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455
PU2	326	393	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	376	438
PU3	0	0	0	130	130	130	130	130	130	130	130	130	130	130	130	0	0	0	0	0	130	130	130	130
PU4	0	0	0	0	130	130	130	130	130	130	130	130	130	130	130	0	0	0	0	0	130	130	130	130
PU5	0	0	108	121	139	141	126	144	130	162	162	162	162	158	128	140	90	162	162	162	162	0	0	0
PU6	0	0	0	0	0	0	0	0	0	0	80	80	62	0	0	0	0	37	71	80	0	0	0	0
PU7	0	0	0	0	0	0	0	0	0	68	58	61	25	0	0	0	0	25	25	77	0	0	0	0
PU8	0	0	0	0	0	0	0	0	0	0	0	55	0	0	0	0	0	0	0	55	0	0	0	0
PU9	0	0	0	0	0	0	0	0	0	0	0	55	0	0	0	0	0	0	0	55	0	0	0	0
PU10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PU11	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455
PU12	326	393	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	376	438
PU13	0	0	0	0	0	130	130	130	130	130	130	130	130	130	130	0	0	0	0	0	0	0	0	0
PU14	0	0	0	0	0	130	130	130	130	130	130	130	130	130	130	0	0	0	0	0	0	0	0	0
PU15	0	0	108	121	139	141	126	144	130	162	162	162	162	158	128	140	90	162	162	162	162	0	0	0
PU16	0	0	0	0	0	0	0	0	0	0	80	80	62	0	0	0	0	0	71	80	79	0	0	0
PU17	0	0	0	0	0	0	0	0	0	68	58	61	0	0	0	0	0	0	25	77	25	0	0	0
PU18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PU19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	55	0	0	0	0
PU20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0
PU21	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455
PU22	326	393	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	376	438
PU23	0	0	0	0	0	130	130	130	130	130	130	130	130	130	0	0	0	0	0	130	130	130	130	130
PU24	0	0	0	0	0	0	130	130	130	130	130	130	130	130	130	0	0	0	0	0	0	0	0	0
PU25	0	0	0	121	139	141	126	144	130	162	162	162	162	158	128	140	90	162	162	162	162	0	0	0
PU26	0	0	0	0	0	0	0	0	0	0	80	80	62	0	0	0	0	0	71	80	79	0	0	0
PU27	0	0	0	0	0	0	0	0	0	68	58	61	0	0	0	0	0	25	25	77	0	0	0	0
PU28, PU29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	55	0	0	0	0
PU30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PU31	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	455	0
PU32	0	0	0	0	0	0	0	0	455	455	455	455	455	455	455	455	455	455	455	455	455	455	0	0
PU33	0	0	0	0	0	0	130	130	130	130	130	130	130	130	0	0	0	0	0	130	130	130	130	130
PU34	0	0	0	0	0	0	0	130	130	130	130	130	130	0	0	0	0	0	130	130	130	130	130	0
PU35	0	0	0	121	139	141	126	144	130	162	162	162	162	158	128	140	90	162	162	162	162	110	0	0
PU36	0	0	0	0	0	0	0	0	0	0	0	80	62	20	0	0	0	0	71	80	79	0	0	0

Table 16 (continued)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
PU37		0	0	0	0	0	0	0	0	68	58	61	0	0	0	0	0	25	25	77	0	0	0	0
PU38		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PU39, PU40		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total operational cost =2,205,546 $										Total start-up cost = 18,380$										Total cost = 2,223,926$				

literature, the values of RD_i and RU_i are same and is to be taken as the 20% of maximum capacity of unit at each hour [20, 43]. Hence, the values of RD_i and RU_i for units in 10-unit system are set to 91, 91, 26, 26, 26, 33, 17, 16, 11, 11, and 11, respectively. The output power of generating unit is limited due to its ramp rate constraints. Therefore, the generation cost will be increases as compared to the cost without ramp rate constraints. The effectiveness of proposed BWOA is tested on 10, 20, 40, 80, and 100 units having ramp rate constraints. Table 19 shows the effect of ramp rate constraints on the performance of BWOA. The best cost obtained from BWOA is \$565,771, which is \$1835 more than the cost without ramp constraints. It can be inferred from table that the BWOA is able to handle the ramp rate constraints.

5.5 Complex power systems

In order to demonstrate the robustness of BWOA, it is tested on two complex power systems. These are IEEE 118-bus system and a real Taiwan Power 38-unit system.

5.5.1 IEEE 118-bus system

IEEE 118-bus system consists of 54 generating units. The generator data and load profile are taken from [46]. The comparison results of proposed technique with other existing techniques in terms of total operating cost are given in Table 20. It is inferred from table that the total cost obtained from BWOA is \$1,643,089, which is far better than the other techniques. Due to better trade-off between exploration and exploitation, the BWOA provide better quality solution for IEEE 118-bus system problem. The cost obtained with HGICA algorithm for this system is \$1,644,115, which is \$1026 more than that of proposed approach. It can be concluded that HGICA is the second-best algorithm among the other techniques. The cost obtained from BOWA is \$1,644,128 which is \$13 more than that of HGICA. The results reveal that the BWOA provides better cost as compared with other reported techniques.

5.5.2 Taiwan power system

The Taiwan Power 38-unit system is taken from [47]. The same conditions have been considered to execute the UCP, which are mentioned in [47]. Table 21 shows the comparison between proposed method and other existing techniques. It is observed from Table 21 that the proposed BWOA provides best solution as compared to existing techniques. The best cost obtained from BWOA is \$195,904,123. The best cost obtained from HGICA is \$195,912,018, which is \$7895 more than that of proposed

approach. HGICA is the second-best algorithm among the above-mentioned techniques. It can be concluded that the BWOA is able to solve Taiwan power system more efficiently than the other existing techniques.

Table 17 Comparison for results obtained from 40-unit system without ramp rate constraints

Method	Best cost ($)	Average cost ($)	Worst cost ($)
LCA-PSO [27]	2,277,396	NA	NA
EP [22]	2,249,093	2,252,612	2,256,085
LR [15]	2,258,503	2,258,503	2,258,503
ESA [20]	2,250,012	2,252,125	2,254,539
BDE [40]	2,245,700	2,246,600	2,247,284
BPSO [21]	2,243,210	2,244,634	2,245,982
IBPSO [11]	2,243,728	NA	NA
GA [19]	2,252,909	2,262,585	2,269,282
BFWA [42]	2,248,228	2,248,572	2,248,645
BGWA [13]	2,252,475	2,257,866	2,263,333
HGICA [25]	2,239,186	2,242,612	2,246,085
BWOA	**2,223,926**	**2,228,753**	**2,231,547**

Bold values indicate the best results

6 Conclusions

In this paper, a novel binary version of the whale optimization algorithm namely BWOA has been proposed, which utilizes the concept of transfer function. The proposed approach has been tested on a set of various benchmark test functions. It has been compared with six well-known metaheuristic techniques. The statistical test has also been carried out to prove the statistical significance of proposed approach. The experimental results reveal that BWOA outperforms the existing metaheuristic algorithms. Furthermore, BWOA has been applied on unit commitment problem. The proposed approach has been tested on 4, 10, 20, 40, 80, and 100 power units. It has also been tested on IEEE 118-bus system and Taiwan 38 bus system. The simulation results show that the proposed method is effective for solving UCP. The total production costs over schedule time obtained from proposed BWOA are less expensive than other existing methods.

Table 18 Comparison of results obtained for 80- and 100-unit systems without ramp rate constraints

Method	80-unit system			100-unit system		
	Best cost ($)	Average cost ($)	Worst cost ($)	Best cost ($)	Average cost ($)	Worst cost ($)
LCA-PSO [27]	4,554,346	NA	NA	5,706,201	NA	NA
EP [22]	4,498,479	4,505,536	4,512,739	5,623,885	5,633,800	5,639,148
LR [15]	4,526,022	4,526,022	4,526,022	5,657,277	5,657,277	5,657,277
ESA [20]	4,498,076	4,501,156	4,503,987	5,617,876	5,624,301	5,628,506
BDE [40]	4,489,022	4,490,456	4,491,262	5,609,341	5,609,984	5,610,608
BPSO [21]	4,491,083	4,491,681	4,492,686	5,610,293	5,611,181	5,612,265
IBPSO [11]	4,488,351	NA	NA	5,608,792	NA	NA
GA [19]	4,507,692	4,525,204	4,552,982	5,626,361	5,669,362	5,690,086
BFWA [42]	4,491,284	4,492,550	4,493,036	5,610,954	5,612,422	5,612,790
BGWA [13]	4,495,173	4,506,362	4,513,026	5,628,975	5,637,659	5,643,899
HGICA [25]	4,485,936	4,487,958	4,489,283	5,604,022	5,608,561	5,613,260
BWOA	**4,478,412**	**4,479,584**	**4,480,924**	**5,599,281**	**5,601,975**	**5,603,122**

Bold values indicate the best results

Table 19 Performance evaluation of proposed approach with ramp rate constraints

Number of power units	Cost ($)		
	Best cost ($)	Average cost ($)	Worst cost ($)
10	565,771	565,873	565,978
20	1,126,625	1,127,031	1,127,640
40	2,235,113	2,236,342	2,236,937
80	4,490,172	4,491,236	4,491,882
100	5,614,038	5,616,185	5,619,514

Table 20 Comparison of results obtained from IEEE 118-bus system

Method	Operating cost ($)
Semi-definite programming (SDP) [46]	1,645,445.00
Artificial bee colony Lagrangian relaxation (ABC-LR) [48]	1,644,269.00
Binary real-coded firefly algorithm (BRCFF) [49]	1,644,141.00
Firefly algorithm with multiple workers (FAMW) [50]	1,644,134.00
Binary grey wolf algorithm (BGWA) [13]	1,644,128.00
Hybrid genetic imperialist competitive algorithm HGICA [25]	1,644,115.00
Binary whale optimization (BWOA)	**1,643,089.00**

Bold values indicate the best results

Table 21 Comparison of results obtained from Taiwan Power 38-unit system

Method	Minimum operating cost (million $)
Dynamic programming (DP) [47]	210.500000
Lagrangian relaxation (LR) [47]	209.000000
Simulated annealing (SA) [47]	207.800000
Real-coded genetic algorithm (RGA) [51]	204.600000
Fuzzy adaptive particle swarm optimization (FAPSO) [52]	196.730000
Binary real-coded firefly algorithm (BRCFF) [49]	195.916304
Firefly algorithm with multiple workers (FAMW) [50]	195.914900
Binary grey wolf algorithm (BGWA) [13]	195.914495
Hybrid genetic imperialist competitive algorithm HGICA [25]	195.912018
Binary whale optimization (BWOA)	195.904123

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix 1

See Fig. 4 and Tables 22, 23, 24, 25, 26, 27, 28 and 29.

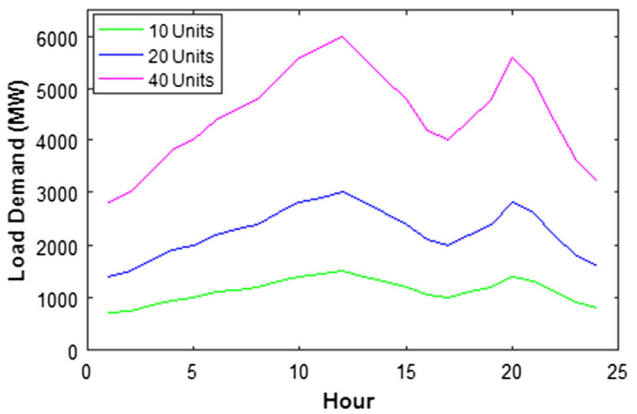


Fig. 4 Load demand pattern for power 10, 20, and 40-unit test systems

Table 22 Unimodal benchmark test functions

Benchmark test function	D	Range	$F_{\min}$
$F_1(Y) = \sum_{i=1}^{D} y_i^2$	5	$[-100, 100]$	0
$F_2(Y) = \sum_{i=1}^{D} \lvert y_i \rvert + \prod_{i=1}^{D} \lvert y_i \rvert$	5	$[-10, 10]$	0
$F_3(Y) = \sum_{i=1}^{D} \left(\sum_{j=1}^{i} y_j \right)^2$	5	$[-100, 100]$	0
$F_4(Y) = \max_i \{ \lvert y_i \rvert, \ 1 \le i \le D \}$	5	$[-100, 100]$	0
$F_5(Y) = \sum_{i=1}^{D-1} \left[100 \left(y_{i+1} - y_i^2 \right)^2 + (y_i - 1)^2 \right]$	5	$[-30, 30]$	0
$F_6(Y) = \sum_{i=1}^{D} (\lfloor y_i + 0.5 \rfloor)^2$	5	$[-100, 100]$	0
$F_7(Y) = \sum_{i=1}^{D} i y_i^4 + random[0, 1)$	5	$[-1.28, 1.28]$	0

Table 23 Multimodal benchmark test functions

Objective function	D	Range	$F_{\min}$
$F_8(Y) = -\frac{1}{D}\sum_{i=1}^{D}\left(y_i \sin\left(\sqrt{\lvert y_i\rvert}\right)\right)$	5	$[-500, 500]$	-2094.914
$F_9(Y) = \sum_{i=1}^{D}\left[y_i^2 - 10\cos(2\pi y_i) + 10\right]$	5	$[-5.12, 5.12]$	0
$F_{10}(Y) = -20\exp\left(-0.2\sqrt{\frac{1}{D}\sum_{i=1}^{D} y_i^2}\right) - \exp\left(\frac{1}{D}\sum_{i=1}^{D}\cos(2\pi y_i)\right) + 20 + e$	5	$[-32, 32]$	0
$F_{11}(Y) = \frac{1}{4000}\sum_{i=1}^{D} y_i^2 - \prod_{i=1}^{D}\cos\left(\frac{y_i}{\sqrt{i}}\right) + 1$	5	$[-600, 600]$	0
$F_{12}(Y) = \frac{\pi}{D}\left\{10\sin(\pi z_1) + \sum_{i=1}^{D-1}(z_i - 1)^2\left[1 + 10\sin^2(\pi z_{i+1})\right] + (z_D - 1)^2\right\} + \sum_{i=1}^{D} u(y_i, 10, 100, 4)$ *where* $u(y_i, a, k, m) = \begin{cases} k(y_i - a)^m, & y_i > a, \\ 0, & -a \le y_i \le a, \\ k(-y_i - a)^m, & y_i < -a \end{cases}$ $z_i = 1 + \frac{1}{4}(y_i + 1)$	5	$[-50, 50]$	0
$F_{13}(Y) = 0.1\left\{\sin^2(3\pi y_1) + \sum_{i=1}^{D-1}(y_i - 1)^2\left[1 + \sin^2(3\pi y_{i+1})\right] + (y_D - 1)^2\left[1 + \sin^2(2\pi y_D)\right]\right\} + \sum_{i=1}^{D} u(y_i, 5, 100, 4)$	5	$[-30, 30]$	0

Table 24 Multimodal benchmark test functions with fixed dimension

Objective function	D	Range	F_{min}
$F_{14}(Y) = \left(\frac{1}{500} + \sum_{j=1}^{25}\frac{1}{j + \sum_{i=1}^{2}\left(x_i - a_{ij}\right)^6}\right)^{-1}$	2	$[-65, 65]$	1
$F_{15}(Y) = \sum_{i=1}^{11}\left[a_i - \frac{y_1\left(b_i^2 + b_i y_2\right)}{b_i^2 + b_i y_3 + y_4}\right]^2$	4	$[-5, 5]$	0.00030
$F_{16}(Y) = 4y_1^2 - 2.1y_1^4 + \frac{1}{3}y_1^6 + y_1 y_2 - 4y_2^2 + 4y_2^4$	2	$[-5, 5]$	-1.0316
$F_{17}(Y) = \left(y_2 - \frac{5.1}{4\pi^2}y_1^2 + \frac{5}{\pi}y_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos y_1 + 10$	2	$[-5, 5]$	0.398

Table 24 (continued)

Objective function	D	Range	F_{min}
$F_{18}(Y) = \left[1 + (y_1 + y_2 + 1)^2\left(19 - 14y_1 + 6y_1y_2 + 3y_2^2\right)\right] \times \left[30 + (2y_1 - 3y_2)^2 \times \left(18 - 32y_1 + 12y_1^2 + 48y_2 - 36y_1y_2 + 27y_2^2\right)\right]$	2	$[-2, 2]$	3
$F_{19}(Y) = -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{3} a_{ij}\left(x_j - p_{ij}\right)^2\right)$	3	[1, 3]	− 3.86
$F_{20}(Y) = -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{6} a_{ij}\left(x_j - p_{ij}\right)^2\right)$	6	[0, 1]	− 3.32
$F_{21}(Y) = -\sum_{i=1}^{5} \left[(Y - a_i)(Y - a_i)^T + c_i\right]^{-1}$	4	[0, 10]	− 10.1532
$F_{22}(Y) = -\sum_{i=1}^{7} \left[(Y - a_i)(Y - a_i)^T + c_i\right]^{-1}$	4	[0, 10]	− 10.4028
$F_{23}(Y) = -\sum_{i=1}^{10} \left[(Y - a_i)(Y - a_i)^T + c_i\right]^{-1}$	4	[0, 10]	− 10.5363

Table 25 Composite benchmark test functions

Objective function	D	Range	$F_{\min}$
$F_{24}(CF1)$: $f_1, f_2, f_3, \ldots, f_{10}$ = Sphere Function $[\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_{10}] = [1, 1, 1, \ldots, 1]$ $[\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{10}] = [5/100, 5/100, 5/100, \ldots, 5/100]$	10	$[-5, 5]$	0
$F_{25}(CF2)$: $f_1, f_2, f_3, \ldots, f_{10}$ = Griewank's function $[\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_{10}] = [1, 1, 1, \ldots, 1]$ $[\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{10}] = [5/100, 5/100, 5/100, \ldots, 5/100]$	10	$[-5, 5]$	0
$F_{26}(CF3)$: $f_1, f_2, f_3, \ldots, f_{10}$ = Griewank's Function $[\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_{10}] = [1, 1, 1, \ldots, 1]$ $[\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{10}] = [1, 1, 1, \ldots, 1]$	10	$[-5, 5]$	0

Table 25 (continued)

Objective function	D	Range	$F_{\min}$
$F_{27}(CF4)$: f_1, f_2 = Ackley's Function, f_3, f_4 = Rastrigin's Function f_5, f_6 = Weierstrass Function, f_7, f_8 = Griewank's Function f_9, f_{10} = Sphere Function $[\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_{10}] = [1, 1, 1, \ldots, 1]$ $[\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{10}] = [5/32, 5/32, 1, 1, 5/0.5, 5/0.5, 5/100, 5/100, 5/100, 5/100]$	10	$[-5, 5]$	0
$F_{28}(CF5)$: f_1, f_2 = Rastrigin's Function, f_3, f_4 = Weierstrass Function f_5, f_6 = Griewank's Function, f_7, f_8 = Ackley's Function f_9, f_{10} = Sphere Function $[\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_{10}] = [1, 1, 1, \ldots, 1]$ $[\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{10}] = [1/5, 1/5, 5/0.5, 5/0.5, 5/100, 5/100, 5/32, 5/32, 5/100, 5/100]$	10	$[-5, 5]$	0
$F_{29}(CF6)$: f_1, f_2 = Rastrigin's Function, f_3, f_4 = Weierstrass Function f_5, f_6 = Griewank's Function, f_7, f_8 = Ackley's Function f_9, f_{10} = Sphere Function $[\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_{10}] = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$ $[\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{10}] = \begin{bmatrix} 0.1 \times 1/5, 0.2 \times 1/5, 0.3 \times 5/0.5, 0.4 \times 5/0.5, 0.5 \times 5/100, \\ 0.6 \times 5/100, 0.7 \times 5/32, 0.8 \times 5/32, 0.9 \times 5/100, 1 \times 5/100 \end{bmatrix}$			

Table 26 Load demand for 4-power unit system (MW)

Hour	1	2	3	4	5	6	7	8
Demand	450	530	600	540	400	280	290	500

Table 27 Test data for 4-power unit system

Unit	P_i^{Max} (MW)	P_i^{Min} (MW)	a (\$/MW^2)	b (\$/MW^2)	c (\$/MW^2)	MUT (h)	MDT (h)	HSC (\$)	CSC (\$)	CSH (h)	IS (h)
PU1	300	75	684.74	16.83	0.0021	5	4	500	1100	5	8
PU2	250	60	585.62	16.95	0.0042	5	3	170	400	5	8
PU3	80	25	213	20.74	0.0018	4	2	150	350	4	− 5
PU4	60	20	252	23.6	0.0034	1	1	0	0.02	0	− 6

Table 28 Load demand for 10-power unit system (MW)

Hour	1	2	3	4	5	6	7	8	9	10	11	12
Demand	700	750	850	950	1000	1100	1150	1200	1300	1400	1450	1500
Hour	13	14	15	16	17	18	19	20	21	22	23	24
Demand	1400	1300	1200	1050	1000	1100	1200	1400	1300	1100	900	800

Table 29 Test data for 10-power unit system

Unit	P_i^{Max} (MW)	P_i^{Min} (MW)	a (\$/MW^2)	b (\$/MW^2)	c (\$/MW^2)	MUT (h)	MDT (h)	HSC (\$)	CSC (\$)	CSH (h)	IS (h)
PU1	455	150	1000	16.19	0.00048	8	8	4500	9000	5	8
PU2	455	150	970	17.26	0.00031	8	8	5000	10000	5	8
PU3	130	20	700	16.60	0.0002	5	5	550	1100	4	− 5
PU4	130	20	680	16.50	0.00211	5	5	560	1120	4	− 5
PU5	162	25	450	19.70	0.00398	6	6	900	1800	4	− 6
PU6	80	20	370	22.26	0.00712	3	3	260	520	2	− 3
PU7	85	25	480	27.74	0.00079	3	3	260	520	2	− 3
PU8	55	10	660	25.92	0.00413	1	1	30	60	0	− 1
PU9	55	10	665	27.27	0.00222	1	1	30	60	0	− 1
PU10	55	10	670	27.29	0.00173	1	1	30	60	0	− 1

References

1. Goldberg D (1989) Genetic algorithms in search, optimization, and machine. Addison-Wesley, Boston
2. Kennedy N, Eberhart RC (1995) Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks, Perth, pp 1942–1948
3. Dorigo M, Birattari M, Stutzle T (2006) Ant colony optimization. IEEE Comput Intell Mag 1(4):28–35
4. Yang X-S, Gandomi AH (2012) Bat algorithm: a novel approach for global engineering optimization. Eng Comput 29(5):464–483
5. Rashedi E, Nezamabadi-pour H, Saryazdi S (2009) GSA: a gravitational search algorithm. Inf Sci 179:2232–2248
6. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. IEEE Trans Evolut Comput 1:67–82
7. Mirjalili S, Lewis A (2016) The whale optimization algorithm. Adv Eng Softw 95:51–67
8. Wang X, Yang J, Teng X, Xia W, Jensen R (2007) Feature selection based on rough sets and particle swarm optimization. Pattern Recognit Lett 28:459–471
9. Srinivasa KG, Venugopal KR, Patnaik LM (2007) A self-adaptive migration model genetic algorithm for data mining applications. Inf Sci 177(20):4295–4313
10. Wu T-H, Chang C-C, Chung S-H (2008) A simulated annealing algorithm for manufacturing cell formation problems. Expert Syst Appl 34(3):1609–1617
11. Yuan X, Nie H, Su A, Wang L, Yuan Y (2009) An improved binary particle swarm optimization for unit commitment problem. Expert Syst Appl 36(4):8049–8055
12. Amiri M, Khanohamadi S (2013) A primary unit commitment approach with a modification process. Appl Soft Comput 13:1007–1015
13. Panwar LK, Reddy SK, Verma A, Panigrahi BK, Kumar R (2018) Binary grey wolf optimizer for large scale unit commitment problem. Swarm Evol Comput 38:251–266
14. Senjyu T, Shimabukuro K, Uezato K, Funabashi T (2003) A fast technique for unit commitment problem by extended priority list. IEEE Trans Power Syst 18(2):882–888
15. Ongsakul W, Petcharaks N (2004) Unit commitment by enhanced adaptive Lagrangian relaxation. IEEE Trans Power Syst 19:620–628
16. Venkatesh B, Jamtsho T, Gooi HB (2007) Unit commitment—a fuzzy mixed integer linear programming solution. IET Gener Transm Distrib 1(5):836–846
17. Cohen AI, Yoshimura M (1983) A Branch-and-Bound algorithm for unit commitment. IEEE Trans Power Appar Syst 102(2):444–451
18. Su C-C, Hsu Y-Y (1991) Fuzzy dynamic programming: an application to unit commitment. IEEE Trans Power Syst 6(3):1231–1237
19. Kazarlis SA, Bakirtzis AG, Petridis V (1996) A genetic algorithm solution to the unit commitment problem. IEEE Trans Power Syst 11(1):83–92
20. Simopoulos DN, Kavatza SD, Vournas CD (2006) Unit commitment by an enhanced simulated annealing algorithm. IEEE Trans Power Syst 21(1):68–76
21. Jeong YW, Park JB, Jang SH, Lee KY (2009) A new quantum inspired binary PSO for thermal unit commitment problems. In: Proceedings of IEEE international conference on intelligent system applications to power systems, Curitiba, Brazil, pp 1–6
22. Juste KA, Kita H, Tanaka E, Hasegawa J (1999) An evolutionary programming solution to the unit commitment problem. IEEE Trans Power Syst 14(4):1452–1459
23. Hadji MM, Vahidi B (2012) A solution to the unit commitment problem using imperialistic competition algorithm. IEEE Trans Power Syst 27(1):117–124
24. Balci H, Valenzuela J (2004) Scheduling electric power generations using particle swarm optimization combined with the Lagrangian relaxation method. Int J Appl Math Comput Sci 14(3):411–421

25. Saber NA, Salimi M, Mirabbbasi D (2016) A priority list based approach for solving thermal unit commitment problem with novel hybrid genetic-imperialist competitive algorithm. Energy 117(1):272–280
26. Kamboj VK, Bath SK, Dhillon JS (2016) Implementation of hybrid harmony search/random search algorithm for single area unit commitment problem. Int J Electr Power Energy Syst 77:228–249
27. Wang B, Li Y, Watada J (2011) Re-scheduling the unit commitment problem in fuzzy environment. In: Proceedings of IEEE international conference on fuzzy systems, Taipei, Taiwan, pp 1090–1095
28. Kamboj VK (2016) A novel hybrid PSO–GWO approach for unit commitment problem. Neural Comput Appl 27(6):1643–1655
29. Mallipeddi R, Suganthan PN (2014) Unit commitment—a survey and comparison of conventional and nature inspired algorithms. Int J BioInspir Comput 6(2):71–90
30. Rashedi E, Nezamabadi-pour H, Saryazdi S (2010) BSGA: binary gravitational search algorithm. Nat Comput 9(3):727–745
31. Hussien AG, Houssein EH, Hassanien AE (2017) A binary whale optimization algorithm with hyperbolic tangent fitness function for feature selection. In: Proceedings of IEEE international conference on intelligent computing and information systems, Egypt, pp 166–172
32. Eid HF (2017) Binary whale optimization: an effective swarm algorithm for feature selection. Int J Metaheuristics 7(1):67–79
33. Digalakis JG, Margaritis KG (2002) An experimental study of benchmarking functions for genetic algorithms. Int J Comput Math 79(4):403–416
34. Liang JJ, Suganthan PN, Deb K (2005) Novel composition test functions for numerical global optimization. In: Proceedings of IEEE swarm intelligence symposium, Pasadena, CA, pp 68–75
35. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–681
36. Kennedy J, Eberhart RC (1997) A discrete binary version of the particle swarm algorithm. In: Proceedings of IEEE international conference on computational cybernetics and simulation, Orlando, pp 4104–4108
37. Mirjalili S, Mirjalili SM, Yang X-S (2014) Binary bat algorithm. Neural Comput Appl 25(3–4):663–681
38. Mirjalili S (2016) Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Comput Appl 27(4):1053–1073
39. Emary E, Zawbaa HM, Hassanien AE (2016) Binary grey wolf optimization approaches for feature selection. Neurocomputing 172:371–381
40. Jeong Y-W, Lee W-N, Kim H-H, Park J-B, Shin J-R (2009) Thermal unit commitment using binary differential evolution. J Electr Eng Technol 4(3):323–329
41. Sriyanyong P, Song YH (2005) Unit commitment using particle swarm optimization combined with Lagrange relaxation. In: IEEE power engineering society general meeting, San Francisco, USA, pp 1–8
42. Panwar LK, Reddy SK, Kumar R (2015) Binary fireworks algorithm based thermal unit commitment. Int J Swarm Intell Res 6(2):87–101
43. Singhal PK, Naresh R, Sharma V (2015) A novel strategy-based hybrid binary artificial bee colony algorithm for unit commitment problem. Arab J Sci Eng 40(5):1455–1469
44. Khanmohammadi S, Amiri M, TarafdarHaque M (2010) A new three-stage method for solving unit commitment problem. Energy 35(7):3072–3080
45. Roque LAC (2016) Optimization methods for the unit commitment problem in electric power systems. Dissertation, University of Porto
46. Bai X, Wei H (2009) Semi-definite programming-based method for security-constrained unit commitment with operational and optimal power flow constraints. IET Gener Transm Distrib 3(2):182–197
47. Huang KY, Yang HT, Huang CL (1998) A new thermal unit commitment approach using constraint logic programming. IEEE Trans Power Syst 13(3):936–945
48. Chandrasekaran K, Hemamalini S, Simon SP, Padhy NP (2009) Thermal unit commitment using binary/real coded artificial bee colony algorithm. Electr Power Syst Res 84:109–119
49. Chandrasekaran K, Simon SP (2013) Optimal deviation based firefly algorithm tuned fuzzy design for multi-objective UCP. IEEE Trans Power Syst 28:460–471
50. Koodalsamy B, Veerayan MB, Koodalsamy C, Simon SP (2016) Firefly algorithm with multiple workers for the power system unit commitment. Turk J Electr Eng Comput Sci 24:4773–4789
51. Sun L, Zhang Y, Jiang C (2006) A matrix real-coded genetic algorithm to the unit commitment problem. Elect Power Syst Res 76(9–10):716–728
52. Saber AY, Senjyu T, Yona A, Funabashi T (2007) Unit commitment computation by fuzzy adaptive particle swarm optimization. IET Gener Transm Distrib 1:456–465

arXiv:1810.04250v2 [cs.CV] 11 Oct 2018

Bird Species Classification using Transfer Learning with Multistage Training

Sourya Dipta Das*[1] and Akash Kumar*[2]

[1] Jadavpur University, Kolkata , India , dipta.juetce@gmail.com
[2] Delhi Technological University, Delhi , India
kmr2907akash@gmail.com

Abstract. Bird species classification has received more and more attention in the field of computer vision, for its promising applications in biology and environmental studies. Recognizing bird species is difficult due to the challenges of discriminative region localization and fine-grained feature learning. In this paper, we have introduced a Transfer learning based method with multistage training. We have used both Pre-Trained Mask-RCNN and a ensemble model consists of Inception Nets(Inceptionv3 net & InceptionResnetv2) to get the both localization and species of the bird from the images. Our final model achieves an F1 score of 0.5567 or 55.67% on the dataset provided in CVIP 2018 Challenge.

Keywords: Bird species classification · Deep Networks · Transfer Learning · Multistage Training · Object Detection

[3]

1 Introduction

Bird species are recognized as useful biodiversity indicators. They are responsive to changes in sensitive ecosystems, whilst populations-level changes in behavior are both visible and quantifiable. Suffered from great species variation, it is difficult for non-professionals to identify the sub-category of a bird only by its appearance. However, it is exhausting to annotate all the images by human beings with expert knowledge. Thus, an automatic classification system for bird species are needed, which will be great convenience for many practical applications. For researchers working outdoors, shoot photos can be classified and analyzed immediately by the system, illustration books are no more needed. For the public, the system could provide much fun when combined with culture information like poems and legends. It will arouse peoples interest in birds and could benefit the protections of birds. Apart from that, Classifying bird species is an interesting problem for Fine-grained categorization, also known as subcategory recognition, which is a subfield in object recognition. In recent years, fine-grained classification stood out from basic-level classification, bringing promising applications and new challenges to computer vision society.

[3] Code is available at: https://github.com/AKASH2907/bird-species-classification

In this Bird Species Classification Challenge, our main focus was to classify birds from high resolution photographs taken from camera. In this task, to improve classification task, we have also provided localization of birds in the respective images with their class labels. The Main Challenges involving in this problem are given below.

1. large Intensity variation in images as pictures are taken in different time of a day (like morning, noon, evening etc.)
2. various poses of Bird (like flying, sitting with different orientation)
3. bird localization in the image as there are some images in which there are more than one bird in that image
4. Large Variation in Background of the images
5. various type of occlusions of birds in the images due to leaf or branches of the tree
6. Size or portion of the bird covered in the images
7. less no of sample images per class and also class imbalance

We have proposed a end to end deep learning based approach using transfer learning to learn both micro and macro level features from Bird ROIs. We have used pre-trained Mask-RCNN to get the Bird ROIs from the images and used Multistage training method to remove class imbalance partially & boost up the accuracy of our model. Further information about model is given in the respective section.

2 Related Work

In recent years, there are number of existing works to automate the classification of bird using audio data rather images. For this Purpose, Feature Extraction from audio signals has some advantages like species have distinctive calls and no line of sight is need for detection. However, there are some disadvantages also like an individual bird may emit no audio at all for extended period of time and cant able to count the no of birds precisely. Due to this reasons, there are growing number of studies to use computer vision and image-based techniques for this problem [2] [5] [4]. Atanbori et al [4] proposed a method to use motion features including curvature and wing beat frequency. Combined with Normal Bayes classifier and a Support Vector Machine classifier. Cheng et al [2] introduced discriminative features for bird species classification based on parts of birds and Marini et al [5] proposed an approach to use a color segmentation to eliminate background elements and compute normalized color histograms to extract feature vector for classification. Bird Species Classification with visual data is also important in domain of fine grained classification and there are significant contribution to bird species classification problem respect to this domain [1] [3] [6] [7]. There are also some few shot and deep learning based approaches [1] [6] [15] regarding this problem which have achieved considered amount of accuracy in their respective datasets.

3 Dataset

In this paper, we have used the dataset from the CVIP 2018 Bird Species challenge. Training dataset consists of 150 images with 16 species of birds and testing dataset contains 158 images. The dataset contains high resolution images ranging from 800 × 600 to 4000 × 6000. The dataset is class imbalanced with 5 images in one species to 20 images in other species.

4 Proposed Approach

4.1 Data Augmentation

To increase the number of training samples per class and reduce the affect of class imbalance, data augmentation is used. Relevant image augmentation techniques are chosen according bird type of each classes. Those techniques are Gaussian Noise, Gaussian Blur, Flip, Contrast, Hue, Add (add some values to each channel of the pixel), Multiply (multiply some values to each channel of the pixel), Sharp, Affine transform. As the dataset was quite small, the networks trained on the dataset, overfitted the dataset and does not generalize well on 150 images. After data augmentation, training dataset increased from 150 images to 1330 images.

Table 1. Table for Data Augmentation techniques for each of bird species

Species	Gaussian Noise	Gaussian Blur	Flip	Contrast	Hue	Add	Multiply	Sharp	Affine	Total
blasti	Yes	Yes	Yes	Yes	Yes	No	No	No	No	90
bonegl	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	78
brhkyt	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	65
cbrtsh	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	91
cmnmyn	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	91
gretit	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	78
hilpig	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	80
himbul	Yes	Yes	Yes	Yes	Yes	No	No	No	No	99
himgri	Yes	Yes	Yes	Yes	Yes	No	No	No	No	100
hsparo	Yes	Yes	Yes	Yes	Yes	No	No	No	No	81
indvul	Yes	Yes	Yes	Yes	Yes	No	No	No	No	81
jglowl	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	78
lbicrw	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	78
mgprob	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	78
rebimg	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	80
wcrsrt	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	80

4.2 Bird ROI(Region of Interest) Detection

To eliminate background elements or regions and also extract features from only body of the birds, Pretrained Object Detection deep nets are used. In this Model,

we have used Mask R-CNN [8] to localize birds in each image from both test & training dataset.We have used the pretrained weights of Mask R-CNN,trained on the COCO dataset [16] which contains 1.5 million object instances with 80 object categories(including birds).

4.3 Transfer Learning

Here, we have used transfer learning based approach to learn both micro and macro level feature extracted from bird images for classification.we have used ImageNet [10] pretrained weights to initialize our Deepnet model for training. ImageNet contains 1.2 million images belonging to 1000 classes.Training using pretrained ImageNet weights help us to learn fine-grained as well as global level features before hand and learn the deepnet more specific & discriminative features for each bird species which leads to increase the accuracy of our model.

Fig. 1. Mask R-CNN cropped birds images

4.4 Ensemble Model Architecture

We have used InceptionResNetV2 [12] & InceptionV3 [13] deepnet architectures to create a ensemble model as our classification model. The prediction vector from Inception V3 [13] and Inception ResNet V2 [12] weights are generated for each image at the time of testing. There are two cases with Mask R-CNN:

1) Birds Detection: If the Mask R-CNN detect birds in the image, then a batch of cropped bird images are created. The whole batch for that particular image is evaluated using both network weights. Both the prediction vectors are compared and then the specie is assigned based on the prediction value with the highest weight or prediction confidence value of the specie of bird is finally predicted.

2) No Bird Detection: Whole Image is predicted for the specie of bird using both the architecture weights. Though, the number of such cases are very less. The specie with the highest predicted value is added to final prediction vector.

Fig. 2 illustrates the overall process of bird detection using Mask R-CNN and species classification using ImageNet models.

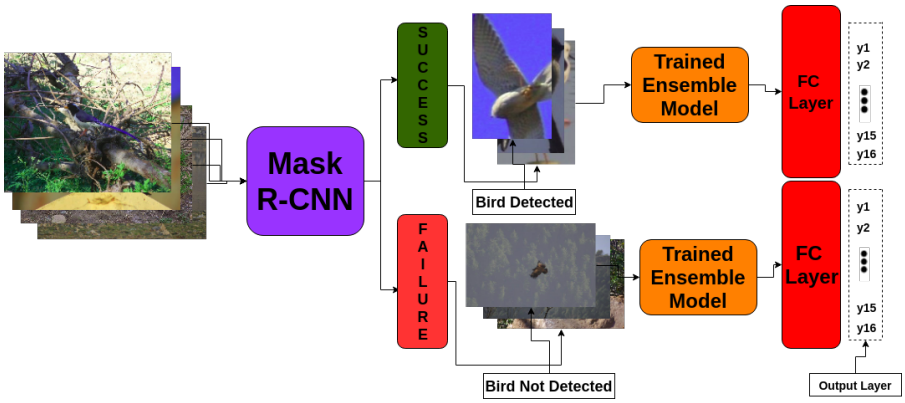


Fig. 2. Overall Architecture of our bird species classification system

5 Experiments

5.1 Multi-Stage Training

We used multi-stage training to improve the accuracy of the model. Firstly, we trained the Inception V3 architecture and then Inception ResNet V2 architecture on data augmented original images. In the second stage, we used the pretrained weight on original images to train on cropped images generated from Mask R-CNN. All the images are resized to 416 ×416. The accuracy of the model increased by 2-3 % after training on the cropped images. The multi-stage training helps to learn fine-grained features using cropped images of birds and the original images are used to learn the global spatial features present in the image.

5.2 Testing

At the time of testing, the images are passed from Mask R-CNN pre-loaded with COCO weights. The cases of Mask R-CNN is discussed in the Ensemble Model Architecture. Here, we have used 'categorical cross-entropy' as loss function & Adam as a Optimizer to train both deep networks.For fine-tuning of the deep-net models, we have tested on various types of activation functions and Swish activation function [14] performed best among all of them.

6 Evaluation

Evaluation Metric For the challenge, we used three evaluation metrics -

1. Precision: It is the ratio of correctly predicted positive observations to the total observations. It is defined as:

$$\frac{TruePositives}{TruePositives + FalsePositives} \tag{1}$$

2. Recall: It is the ratio of correctly predicted positive observations to all the observations in the relevant class.

$$\frac{TruePositives}{TruePositives + FalseNegatives} \tag{2}$$

3. F1-score: It is the harmonic mean of precision and recall.

$$2 * \frac{precision * recall}{precision + recall} \tag{3}$$

The evaluation metrics are calculated from Confusion matrix. In confusion matrix, True positives is equal to sum of diagonal elements. False positives is equal to the sum of each column excluding diagonal elements. False negatives is equal to the sum of each row elements excluding diagonal elements.

7 Results

Table-2 contains the F1-scores obtained from different architectures trained on original images and Mask R-CNN crops as discussed in Multi-Stage Training. The Inception ResNet V2 model trained firstly on resized original images and then on Mask R-CNN cropped images gives the best results at the time of training. The final trained weights of this model was used to predict the specie of bird.

Table 2. Accuracy during Multi-Stage Training on InceptionResnetV2 & InceptionV3 Models.

Model Architecture	**Data Subset**	**Train**	**Validation**	**Test**
Inception V3	Images	91.26	12.76	30.95
	Images + Crops	93.97	15.50	41.66
Inception Resnet V2	Images	97.29	29.17	47.96
	Images + Crops	92.29	33.69	49.09

Table-3 contains the class averaged precision, recall and F1 scores. In ensemble method, the prediction vector for both Inception V3 and Inception Resnet V2 is compared for each class predicted and then it was appended to the final prediction file.

Table 3. Evaluation Metrics (in %) on Test Dataset

Model Architecture	**Precision**	**Recall**	**F1**
Mask R-CNN + InceptionV3	48.61	45.65	47.09
Mask R-CNN + InceptionResnetV2	53.62	48.72	51.05
Mask R-CNN + Ensemble Model	**56.58**	**54.8**	**55.67**

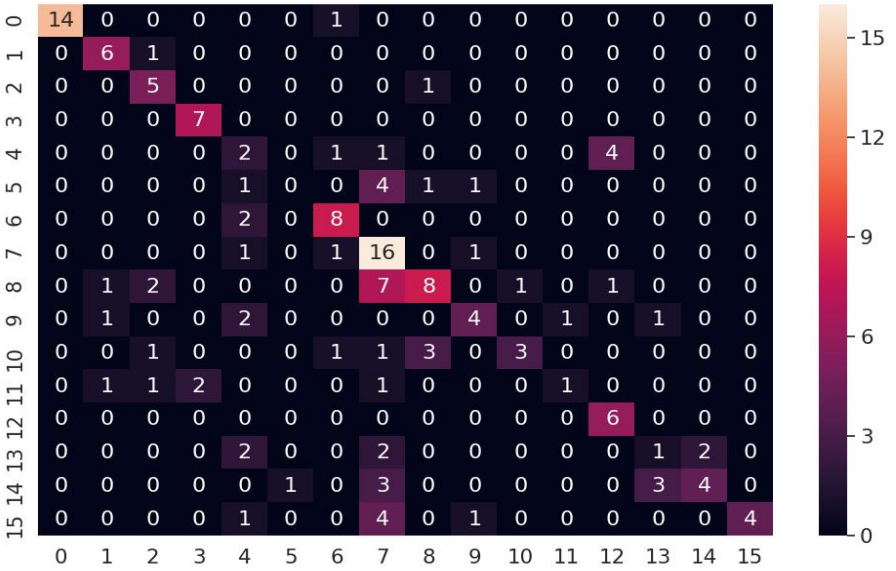


Fig. 3. Confusion Matrix of Mask R-CNN + Ensemble Model

8 Drawbacks

From the confusion matrix, cmnmyn, gretit, hsparo, indvul, mgprob and rebimg classes have poor accuracy.Few training samples per class with large variance and class imbalance are the Main Reason for those misclassifications. Though there are some other reasons like small bird ROIs , similarity in bird body part's colour & background colour and lighting condition change in training and testing dataset. For those conditions deep nets can't be able to learn those discriminative features (both micro features like colour, gradients,textures etc. and macro features like shape , colour patch etc.). Various lighting condition (like picture taken during daylight ,dawn, dusk, evening etc.) affected our model most as due to low background light, many micro features of the bird like colour,texture , gradients etc. are lost. Different poses of bird also reduced our model's accuracy though it is compensated with micro level features of the bird. Possible solution for those problems are discussed further in the later section.

9 Conclusion & Future Work

In this paper, we have proposed a method to both localize and classify the specie of the bird from high-definition photographs taken from camera by using a end-to-end approach with Mask R-CNN, transfer learning and multi-stage training. We have considered this challenge and as a few shot classification problem and

Fig. 4. Few example of images where Mask R-CNN fails to detect birds

proposed to use bird localization in images to boost up the accuracy of our model. Transfer learning helped our model to learn more specific and converge loss function more quickly with good accuracy on test dataset. In future, we are planning to extend this work using Part Model based approach with NTM (Neural Turing Machine) and Visual Attention Network.

References

1. J. Fu, H. Zheng and T. Mei, "Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 4476-4484. https://doi.org/10.1109/CVPR.2017.476
2. Cheng Pang, Hongxun Yao, and Xiaoshuai Sun. 2014. Discriminative Features for Bird Species Classification. In Proceedings of International Conference on Internet Multimedia Computing and Service (ICIMCS '14). ACM, New York, NY, USA, , Pages 256 , 5 pages.https://doi.org/10.1145/2632856.2632917
3. Z. Ge, C. McCool, C. Sanderson, A. Bewley, Z. Chen and P. Corke, "Fine-grained bird species recognition via hierarchical subset learning," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, 2015, pp. 561-565. https://doi.org/ 10.1109/ICIP.2015.7350861
4. Atanbori, John & Duan, Wenting & Murray, John & Appiah, Kofi & Dickinson, Patrick. (2016). Automatic classification of flying bird species using computer vision techniques. Pattern Recognition Letters. 81. 53-62. https://doi.org/10.1016/j.patrec.2015.08.015.
5. Marini, Andreia & Facon, Jacques & Koerich, Alessandro. (2013). Bird Species Classification Based on Color Features. Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013. 4336-4341. https://doi.org/10.1109/SMC.2013.740.
6. Branson, Steve et al. Bird Species Categorization Using Pose Normalized Deep Convolutional Nets. CoRR abs/1406.2952 (2014): n. pag.
7. Yang, Shulin et al. Unsupervised Template Learning for Fine-Grained Object Recognition. NIPS (2012).
8. He, Kaiming et al. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV) (2017): 2980-2988.
9. A. S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition", CVPR Workshop on Deep Vision, 2014.
10. A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks", NIPS, 2012.

11. C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9. https://doi.org/ 10.1109/CVPR.2015.7298594
12. Szegedy, Christian et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. AAAI (2017).
13. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, United States, 2016, pp. 2818-2826.https://doi.org/10.1109/CVPR.2016.308
14. Ramachandran, Prajit & Zoph, Barret & V. Le, Quoc. (2017). Swish: a Self-Gated Activation Function.
15. Vinyals, O., Blundell, C., Lillicrap, T.P., Kavukcuoglu, K., & Wierstra, D. (2016). Matching Networks for One Shot Learning. NIPS.
16. Lin TY. et al. (2014) Microsoft COCO: Common Objects in Context. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham
17. Sainin, Mohd & Alfred, Rayner & Adnan, Fairuz & Ahmad, Faudziah. (2018). Combining Sampling and Ensemble Classifier for Multiclass Imbalance Data Learning. 262-272. https://doi.org/10.1007/978-981-10-8276-4_25.

Volume 6, Issue 3(2018) 172-174 ISSN 2347-3258
International Journal of Advance Research and Innovation

Cement Industry and Strategies for Mitigating Carbon Emissions: An Overview

Ahana Ghosh*, Anubha Mandal
Department of Environmental Engineering, Delhi Technological University, Delhi, India.

Article Info:
Article history
Received 18 August 2018
Received in revised form:
28 August 2018
Accepted 2September 2018
Available online 15 September 2018

Keywords: Cement, carbon, strategies, carbon tax, industrial, environment

Abstract
Cement is the largest mass manufactured man made product on earth. The demand for cement is on a continual rise, as more and more developing countries strive for better infrastructure. This demand has, however, entailed an unacceptable increase in the carbon emissions as the cement manufacturing industry is one of the most carbon releasing industries in the world; responsible for more than 5% of the global carbon emissions. The dangerously high levels of Carbon Dioxide have contributed to a large scale climate change which has global repercussions. The need of the hour is an effective yet inexpensive mechanism to trim down the carbon emissions from the cement factories. In this paper, the main industrial as well as the governmental strategies for alleviating the carbon emissions of the cement industry are reviewed, focusing on the carbon taxation for the latter. This review has observed a comprehensive literature in term of the peer reviewed journals, research papers, industry reports, authentic websites etc on the cement industry and the strategies to reduce the carbon emissions.

1. Introduction

The manufacture of most industrial materials has some form of impact on the environment. Research is being conducted to reduce this impact and promote sustainable development.

One such industrial material which has a substantial effect on the environment, specifically on the carbon levels in the atmosphere is the vastly manufactured and consumed raw material: cement.

Currently, the cement industry accounts for almost 5%-7% [1] of the global CO_2 production. Over the recent years, cement production has witnessed an exponential increase in developing countries to meet the needs of a rapidly urbanizing civilization [2] and the carbon levels in the environment have seen a proportional increase. If nothing is done to control them, this ubiquitous industry will account for nearly 33% of the global carbon levels by 2050. [3]

This is indeed a cause for worry, as CO_2 is a greenhouse gas which when present in large quantities in the atmosphere can contribute to dangerous phenomena like global warming and climate change.[4] In such a situation, it is the need of the hour to find economically viable methods to help propagate a low CO_2 emitting cement industry.

It has been estimated that fossil fuel combustions account for a mere 30% of the amount of CO_2 generated during the production of cement whereas the calcination of the limestone accounts for almost 60% of the total carbon emitted.[5] This is both good as well as bad news. The carbon emissions cannot be controlled in the cement industry by simply increasing the energy efficiency; the problem must be tackled at the very base itself, by adjusting or changing the constituents of the cement, while keeping in mind the economic feasibility of the cement.

This paper gives an overview of the various major strategies that the government as well as the cement industry have considered to lessen the amount of carbon released from the cement production process. The rest of the paper is organized as follows: The next section reviews the industrial plans and the subsequent section focuses on the governmental initiatives. After that, the conclusions are discussed and suggestions are made for the areas that merit further research.

2.Industrial Plans

If the carbon emissions are to be reduced, it is essential that the industries take initiative without any external pressure. For this, awareness as well as incentives by the government is necessary. Discussed below are some of the strategies applied by different cement companies to mitigate their emissions.

2.1 Carbon Capture and Storage

Carbon Capture and Storage (CCS) is one of the foremost technologies that are available in the industry today. It is the process of capturing the carbon from the source site (like a cement plant) and then transporting it to a storage facility and depositing it in a location where it will not impair the environment. Over the last two decades, the feasibility of this method has been researched upon considerably. [6] Research has shown that the expenditure of this strategy would cause in increase in the cost of cement production 2-3 times. [7] As such, it is yet to be proven for large scale use and is mainly suitable for those industries which do not have any other option to reduce their carbon footprint.

2.2 Use of Supplementary Cementitious Materials

These materials can be used to replace a certain percentage of the clinker that is used to manufacture the cement and hence reduce the carbon content of the clinker. These materials are mostly the by-products of other industries like fly ash, calcined clay, natural pozzolans or geopolymers and silica fume. Due to their advantages like cost effectiveness [2], long term durability [8] and ease of use, they have been used since the 1990s in the cement industry [9]. However, they come with their own set of disadvantages. The availability of these materials varies regionally [7] and the potential for their usage for most of these materials has already been explored. Hence there is little scope for further carbon reduction using these methods. Moreover, the biggest question that most research done till now fails to answer is that exactly upto how much percentage can the substitution be done without compromising on the durability of the cement? [10, 2] Furthermore, the cost analysis in terms of the labour required is still vague in most literature. For example, geopolymers are hailed to have widespread advantages like reduction of the CO_2 emissions by 44-64% [11] over Ordinary Portland Cement, increased durability and better workability [12]. However, the disadvantages of these geopolymers like the fact that the making of the geopolymer concrete requires handling of the hazardous wastes and hence requires specialized training [13] and other technical difficulties were not accounted for in the cost analysis. Lastly, field studies have not yet proved conclusive for most of these materials and further research is needed to improve their usage in cement.

2.3 Use of Cement made from Alternative Clinkers

The main source of the carbon emission during cement manufacture is the stage at which the calcination of the CaO occurs. [14]. Hence, cement made from the alternative clinkers may allow for a substantial reduction in the carbon production. The paper by Gartner and Sui [15] provides an exhaustive analysis of the alternatives to Portland cement. Some of their conclusions are given in Table 1..

2.4 Use of Alternative Fuels

Fuels account for only 25-30% [16] of the carbon released during the cement manufacturing process and hence changing their compositions does not have a massive impact on the carbon emissions. The area does not have much more scope for exploitation [7] as a variety of alternative fuels have already been tried and tested and are currently being used in the cement industry.

***Corresponding Author,**
E-mail address: ahanaghosh1997@yahoo.com
All rights reserved: http://www.ijari.org

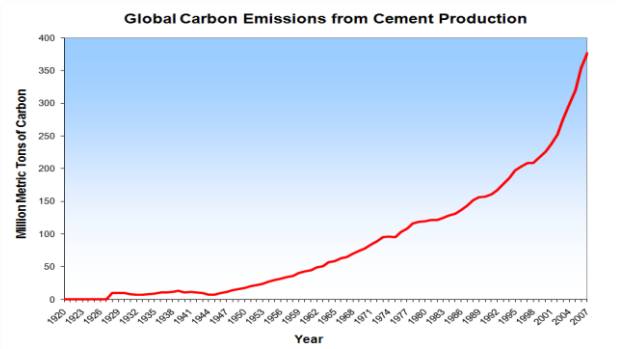


Fig 1. Global Carbon Emissions from Cement Production (Boden, T.A., G. Marland, and R.J. Andres. 2010. Global, Regional, and National Fossil-Fuel CO2 Emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. doi 10.3334/CDIAC/00001_V2010)

Table 1: Comparison of Alternative Clinkers

CEMENT	ADVANTAGE	DISADVANTAGE
1] Belite rich Portland Cement	Carbon reduction upto 10% per unit clinker	Takes more time to gain strength as compared to OPC
2] Belite Calcium Sulfo-aluminate (CSA) Cements	Carbon reduction upto 20% per unit clinker.	The cost of the raw material is very high.
3]Magnesium Based Cements	Made from ultramafic rocks which have the inherent ability to capture carbon. Globally abundant raw materials.	As of now, no energy efficient industrial manufacturing process has been invented.

3. Governmental Policies

Due to the difficulty faced by the cement industry in reducing the carbon emissions, the attention of the worldwide leaders and organizations as well as the governments of individual countries have riveted towards policies and strategies that could reduce the domestic greenhouse gases emissions. They can be very broadly divided into 3 types of approaches: - Voluntary approaches (VA), Trading and Carbon Taxes (CT). In this section, after a brief overview of VA and Trading, we shall focus on the discussion of Carbon Taxes.

3.1 Voluntary Approaches

VAs can be classified into one of four types: unilateral commitments by industry; private agreements between industry and stakeholders; environmental agreements negotiated between industry and government; voluntary programmes developed by government that individual firms can join. [17]. While this policy has been applauded for its flexibility and it's relatively less effect on the competitiveness of the companies which produce cement, it has also been criticized for the laxity in the coverage of the industries and the ineffectiveness of the implementation.

3.2 Carbon Trading

Carbon trading is a flexible mechanism introduced by the Kyoto Protocol, which limits the carbon emission from the industries by granting companies the permit to emit only a certain amount of carbon. This cap and trade mechanism, although effective to a certain degree, as proved by the studies of Shammin and Bullard (2009), is also a very complicated and expensive mechanism to implement. This system requires a completely new administrative system to aid the establishment of a competent trading market. [18] Moreover, the carbon trading mechanism merely shifts the production of carbon from one country to the other; which cannot be a permanent solution to the carbon problem as this issue is a global issue.

3.3 Carbon Taxes

Due to the vast carbon footprint of the cement industry, the governments of various countries have tried implementing a tax known as Carbon Tax in order to hold the cement companies answerable for the carbon that they generate. If set high enough, it becomes a potent financial incentive that motivates switches to clean energy across the economy, simply by making it more economically rewarding to move to less carbon intensive manufacturing methods. Carl and Fedor (2016) have come to the conclusions that an effective carbon tax with a good rate has been preferred over other governmental policies, especially since the last decade. Research has been done exhaustively in quite a few areas relating to the carbon tax. Certain conclusions from them are discussed below.

Research done states that the different impacts of the carbon tax could arise from the fact that different rates are followed in different countries [19] and that the design of a proper carbon tax may be able to alleviate the negative impacts of its implementation. [20]. Moreover, the carbon tax can be considered an efficient system only when the tax is set high enough that companies have enough incentive to switch to a lower carbon intensive manufacturing process.[20]. However in China, an analysis done to evaluate the preference of companies to carbon tax found that companies prefer a low rate of carbon tax in the beginning; about 1 to 4 US dollars per tonne of CO_2 emitted.[21]. This, they claim, will reduce the carbon emissions and at the same time, not affect the competitiveness of the companies. A research done based on the Saudi Arabian cement industry suggests a compromise at 27 US dollars per ton of carbon [6] claiming that at this rate the profit of the industries would not be compromised and the emissions would significantly be reduced. Suggestions have also been made of starting at a low carbon tax rate and then gradually moving to high rates to avoid a sudden economic pressure on the industries. [22]

A survey done in the University of Geneva evaluates carbon taxes with respect to their competitiveness, distributional and environmental effects. [23]. Competitiveness indicates the ability of a company to sell its goods and services in the domestic as well as the global market. Distributional effects can be regressive i.e. the bulk of the tax falls more on the low income population or progressive i.e households with higher income pay proportionately more. The studies conclude that revenue recycling may be an interesting method offset the losses due to reduction in competitiveness and that in general; carbon taxes are indeed regressive unless subsidies are provided to the low carbon intensive companies. Moreover, carbon leakage is a said to be major a cause for concern [24] as companies may simply shift the manufacture of carbon intensive merchandise to countries without a carbon tax or with lax governmental regulations. The conclusions drawn from this discussion are presented below.

4. Conclusions

The main conclusions of the above review can be summarized as follows:

1. From the discussion on industrial policies, it can be concluded that CCS is not yet a feasible option and could be made more cost effective by methods like Carbon Capture and Usage, where the stored carbon can then be used profitably.
2. Supplementary cementitious materials are already in use in the cement industry and further research has to be done to ascertain the exact amount of substitution that can be done in the cements without affecting their strength. If the strength is affected by the substitution, the substitution becomes redundant as more amounts of the same clinker will be needed to achieve the same strength.
3. Alternative fuels for the cement industry have already been explored exhaustively and not much potential for further CO_2 reduction exists in that area. However, using alternative fuels may have other advantages to the environment.
4. The use of alternative cement clinkers with a different composition is one area that merits further research. If the cost analysis of different types of cements is done, they could begin replacing the OPC in the markets. For this, the market bankability of the Portland cement has to be taken into account. It would definitely help if experts and educationalists raised

awareness about the benefits of using different types of cements and hence increase their reliability.

5. From the governmental policies it can be concluded that while Voluntary Approaches are flexible mechanisms, their effectiveness is very difficult to gauge and may require more standardization.
6. Carbon trading while having the advantage of giving flexibility to the companies who can decide which method they want to use to reduce their carbon emissions, has the disadvantage of not being transparent and being very easy to evade. Moreover, the method is expensive to execute and often complicated.
7. Carbon taxes on the other hand are simple enough in theory but may give rise to multiple complications. They are transparent and easy to implement. However, the imposition of carbon taxes raises quite a few problems. If set too high, they may affect the competitiveness of the industry and raise carbon leakage issues, if set too low, they lose their effectiveness as they do not provide sufficient incentive to the companies to reduce the emissions. If the companies are dissatisfied with the imposition of this tax, it may promote them to produce carbon in covert operations, which may cause more damage to the environment. Policies by the governments need to address the issue of competiveness directly through compensation mechanisms for the deserving companies. Moreover, most research finds that carbon taxes are regressive in nature.
8. In such a scenario, the design of a carbon tax is highly important and very few studies exist where the design of the carbon tax is given sufficient consideration. The design of the carbon tax should also be done taking into account different cements and technologies available. Moreover, most studies do not account for uncertainty and perform no risk analysis. This could be an important missing link as every factor affecting the carbon tax is subject to a lot of variability.
9. The promotion of a less carbon intensive cement industry requires that the government and the industries work together through a mixture of strategies and technologies based on the different domestic conditions of various countries.

Acknowledgements

I would like to thank Dr. Anubha Mandal for sparing her valuable time and mentoring me throughout this paper.

References

[1] PK Mehta. Reducing the Environmental Impact of Cement", Concrete International, 2001.

[2] A Bilodeau, VM Malhotra. High Volume Fly Ash System: Concrete Solution for Sustainable Development, ACI Materials Journal, Title No. 97-M6, 2000.

[3] J Davidovits. Global Warming Impact on the Cement and the Aggregates Industries, World Resource Review, 6(2), 1996, 263-278.

[4] Z Chen, P Nie. Effect of carbon tax on social welfare: A case study in China, Applies Energy 183, 2016, 1607-1615.

[5] X Liu, Y Fan, C Wang. An estimation of the effect of carbon pricing for CO_2 mitigation in China's cement industry, Applied Energy 185, 2016, 671-686.

[6] W Matar, AMElshurafa. Striking a balance between profit and carbon dioxide emissions in the Saudi cement industry, International Journal of Greenhouse Gas Control 61, 2017, 111-123.

[7] KLScrivener, VM John, EM Gartner. Eco-efficient cements: Potential, Economically viable solutions for a low CO_2, cement based industry, UNEP report, 2015.

[8] C Valderrama. Implementation of the best available techniques in cement manufacturing: a life cycle assessment study", Journal of Cleaner Production 25, 2011, 60-67.

[9] N Muller, J Harnisch. A blueprint for a climate friendly cement industry,2008http://assets.panda.org/downloads/english_report_lr_pdf.pdf

[10] YC Diaz. Limestone Calcined clay cement as a low carbon solution to meet expanding cement demand in emerging economies, Development Engineering 2, 2017, 82-91.

[11] BC McLellan. Costs and carbon emissions for geopolymer pastes in comparison to ordinary Portland cement, Journal of Cleaner Production 19, 2011, 1080-1090.

[12] M Bediako. Maximizing the Sustainability of Cement Utilization in Building Projects through the Use of Greener Materials.", Journal of Engineering, ID 1375493, 2016, 6.

[13]http://civilengineersforum.com/geopolymer-concrete advantages-disadvantages/

[14] S Mishram, NA Siddiqui/ A Review on Environmental and Health Impacts of the Cement Manufacturing Industry, International Journal of Geology, Agriculture and Environmental Sciences, 2, 2014.

[15] World Steel Association, 2015, Worldsteel in figures, Brussels.

[16] Eugeniusz Mokrzycki, Alicja Uliasz- Bochenczyk, Alternative Fuels for the Cement Industry, Applied Energy 74, 2003 95-100.

[17] S Bygrave, J Ellis. Policies to Reduce Greenhouse Gas Emissions in Industry - Successful Approaches and Lessons Learned: Workshop Report,OECD and IEA report, 2003.

[18] ZE Gevrek, A Uyduranonglu. Public Preferences to Carbon Tax Attributes, Ecological Economics 118, 2015, 186-197.

[19] B Lin, X Li. The effect of carbon tax on per capita CO_2 emissions", Energy Policy 39, 2011, 5137-5146.

[20] TC Kuo, IH Hong, SC Lin. Do carbon taxes work? Analysis of governmental policies and enterprise strategies in equilibrium", Journal of Cleaner Production 139, 2016, 337-346.

[21] X Liu. An analysis of company choice policy preference to carbon tax policy in China, Journal of Cleaner Production 103, 2014, 393-400.

[22] Z Zhang. How to improve the performance of Carbon Tax in China, Journal of Cleaner Production 142, 2016, 2060-2072.

[23] A Baranzini. A future for carbon taxes", Ecological Economics 32, 2000, 395-412.

[24] R Sathre, L Gustavsson. Effects of energy and carbon taxes on building material competitiveness, Energy and Buildings 39, 2000, 488-494.

Volume 6, Issue 3(2018) 205-208 ISSN 2347-3258
International Journal of Advance Research and Innovation

Comparative Study of Mechanical Properties of TIG Welded Joints of Similar and Dissimilar Grades of Stainless Steel Material

Roop Lal, Mohd Shuaib, Vikal Paliwal
Department of Mechanical Engineering, Delhi Technological University, Delhi, India

Article Info
Article history
Received 18 August 2018
Received in revised form
28 August 2018
Accepted 02September 2018
Available online 15 September 2018

Keywords: Tig welding, AISI 304, AISI 316, Tensile strength, Hardness,

Abstract
This paper discuss and experimentally analyse the mechanical properties of TIG welded joints of similar and dissimilar grades of stainless steel and to find out the best joint. In this study. Different combinations from samples of AISI 304 and AISI 316 are welded under standard conditions by using Tungsten Inert Gas welding (TIG Welding) technique,.. The welded joints were tested for Tensile strength, Hardness & toughness and best combination is pointed out. The three combinations analysed were AISI 304-304, AISI 316-316 and AISI 304-316. It was found that the tensile strength and toughness of AISI 304-304 combination was highest and maximum hardness is observed in AISI 304-316 combination. The results indicate that there is not a single combination which excels in all tests.

1. Introduction

The Gas Tungsten Arc Welding (GTAW) process is also referred as TIG, or Heliarc. Helium gas had been used for the first time during this welding process, from where the term Heliarc has been produced. Until the 19th century, forge welding was the only technique known for joining metal pieces [2]. In the late 1930s and 1940s, an aircraft industry had used this technique for welding of magnesium.

TIG welding is a metal joining process in which electricity is used to form an electric arc between the tip of a non-consumable tungsten electrode and the work piece [3]. The heat from the arc melts the metal pieces and with the help of filler material joint is formed. In order to provide the shield to an arc, inert gas has been used which passes from the GTAW torch, this shield also prevent the heat affected zone, molten metal and tungsten electrode from an atmospheric contamination. Generally, in the TIG welding, Argon and Helium gases are the preferred due to their inert gases as they do not react with metals that need to being joined. Along with this benefit, the shielding gases also used in order to cover the weld pool and thereby avoiding the contamination of weld portion.

TIG welding is different from most other welding processes because of its process of forming arc and adding filler material. While performing the TIG Welding one hand is used for holding the TIG torch that will produce an arc and the other hand is required to add the filler metal to the weld joint. Here tungsten is used because of its properties which makes an arc to maintain its temperature up to 11000^0F. Due to its high melting point and excellent electrical conductivity it is used for welding with a hotter arc then the actual melting point of the tungsten. Although it is cleaner and more versatile that results in a high quality welds and can be works on steel, aluminium, brass and many other metals but this is more expensive than a normal welding.

Also the Stainless Steel has been used here which is an alloy of an iron with a chromium (minimum 10.5%). This chromium forms a passive layer which is a thin layer of oxide on the surface of the steel which prevents the corrosion of the surface. There are also other elements present in it which makes it useful for variety of purposes. Some of the elements are carbon, silicon, manganese, nickel, molybdenum, sulphur, etc.

2. Experimental Procedure

The two grades of austenitic stainless steel i.e. AISI 304 and AISI 316 are used. The material was in the form of sheets of 1mm thickness. Both of these have same elements in them except for the presence of molybdenum which is only present in AISI 316 thus making it useful in the harsh chemical environment. The details of their respective composition are shown in following table 1 [8]

The process started with welding of two rectangular piece if AISI 304. The same process was repeated with AISI 316. At last two rectangular sheets of AISI 304 an AISI 316 are welded together. Prior to the welding, the specimen was cleaned properly with sand paper followed by cleaning with acetone. For welding a standard TIG welding apparatus was used. The specimen was kept on a horizontal surface and clamped properly before welding. The work piece was connected to positive terminal and electrode with negative terminal which gave us Direct current straight polarity setting. Under the protective environment of Argon gas with a flow rate of 15L/min, the welding was started. A dimension of the specimen for EDM cutting has been shown as such in the figure 1. The standard dimensions used were 120mm x 100mm x 1mm.

Table 1: Chemical compositions of specimens

	Elements	Fe	C	Mn	Si	Mo	Cr	Cu	Ni	
Composition (% by weight)	AISI 304	70.78	0.025	1.140	0.14	0.36	18.40	0.18	8.19	
	AISI 316	67.69	0.018	1.28	0.38	2.42	16.63	0.21	10.85	

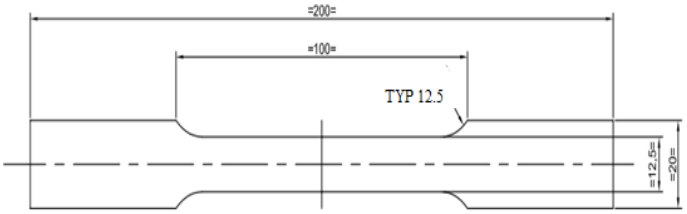


Fig.1. Dimension of Specimen

For each case a different filler rod was used which have same composition as parent material. The filler rods used in each case have different composition which is shown in the following table 2.

Table 2: Chemical composition of filler material (% weight)

	C	Si	S	P	Mn	Ni	Cr	Mo
ER316	0.019	0.470	0.003	0.024	1.570	12.20	18.5	2.30
	C	**Mo**	**Nb**	**-**	**Mn**	**Ni**	**Cr**	**-**
ER304	0.063	0.22	0.01	-	0.9	8.11	18.4	-

The voltage was kept constant while current was varied. The variation of current in each case along with other parameters used in welding is shown in table 2. After the welding, we took out the three pieces for testing from each of the combination with the help of an EDM machine.

After getting the required dimension, we need to test different mechanical properties such as tensile test, Hardness and Toughness on different machine set-up.

3. Experimental Test

3.1. Tensile test

It is basically the property of a material that tells the maximum amount of tensile stress that it can take before failure, such as permanent deformation [2]. It specifies the point when a material goes from elastic to plastic deformation. The tensile strength of the specimen was

Corresponding Author,
E-mail address: rooplalrana@dce.ac.in; shuaib89141@gmail.com; vickyvp20@gmail.com
All rights reserved: http://www.ijari.org

determined by UTM (Universal Testing Machine) as shown in fig. 4.14. For that a H50K-S UTM from Tinius Olsen of 50kN Capacity is used. The speed of the UTM was kept at 0.5mm/min. The tensile strength of the metal is proportional to its ductility, more is the tensile strength then more will be its ductility and vice versa [7]. It was observed form the test that the strength of welded joint increased proportionally with the welding current in all the cases. For different combinations of metals, we have used different set of the parameters in order to achieve optimum quality welding. It is imperative to note that while performing the test for the tensile strength of the specimens, the parting off took place away from heat affected zone in all the three cases. This shows that the joint obtained was more than 100% efficient. The universal testing machine has been shown in the following figure 2.

Fig. 2. Universal Testing Machine

3.2. Hardness

It is the property of a material that enables the material to resist indentation of localized displacement. It is the resistance to indentation, wear and abrasion. The strength coefficient and strain hardening exponent gives a measure of increase in hardness due to plastic deformation [8]. Hardness tests are used as a means of controlling the properties of materials. Once the hardness is calculated, by altering the composition of various compounds of material a certain desired properties can be obtained. Here we have uses the Brinell hardness machine. In the Brinell tests, the specimen is mounted on the anvil of the machine and a load of 750 kg is applied against a hardened steel ball which is in contact with the surface of the specimen being tested. The steel ball is 5 mm in diameter. The load is allowed to remain 5 second and is then released, and the depth of the depression made by the ball on the specimen is measured. The Resultant Brinell hardness number was obtained by the eq. (1):

$$\frac{P}{\frac{\pi D}{2}(D-\sqrt{D^2-d^2})} \qquad (1)$$

Where P is applied load, D is diameter of steel ball and d is diameter of impression.

3.3. Toughness

It is the ability of a material to absorb energy. As the ductility of a metal decrease, the toughness of a material decreases with it [6]. The toughness of the specimen is obtained from the stress strain graph formed from the UTM. This is done by measuring the area under the curve of stress strain curve. The toughness is expressed as energy per unit volume.

4. Results and Discussion

4.1 Tensile strength

4.1.1 AISI 304-304

The results obtained from tensile test are shown in Table. From the Stress Vs Strain curve and Force-Displacement graph it was observed that the joint bore the force upto 4880 N before finally undergoing fracture. The fracture was observed from the heat affected zone on the parent metal. The maximum stress the joint experienced was of the order of 390 Mpa which is clearly seen in the stress Vs Strain curve. Furthermore, it was observed that the combination experienced a consolidated elongation of 40.8 %. The graph shows various results obtained.

Table 3: Results obtained in tensile test of AISI 304-304 combination

Width	Thickness mm	Initial Gauge length mm	Final gauge length mm	Area *mm*	Ultimate force N	Ultimate stress MPa	offset @0.2 % N	offset @0.2 % MPa	TE %
12.5	1.00	82.2	115	12.5	4880	390	1730	138	40.8

The stress vs strain curve and force displacement graphs in this case is given by figure 3 and figure 4.

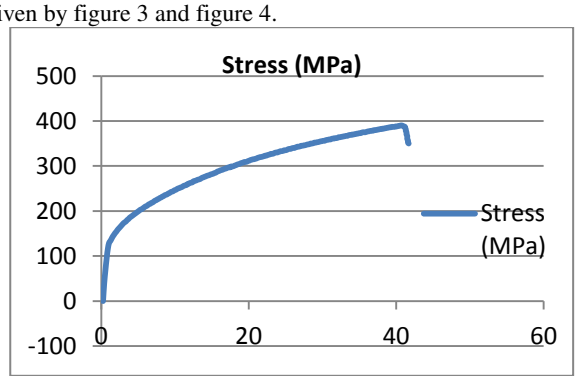


Fig. 3. Stress Vs Strain curve of 304-304 combination

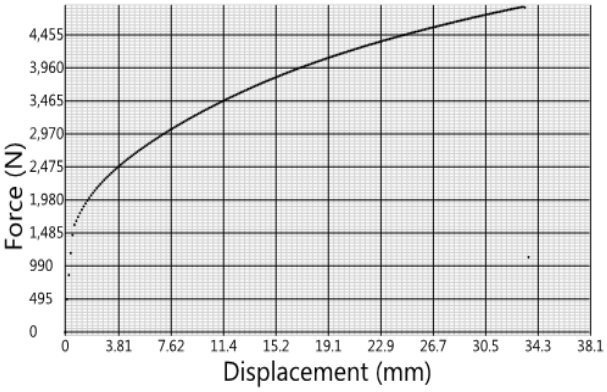


Fig. 4. Force Vs Displacement of 304-304 combination

Fig.5. AISI 304-304 Specimen after Tensile Test

4.1.2 AISI 304-316

From the stress Vs Strain curve and force-displacement graph it was observed that the joint bore the force upto 4480 N before finally undergoing fracture. The maximum stress the joint experienced was of the order of 357 MPa which is clearly seen in the stress Vs Strain curve. Furthermore, it was observed that the combination experienced a consolidated elongation of 28.8 %. The fracture was observed from the heat affected zone on the parent metal. The results obtained are shown in table 4.

Table 4: Results obtained in tensile test of AISI 316-316 combination

Width mm	Thickness mm	Initial Gauge length mm	Final gauge length mm	Area *mm*	Ultimate force N	Ultimate stress MPa	offset @0.2 % N	offset @0.2 % MPa	TE %
12.5	1.00	82.2	106	12.5	4480	357	2060	165	28.8

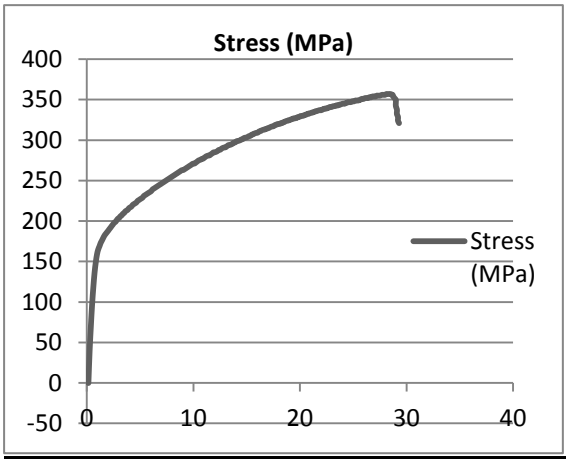


Fig.6. Stress Vs strain curve of 316-316 combination

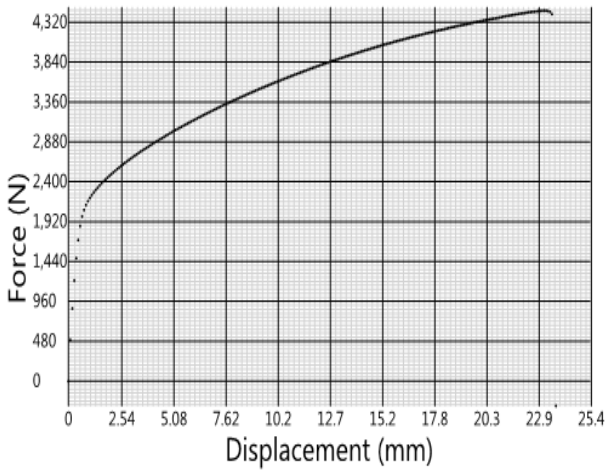


Fig.7. Force Vs displacement curve of 316-316 joint

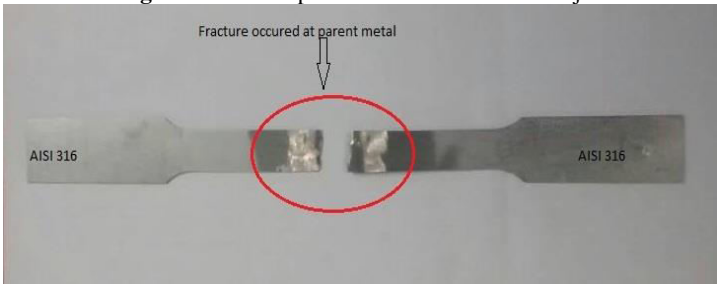


Fig.8. AISI 316-316 Specimen after Tensile Test

4.1.3.AISI 304-316

From the stress Vs Strain curve and force-displacement graph it was observed that the maximum stress the joint experienced was of the order of 300 MPa. The joint bore the force upto 3750 N before finally undergoing fracture which is clearly seen in the stress Vs Strain curve. Furthermore, it was observed that the combination experienced a consolidated elongation of 10.7 %. The joint in this case fractured from parent metal of AISI 316.

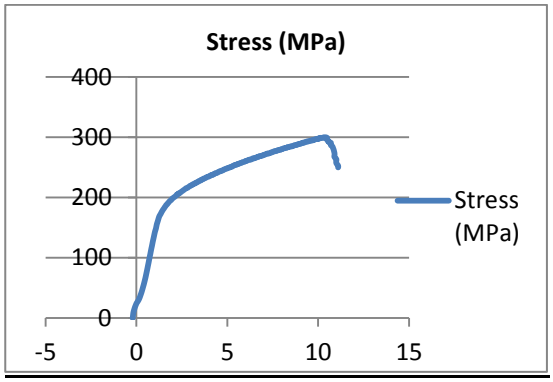


Fig.9. Stress Vs strain curve of 304-316 joint

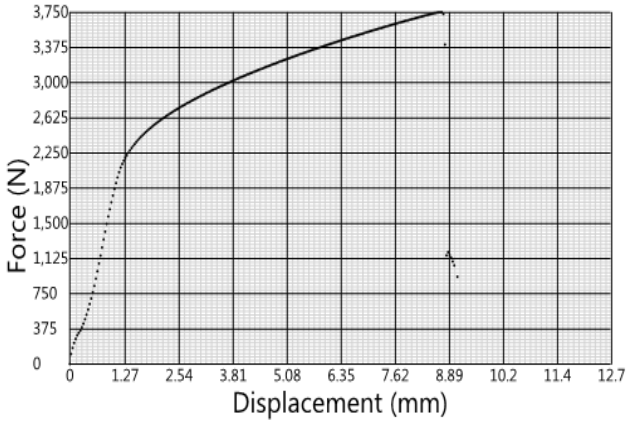


Fig.10. Force Vs displacement curve of 304-316 joint

Table 4: Results obtained in AISI 304-316 combination

Width Mm	Thickness Mm	Initial Gauge length Mm	Final gauge length Mm	Area ***mm***	Ultimate force N	Ultimate stress MPa	offset @0.2 % N	offset @0.2 % MPa	TE %
12.5	1.00	82.2	91.3	12.5	3750	300	2280	182	10.7

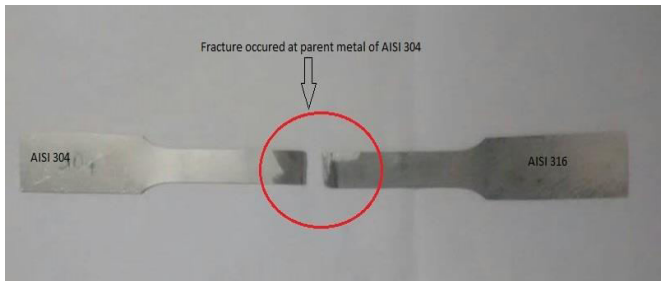


Fig.11. AISI 304-316 Specimen after tensile test

4.2 Hardness

The results obtained from hardness test performed on Brinell Hardness testing machine are as follow. The following figure shows the indentation obtained from the test.

4.2.1 For AISI 304- AISI 304 welded Specimen

Load applied: 750kg
Time for which load is applied: 5second
Diameter of ball: 5mm
Diameter of Impression produced: 2.4mm
Brinell Hardness Number (BHN)

$$= \frac{P}{\frac{\pi D}{2}(D-\sqrt{D^2-d^2})} = 154.85 \text{ BHN}$$

4.2.2 For AISI 316- AISI 316 welded Specimen

Load applied: 750kg
Time for which load is applied: 5second
Diameter of ball: 5mm
Diameter of Impression produced: 2.8mm
Brinell Hardness Number (BHN)

$$= \quad = \frac{750}{\frac{3.14*5}{2}(5-\sqrt{5^2-2.8^2})} = 112.40 \text{ BHN}$$

4.2.3 For AISI 304- AISI 316 welded specimen :

Load applied: 750kg
Time for which load is applied: 5second
Diameter of ball: 5mm
Diameter of Impression produced: 2 mm
Brinell Hardness Number (BHN)

$$= \frac{750}{\frac{3.14*5}{2}(5-\sqrt{5^2-2^2})} = 228.56 \text{ BHN}$$

4.3 Toughness

The toughness was measured from stress vs strain curve. By measuring the area under the curve, the toughness was obtained. The following table shows the toughness obtained in each case.

Table 6 : Results obtained from toughness

Combination	AISI 304-304	AISI 316-316	AISI 304-316
Toughness(Energy per unit volume)	12388.44	8376.50	10495.28

Table 7. Comparative Results

Combination/Properties	Tensile Strength (Mpa)	Toughness (Energy per unit volume)	Hardness (BHN)
AISI 304-304	390	12388.44	154.85
AISI 316-316	357	8376.5	112.4
AISI 304-316	300	10494.28	228.56

4. Conclusions

The major conclusions drawn from this study are as following:

- It is concluded from the experiment that for proper welding the parameters need to be changed according to the working conditions. Appropriate welding cannot be accomplished. By keeping the variables constant in every case,
- It has been observed that specimens should be clamped properly using the fixtures because during solidification of metals, their tendency of shrinking may results in bending of the specimen near the weld area. So in order to prevent such errors they need to be clamped.
- It has been found that the hardness of AISI 304-AISI 316 combination is more than that of AISI 304-AISI 304 and AISI 316-AISI 316 combination. Whereas the tensile strength of AISI 304-AISI 304 combination is found to be more than that of AISI 316-304 and AISI 316-316 combinations.
- It has been observed that there is not any single combination which has the highest magnitude of properties in every case. So in case whenever these kinds of materials are available, one need to make a balance between tensile strength & hardness according to their requirements.

This study can be useful in field of manufacturing and production. For example, now a day the automobile components are made as lighter as possible to minimize fuel consumption by using metals like stainless steel. So in case AISI 304 and AISI 316 are available, the manufacturer can take the help of data obtained from this study and according to his requirement, can choose the right combination for manufacturing.

References

[1] P Mithilesh, D Varun, ARG Reddy, KD Ramkumar, N Arivazhagan, S Narayanan. Investigations on Dissimilar Weldments of Inconel 625 and AISI 304. International Journal of Engineering Science and Technology 75, 2014, 66 – 70.

[2] VA Rao, R Deivanathan. Experimental Investigation for Welding Aspects of Stainless Steel 310 for the Process of TIG Welding. Procedia Engineering 97, 2014, 902 – 908.

[3] P Peasura, A Watanapa. Influence of Shielding Gas on Aluminium Alloy 5083 in Gas Tungsten Arc Welding. Procedia Engineering 29, 2012, 2465 – 2469.

[4] MP Reddy, AASWilliam, MM Prashanth, SNS Kumar, KD Ramkumar, N Arivazhagan, S Narayanan. Assessment of Mechanical Properties of AISI 4140 and AISI 316 Dissimilar Weldments. Procedia Engineering 75, 2014, 29 – 33.

[5] P Peasuraa, A Watanapa. Influence of Shielding Gas on Aluminium Alloy 5083 in Gas Tungsten Arc Welding. Procedia Engineering 29, 2012, 2465 – 2469.

[6] RK Buddu, N Chauhan, PM Raole. Mechanical properties and microstructural investigations of TIG welded 40 mm and 60 mm thick SS 316L samples for fusion reactor vacuum vessel applications. Fusion Engineering and Design 89,2014, 3149–3158.

[7] KD Ramkumar, AChandrasekhar, AK Singh, S Ahuja, A Agarwal, N Arivazhagan, AM Rabel. Comparative studies on the weldability, microstructure and tensile properties of autogenous TIG welded AISI 430 ferritic stainless steel with and without flux. Journal of Manufacturing Processes 20, 2015, 54–69.

[8] RK Desu, HN Krishnamurthy, A Balu, AK Gupta, SK Singh. Mechanical properties of Austenitic Stainless Steel 304L and 316L at elevated temperatures. Journal of Material Research and Technology, 5(1), 2016, 13–20.

[9] N Jeyaprakash, AHaile, MArunprasath. The Parameters and Equipments Used in TIG Welding: A Review. The International Journal of Engineering and Science (IJES), 2015.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 6 (2018) pp. 17-22
© Research India Publications. http://www.ripublication.com

Comparative study of surface coating for automotive application-A review

Ramesh ChandraYadaw[1], Shailesh Kumar Singh[1], Somanth Chattopadhyaya[1] Sanjeev Kumar[2] R.C.Singh[3]

[1]Department of Mechanical Engineering, Indian Institute of Technology (IIT-ISM) Dhanbad, India
[2]Department of Mechanical Engineering, Krishna Engineering College, Ghaziabad, India
[3]Department of Mechanical Engineering, Delhi Technological University, Delhi, India
Corresponding author

Abstract

Automobile sectors are facing major challenges to materials automotive parts with high performances and durability due to international intrinsic norms. A new combinations of the materials with high hardness, good wear resistant, and environment friendly in order to meet the international intrinsic norms. The HVOF coating is considered to be an excellent replacement to chrome plating. The coating process is more environmentally friendly than the chrome plating process. Also, the coating has potentially lower or equivalent production cost when compared to chrome.

Keywords: Thermal spray, Chromium alternatives, automotive applications, tribomechanical properties.

1. Introduction: An automotive industry sector is possibly among the most technologically demanding of all industry segments. In order to face the challenges & meet market competitions, Scientist and researchers are constantly exploring new technologies and materials to enhance tribological behavior, life of the automotive parts, surface morphology, fuel efficiency, reduce weight, lower emissions, corrosive resistant and in the end best economy. In an internal combustion engine, the tribological performance of the piston-ring has long been recognized as important in achieving desired engine efficiency and durability in terms of power loss, fuel consumption, oil consumption, blow by, and even harmful exhaust emissions. Back in the 1940's piston rings were in the main Chromium plated for wear resistance. Although still used in the mid 1960's, this was superseded by the thermal spray process to overcome Fretting wear in the chromium plated rings. The development of the Thermal Spray Process since its introduction has gone from Flame sprayed coatings to include Plasma sprayed coatings and High Velocity Oxygen Fuel Sprayed Coatings.

In our current efforts to find alternatives of chrome plating, we have studied various coating methods such as Chrome plating, thermal sprayed coating (Plasma and flame spray) as well as unlimited no of coating materials. The results from various coating methods were collected.

2. Materials &Methods:

2.1. Types of deposition techniques: Steel and cast iron are widely used metals in construction, mechanical, tooling, aircraft and chemical industries [1 -2], as they have good mechanical properties and relatively cheaper. But its hardness, corrosion resistance and wear restrict their uses in demanding conditions and environments. Deposition techniques are very practical and effective for the improvement of hardness, corrosion resistance and wear properties of these materials [3-4].The various deposition techniques are being used for surface coating purpose, some typical techniques are as follows.

1) Thermal spraying technique
2) Physical vapor Technique
3) Chemical vapor technique
4) Galvanize
5) Baked polymer
6) Hard Chrome plating

2.1.1 Thermal Spraying

Thermally sprayed hard metal coatings have a typical coating thickness within the range 100–500 μm [5]. Thus, thermal spray enables the functionality of hard metals to be realized on the surface of large parts, which cannot be obtained by powder metallurgy for economical and technical reasons.

Thermal spray technology has several different deposition processes that all use heat sources to liquefy the coating materials and to project the liquefied particles on to the surface to be coated. Thermal spray processes can be based on thermal energies obtained from combustion of gases, such as hydrocarbons or hydrogen, or liquid fuels or can be obtained from electric discharges. Besides actual thermal spray processes, such as flame spray with powder or wire material, electric arc spraying, plasma arc spraying, and HVOF spraying, processes entirely based on deposition by kinetic energies are also classified under the simple term thermal spray processes. Some thermal spraying processes are as following.

2.1.1. a. High velocity Oxy Fuel (HVOF):

In this method, the fuel generally propylene, propane, or hydrogen, is mixed with oxygen and burned in a chamber and the products of the combustion are allowed to expand with the help of nozzle where the gas velocities may become supersonic. Powder is introduced along the axis, in the nozzle and is heated and accelerated. The powder is usually fully or partially melted and achieves velocities of near about 540 m/s. Since the powder is exposed to the products of combustion, they may be melted in either a reducing or an oxidizing medium, and oxidation of metallic and carbide is possible.HVOF processes can produce coatings of any metallic or cermet materials.

2.1.1. b. Detonation Gun:

In the detonation gun process, a mixture of oxygen and acetylene, along with a pulse of powder, is introduced into a barrel and detonated by creating spark. Virtually all metallic, ceramic, and cermet materials can be deposited using detonation gun deposition. Detonation gun coatings are used widely for wear and corrosion resistance as well as for many other types of applications.

2.1.1.c. Plasma Spray Technique:

A plasma spray consist of a gas, usually argon, but sometimes including nitrogen, hydrogen, or helium, is allowed to flow between a tungsten cathode and a water-cooled copper anode. An electric arc is created between the two electrodes using a high frequency discharge and then sustained using do power. The arc ionizes the gas, creating high-pressure gas plasma. Powder is generally introduced into the gas stream either just outside the torch or in the diverging exit region of the nozzle (anode). It is heated and accelerated by the high-temperature, high-velocity plasma gas stream. Plasma spray may be used to produce coatings of virtually any metallic, cermet, or ceramic material

A major advantage of the thermal spray processes is the extremely wide variety of materials that can be used to make a coating. Virtually any material that melts without decomposing can be used. A second major advantage is the ability of most of the thermal spray processes to apply a coating to a substrate without significantly heating it.

2.1.2. Physical Vapor Deposition (PVD) & Chemical Vapor Deposition (CVD):

In Physical Vapor Deposition (PVD) and Chemical Vapor Deposition (CVD) process, the coating material is deposited onto the substrate in the form of ions. Only thin coatings ((µm range) can be deposited using these processes. This is the main limitation of CVD and PVD compared to thermal spray processes. However metals, alloys as well as refractory compounds can be deposited using CVD and PVD processes.

2.1.3. Baked polymer:

Baked polymer is used for corrosion resistance and aesthetics. The coating thickness varies from 1-10 µm.The galvanize method has coating thickness 1-5 µm and used for corrosion resistance purpose. [6-7]

2.1.4. Hard chrome plating:

Hard chrome plating is an electrolytic process in which chromium is deposited onto a metal substrate giving it enhanced properties for durability, wear and hardness.

Hard Chrome plating is applied to ferrous and nonferrous materials to improve wear and abrasion resistance, reduce friction, prevent seizing and galling, and to restore the dimensions of undersized parts. Hard chrome plating has been a trusted industry solution for wear, erosion, corrosion resistance and dimensional reclamation for many years. It can be applied at a reasonable cost per unit of surface area, but has limitations on thickness build-up, part size, and in some instances performance in service. Functional hard chrome plating is a critical process associated with manufacturing and maintenance operations on aircraft, vehicles and ships, both in civilian and military sectors.

2.2. Wear Test Methods:

An apparatus for wear testing is termed as wear tester, tribotester or tribometer. From a surface engineering point of view, wear test is carried out to evaluate the potential of using a certain surface engineering technology to reduce wear for a specific application, and to investigate the effect of treatment conditions (processing parameters) on the wear performance, so that optimized surface treatment conditions can be realized.

There are some important types of wear tester, discussed briefly as following.

2.2.1 Pin-on-disc wear tester:

In a pin-on-disc wear tester, a pin is loaded against a flat rotating disc specimen such that a circular wear path is described by the machine. The machine can be used to evaluate wear and friction properties of materials under pure sliding conditions. Either disc or pin can serve as specimen, while the other as counter face. Pin with various geometry can be used. A convenient way is to use ball of commercially available materials such as bearing steel, tungsten carbide or alumina (Al_2O_3) as counter face, so that the name of ball-on-disc is used.

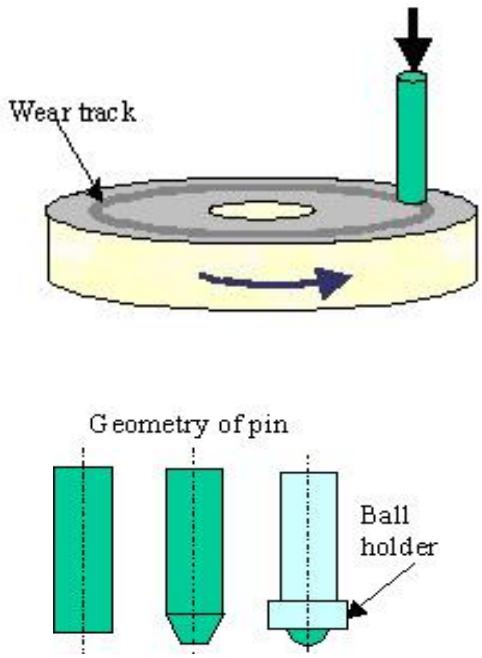


Fig.1 [8]

2.2.2 A rolling sliding wear tester:

Rolling-sliding wear tester is used for investigating wear as well as frictional behavior of materials under rolling, sliding, or a combination of both.

2.2.3 Abrasive wear tester:

The abrasive wear tester is used for abrasive wear testing of materials in reciprocating sliding. This machine has several advantages such as high reproducibility, short test time, simple flat test geometry, easy evaluation, simple operation. This machine has some disadvantages. The friction cannot be measured by this machine, some parameters such as humidity and temperature cannot be controlled.

2.2.4 High temperature tribometer:

The High Temperature Tribometer is made to analyze tribological behavior at elevated thermal conditions. This instrument can accurately simulate in-service conditions at high temperature and features a unique dual heating setup which makes the instrument highly reliable and stable up to 1000°C.

2.2.5 Reciprocating:

The linearly reciprocating tribometer is utilized to perform wear tests of two friction pairs under relatively oscillating movement. This is performed by means of moving the upper friction partner via linear drive in a cyclical translatory motion, while the lower friction partner is pressed under the defined normal force. An essential future of this tribometer is the big variance of applicable sample pairs and probe geometries. In addition to standard geometries (pin-disk; ball-disk) by means of using specific sample holders, the tribological performance of real geometries can be tested with or without lubricants.

2.3 Types of wear: Undesirable removal of material from operating solid surface is known as wear. Wear may be classified as:

2.3.1 Abrasive wear:

Abrasive wear, sometimes called cutting wear, occurs when hard particles slide and roll under pressure, across the tooth surface. Abrasive wear is caused by the passage of relatively hard particles/asperities over a surface. Hard particle sources are: dirt in the housing, sand or scale from castings, metal wear particles, and particles introduced into housing when filling with lube oil.

2.3.2 Adhesive wear:

Adhesive wear is very common in metals. It is dependent on the mutual affinity between the materials. When brass pin under load pushed in indium bloack, and subsequently retracted, some particles of indium transferred on brass pin and loss of indium takes place. These loss occurs due to adhesive force between brass and indium.[9].

2.3.3. Corrosive wear:

The fundamental cause of corrosive wear is a chemical reaction between the material and a corroding medium which can be a chemical reagent, reactive lubricant or even air. When sliding surfaces interact with the environment, reaction takes place and form reaction product like oxide, chloride etc that causes wear. [9]

2.3.4. Erosive wear:

Erosive wear caused by the impact of particles (solid/liquid) against a solid surface. For example dust particles impacting on gas turbine blades and slurry impacting on pump impeller. [9]

2.3.5 Fretting wear:

It refers to small amplitude (1 to 310 μm), with high frequency oscillatory movement mainly originated by vibration. This generally occurs in mechanical assemblies (press fit parts, rivet / bolt joints, strands of wire ropes, rolling element bearings), in which relative sliding on micron level is allowed. It is very difficult to eliminate such movements and the result is fretting. Fretting wear and fretting fatigue are present in almost all machinery and are the cause of total failure of some otherwise robust components.[9]

3. Result & Discussions:

The various deposition techniques are being used to improve wear resistant properties, surface morphology and microstructral properties. Various parameters were collected and compared in the following table 1.The various mechanical properties were compared.

The Fine CrC–NiCr coatings demonstrate superior performance to hard chrome with regard to mechanical and tribological properties. For future successful applications of CrC–NiCr HVOF coatings as alternative to hard chromium, many factors like wear resistance, friction coefficient, costs and environmental issues have to be considered collectively.

The HVOF system produces a near-theoretical dense WC–Co coating that exhibits a high elastic modulus which imparts high compressive residual stress in the substrate. The hard chrome plating process induces a tensile residual stress in the substrate that allows the crack initiation process to start at the periphery of the specimen; thus, the rate of the fatigue process increases and the fatigue strength decreases.

S.No.	Deposition method	Coating materials	Substrate materials	Mechanical properties		References
				Hardness		[10]
1.	PVD	MoN	Al+Si alloy	2000±400 HV		
2.	PVD	pure chromium deposits,	Steels	425-840 HV		[11]
3	PVD	Cr-C, Cr-N	Steels	3500HV	fretting wear, fatigue and corrosion properties superior to those of electrolytic hard Chrome.	[11]
4.	Electrolytic method	Hard chrome	1010 mild steel	1000±200		[12]
5.	HVOF	W(-17(	AISI 43		High Compressive residual stress	[13]
6.	Hard Chrome plating	Hard C	AISI 43		High tensile residual stress	[13]
7.	HVOF	CrC -NiC	Ste		Mechanical & Tribological properties superior than chrome	[14]
8.	Hard Chrome plating	Hard cl	ste		Mechanical & Tribological properties inferior than CrC-NiCr	[14]
9.	Hard Chrome plating	Hard chrome	Mild steel	884±97		[15]
10.	HVOF	WC-12%Co	Mild steel	1293		[16]

Table.1

3.1. Micro structural analysis:

The polished cross-section of the CrC75–NiCr25 Fine-5 μm and Fine-10 μm coatings, measured by HVOF thermal spraying process, are shown in Fig.2. The coating thickness was found to be approximately 150 μm.

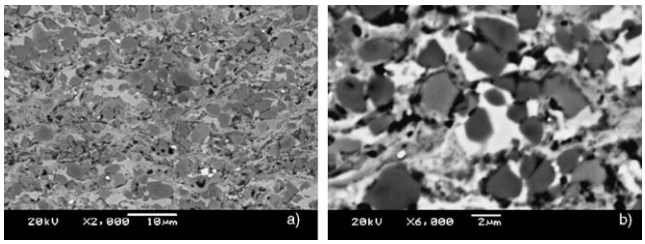


Fig. [14] ,BSE images showing microstructure detail of the CrC75–NiCr25 coating, consisted of a metallic binder Ni–Cr solid solution (bright areas) and dispersed carbide CrC phase (dark areas): 2(a) Fine-10 μm and2 (b) Fine-5 μm coatings

The coatings sprayed from the Fine-10 μm and Fine-5 μm powders show a dense and coherent microstructure, consisted of a metallic binder Ni–Cr solid solution (bright areas) and dispersed carbide CrC phase (dark areas)

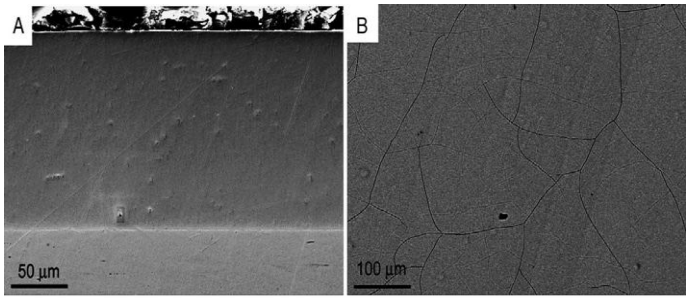


Fig.3

[17] SEM micrographs of EHC coatings. 3A: thick plating on polished substrate, cross-section; 3B: thick plating on polished substrate, surface

EHC coatings are completely non crystalline. The Cross-sectional SEM micrographs (Fig. 3A) exhibit that EHC coatings possess a very homogeneous microstructure, where several typical small vertical cracks are the only detectable feature. No remarkable differences are observable between the various coatings. Excellent adhesion, with almost no interfacial defects, is detectable

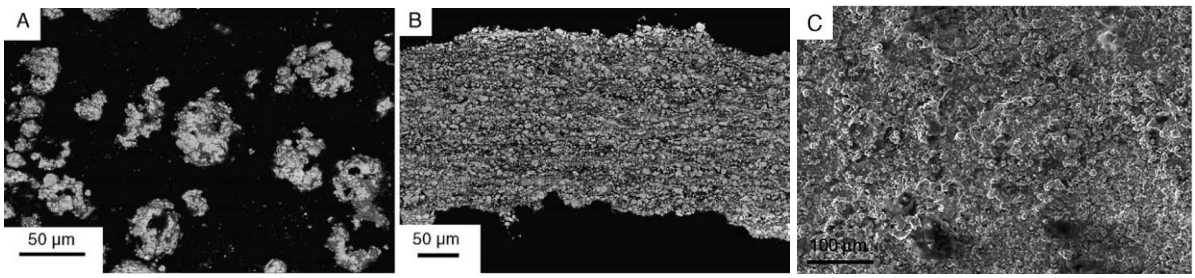


Fig.4 [17] WC-17% Co coating. Fig.4A: SEM micrograph of starting WC-17% Co powders; Fig.4B: SEM micrograph of cross-section; Fig.4C: SEM micrograph of surface

HVOF-sprayed coatings surface is definitely different. The WC-17% Co coating, obtained from a good quality powder with uniform carbide distribution among various grains (Fig. 4A), possesses a very dense microstructure. (Fig. 4B).The coating surface is quite irregular (Fig. 4C), with the metal matrix incorporating slightly protruding carbide grains.

3.2. Effect of type of substrate and heat treatment in Mass loss: There is no effect of substrate type and preheating in mass loss. The heat treatment of coating influenced significantly the mass loss. The heat treated coating had approximately 3 times lower mass loss than as sprayed coating.[18-20]

3.3. Effect of heat treatment on friction and abrasive wear behavior: Temperature ranges played very significant role .Its effect on friction and wear resistance were remarkable. The high temperature heat treatment of coatings was detrimental to friction and abrasive wear resistance. The low temperature heat treatment increased wear resistance of WC-12Co and Cr3C2–25NiCr coatings under benign abrasive wear conditions [21-24].

3.4 Effect of coating materials on health & environment: Hard chrome electroplating is commercially used to produce wear-resistant automobile parts, but the plating bath contains hexavalent chromium, which has adverse health and environmental effects. [17]

3.5 Wear behavior of HVOF and hard chrome coated piston ring: A highly dense chrome carbide based High velocity oxy fuel coated were applied on piston ring. The ring wear behavior was determined to be six times lower than hard chrome coating. The production cast was also less as compared to the hard chrome coating. [25

4. Conclusions:

4.1 Hard chrome coating has adverse effect on health and environment issues.

4.2 The HVOF coatings of Fine CrC–NiCr demonstrate superior performance to hard chrome with regard to mechanical and tribological properties. This may the alternative of hard chrome plating. For future successful applications of CrC–NiCr HVOF coatings as alternative to hard chromium, many factors like wear resistance, friction coefficient, costs and environmental issues have to be considered collectively.

4.3. Wear behavior of HVOF coated piston ring is better than hard chrome coated piston ring.

4.5 The heat treatment of coating influenced significantly the mass loss. The heat treated coating had approximately 3 times lower mass loss than as sprayed coating.

4.6 The high temperature heat treatment of coatings was detrimental to friction and abrasive wear resistance. The low temperature heat treatment increased wear resistance of WC-12Co and Cr_3C_2–25NiCr coatings under benign abrasive wear conditions.

5. Future scope:

There may be some more alternative compositions, e.g. such as nano materials doping can be used with Cr and also based on TiC or use of iron-based or iron-containing binders may be of importance.

As on date, majority of the studies were based on WC-Co, WC-Ni or WC-Co-Cr or Cr_3C_2-NiCr but there was no or very rare study found with WC-NiCr or Cr_3C_2-C_O combinations There may the scope to work with these combinations.

References:

[1] M.F. Ashby, D.R.H. Jones, Engineering Materials: 2. An Introduction to Microstructures, Processing and Design, Pergamon Press, Oxford,1986, p. 4.

[2] H. Skuleva, S. Malinovb, W. Shac,T, P.A.M. Basheer ‘Microstructural and mechanical properties of nickel- base plasma sprayed coatings on steel and cast iron substrates’Surface & Coatings Technology 197 (2005) 177–184

[3] H. Skulev, S. Malinov, P.A.M. Basheer, W. Sha, Surf. Coat. Technol. 185 (2004) 18.

[4] H. Skulev, S. Malinov, P.A.M. Basheer, W. Sha, Surf. Coat. Technol. 185 (2004) 18.

[5] J.R. Davis (Ed.), Surface Hardening of Steels: Understanding the Basics, ASM International, Materials Park, OH, 2002.

[6] L.-M. Berger, Application of hardmetals as thermal spray coatings, Int J RefractMet HardMater (2014).

[7] . Sulzer Metco, An Introduction to Thermal Spray, Issue 4 • c 2013 Sulzer Metco

[8] C. X. Li, The University of Birmingham, UK, Wear Testing and Wear Measurement, Surface Engineering.

[9] Singh, Shailesh Kumar, et al. "Influence of Nano-particle on the Wear behaviour of Thin Film Coatings A Review." International Journal of Applied Engineering Research 13.6 (2018): 4053-4058.

[10] Hanbey Hazar * "Characterization of MoN coatings for pistons in a diesel engine" Materials and Design 31 (2010) 624–627

[11] A. AUBERT AND R. GILLET "HARD CHROME COATINGS DEPOSITED BY PHYSICAL VAPOUR DEPOSITION *"***Thin Solid** Films, 108 (1983) 165-172

[12] Amjad Saleh El-Amoush*, Ayman Abu-Rob, Hmood Edwan, Khalid Atrash, Mohannad Igab "Tribological properties of hard chromium coated 1010 mild steel under different sliding distances" Solid State Sciences 13 (2011) 529e533

[13] A. Ibrahim ,, C.C. Berndt "Fatigue and deformation of HVOF sprayed WC–Co coatings and hard chrome plating" Materials Science and Engineering A 456 (2007) 114–119

[14] Picas JA, Forn A, Mattaus G. HVOF coatings as an alternative to hard chrome for pistons and valves. Wear 2006; 261:477-84.

[15] S ˇ.Houdkova´ • F. Zaha´lka • M. Kasˇparova´ • L. -M. BergerComparative Study of Thermally Sprayed Coatings Under Different Types of Wear Conditions for Hard Chromium Replacement

[16] Shabanaa*, M.M.M. Sarcar b, K.N.S.Sumanc,S Kamaluddind. 'Tribological and Corrosion behavior of HVOF Sprayed WC- Co, NiCrBSi and Cr3C2-NiCr Coatings and analysis using Design of Experiments'Materials Today: Proceedings 2 (2015) 2654 – 2665

[17] Giovanni Bolelli, Valeria Cannillo, Luca Lusvarghi*, Sara Ricco` "Mechanical and tribological properties of electrolytic hard chrome and HVOF-sprayed coatings, Surface & Coatings Technology 200 (2006) 2995 – 3009

[18]. Bolelli, G., Cannillo, V., Lusvarghi, L., Soragni, E., Loreto, A., Valente, T.: A comprehensive characterization of industrially manufactured hard chrome platings and of thermally sprayed alternatives. In: Proceeding of the International Thermal Spray Conference ITSC 2005, Basel, 2–4 May 2005, pp. 1456–1461. DVS-Verlag, Basel, Switzerland, Du¨sseldorf (2005). CD (ISBN 3- 87155-793-5)

[19]. Guilemany, J.M., Espallargas, N., Suegama, P.H., Benedetti, A.V., Fernandez, J.: High-velocity oxyfuel Cr3C2–NiCr replacing hard chromium coatings. J. Therm. Spray Technol. 14, 335–341 (2005)

[20].Wank, A., Schwenk, A., Wielage, B., Grund, T., Friesen, E., Pokhmurska, H.: Behavior of thermally sprayed wear protective coatings exposed to different abrasive wear conditions in comparison to hard chromium platings. In: Conf. Proc. Int. Thermal Spray Conference & Exhibition ITSC 2007, 14–16 May 2007, Beijing, China, pp. 1011–1016. ASM International, Materials Park, Ohio, USA (2007)

[21] Nitesh Vashishtha, S.G. Sapate, Pranay Bagde, A.B. Rathod,Effect of heat treatment on friction and abrasive wear behavior of WC- 12Co and Cr3C2-25NiCr coatings, S0301-679X(17)30477-2

[22]. Thiele S, Sempf k, Roessler KJ, Berger LM, Spatzier J, Thermophysical and microstructural studies on thermally sprayed tungsten carbide-cobalt coatings. J Therm Spray Technol 2011;20:358-65.

[23]. Roy M, Pauschitz A, Polak R, Franek F. Comparative evaluation of ambient temperature friction behaviour of thermal sprayed Cr3C2-25(Ni20Cr) coatings with conventional and non-crystalline grains. Tribol Int 2006;39:29-38.

[24] Kumar, Dalip, Qasim Murtaza, and R. C. Singh. "Sliding wear behavior of aluminum alloy coating prepared by two-wire electric arc spray process." The International Journal of Advanced Manufacturing Technology 85.1-4 (2016): 237-252.

[25] F. Rastegar *, D.E. Richardson "Alternative to chrome: HVOF cermet coatings for high horse power diesel engines, Surface and CoatingTse chnolog9y 0(1997)1 56-163

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/327971500

Determination and Allocation of Optimal Size of TCSC and STATCOM for obstruction alleviation in Power System

Conference Paper · October 2018

CITATIONS
0

READS
12

2 authors, including:

Jitender Kumar
JIMS Engineering Management Technical Campus, Greter Noida, India
11 PUBLICATIONS **19** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Optimal placement & allocation of Statcom View project

All content following this page was uploaded by Jitender Kumar on 30 September 2018.

The user has requested enhancement of the downloaded file.

Determination and Allocation of Optimal Size of TCSC and STATCOM for obstruction alleviation in Power System

Jitender Kumar
Department of Electrical Engineering,
Delhi Technological University,
New Delhi, India
jitender3k@gmail.com

Narendra Kumar
Department of Electrical Engineering,
Delhi Technological University,
New Delhi, India
dnk_1963@yahoo.com

***Abstract*—The traditional technique of power system for measuring, monitoring and identifying the location of the overloaded power line is more complex in nature and time-consuming. The power system basic requirement to limit their parameter within a specified limit and keep close to as Unity either during underloaded or overloaded conditions of Bus. The Flexible AC Transmission System (FACTS) devices are the most effective tool which can be used for the purpose of minimising the impact of variation in load on a power system. Such devices work as the controllers and work effectively in mitigation of varying Voltage profile, enhance steady-state system stability, improve reactive power and power factor, etc to a specific limit. The size of a controller may vary on the basis of their requirement for a specific system. The TCSC and STATCOM are effective controllers of a power system to mitigate the multi-dimension variation of parameters. The proposed tools are implemented and tested on IEEE 14 Bus System individually for its efficacy. The MATLAB software is used for design of IEEE 14 Bus system and implementation of FACTS devices on Simulation Platform.**

Keywords— Traditional method, FACTS Devices, TCSC, STATCOM, Congestion Mitigation

I. INTRODUCTION

The traditional power system is more wide and complex in nature. The controller is static in nature and their design is fixed to a specific value. The controller basically uses Static Capacitor or Static Inductor of pre-defined value. As per normal operation of power system are categorised in three distinguish category as Generation, Transmission and Distribution. As per the Indian scenario, there are the different types of organisation which operate their own task and manage their operation optimally. The power generation will be performed by NTPC, NHPC, NPC, etc. The transmission part is handled by PGCIL, Sterlite Power, etc, and Distribution part of electricity is handled by multiple state organisation. So that complexity of power transfer from generation end to distribution end is a complicated task due to an economy of the entire system as per the customer requirement. The cost to consumer per unit is pre-defined and cost incurred during entire process are varied due to different organisation involved. The individual organisation performance plays a vital role in the economy of the entire power system.

The requirement of electric power for industrial and domestic usage is exponentially increasing as compared to improvement in transmission system capacity. There are bilateral and multilateral contracts between the government and private parties for establishing a proper coordination between generation side and demand side of power supply. The contract is well defined in terms of power transfer from one end to another end with specified power limit. Otherwise, some power lines are either overloaded or under-loaded during normal operation of power system. Such type of operation whenever some lines are overloaded during normal operation is called congestion [2]. It forces the operating mechanism to improve its power transfer capability. Also, to ensure that the system with regards to security, stability and reliability is fruitful for the entire power system which is defined as congestion mitigation management.

In [9] author uses the concept to minimize the transmission line loading by using sensitivity factor and stability indices. The transmission lines during peak hours are under heavy stress of power flow. Such stress can be minimized by using FACTS device like TCSC of a specified value and an optimal location. The sensitivity factor and stability indices are used to allocate the size of FACTS device and for implementing the location. The paper [1] discusses the effect of wind penetration on power generation in association with FACTS devices. The authors apply two types of operation by a reduction in wind and second wind penetration increased and apply FACTS device to compensate voltage and power angle of the system. In paper [3], authors concentrate on voltage profile and control of active & reactive power flow among various buses. They have identified the effect of different FACTS devices on IEEE/CIGRE transmission system model. The different FACTS devices impact on power losses in an interconnected power system.

978-1-5386-6625-8/18/$31.00 ©2018 IEEE

The paper [7] suggests the optimal model of IEEE 14 Bus system on the basis of MATLAB/Simulation software. The authors have proposed [8] a Contingency Severity Index (CSI) and Fast Voltage Stability Index (FVSI) for optimising the size of FACTS devices. Due to that the Cost of Installation of FACTS devices was optimised with reduced overloading on a system.

In [10] optimal location of Shunt FACTS device identified with the help of series FACTS device to get highest possible maximum power transfer and maintain system stability. It has affected the degree of series compensations and decide the optimal allocation of Shunt devices. They proposed the location of FACTS device must be located near to load centre in place of mid-point of a transmission line. The [5] will be focussed on supplementary controller design of STATCOM for oscillation mitigation in power system. The design of controller was based on the requirement of variation applicable to voltage profile and requirement of Reactive power at a specific location of the bus.

The authors have suggested [11] the real power performance index & reactive power loss reduction for identification of the location of FACTS device in a transmission line. The power system will require restructuring for optimal usage of their potential which may lack due to an environmental problem, cost problem and right-of-way during expansion of power network. The FACTS device will help in optimising loaded lines, increase loading capability, minimise system losses, improve system stability, reduce power production cost and fulfil a contractual requirement.

In the literature it has been seen that most of the papers concentrate on the obstruction alleviation during heavy loading of a specific bus of the power system. By use of FACTS devices, will guide in deciding the overloading or under-loading of a congested bus during a loading operation. The TCSC and STATCOM are tested on IEEE 14 Bus System on MATLAB/Simulation platform.

II. FACTS DEVICES MODELLING

A. Thyristor Controlled Series Compensator (TCSC)

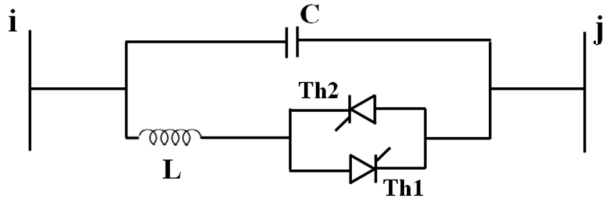


Fig. 1. TCSC

The Thyristor Controlled Series Compensator is known as TCSC [6], which is an advanced version of static Filter [9] operated with the help of power electronic devices. The parallel combination of capacitor and inductor with bidirectional control performed by power electronic devices. Such type of controller helps in varying the overall impedance of TCSC. The overall reactance of device varied from purely inductive to purely capacitive. The device is dynamic in nature and varied their reactance on the basis of requirement of the power system.

The TCSC is a series compensation device used to compensate the loading condition of the transmission line, improving system stability and enhance the reliability of power system. It never requires any interfacing device like a transformer, so simple in nature and economical as compared to other FACTS devices. The TCSC can mitigate the problem of Voltage compensation [1] and Reactive Power compensation during loading conditions. The variations of net reactance depends upon the firing angle of power electronic devices. The firing angle of TCSC varied from 0^o to 90^o for Th_1 and 90^o to 180^o for Th_2 [2]. Due to that, the effect of inductance will vary and accordingly overall impedance of TCSC will be varied.

- If $X_C > X_L$, The TCSC may involve the advantage of both Inductive and Capacitive region.
- If $X_C < X_L$, The TCSC will behave as Capacitive due to shunt connection of Capacitor and low reactance offer by TCSC for power system.
- If $X_C = X_L$, Then a resonance condition achieved which is harmful to power system. It has infinite impedance offer by TCSC and transmission line behave as an open circuit.

The impedance offer by TCSC treated as positive (for inductive) and negative (for capacitive) region. The region of TCSC [4] will varied on the basis of firing angle and can be calculated by

$$X_L = X_l \frac{\pi}{\pi - 2\alpha - \sin\alpha} \quad (1)$$

The x_{LC} can be correspond to as

$$x_{LC} = \frac{x_C x_L}{x_C - x_L} \quad (2)$$

$$Z_{TCSC} = j\, X_{TCSC} \quad (3)$$

$$X_{TCSC} = -x_C + \frac{x_C + x_{LC}}{\pi}[2(\pi - \alpha_{TCSC}) + \sin\{2(\pi - \alpha_{TCSC})\}]\left[\frac{4x_L^2 C}{\pi x_L}\right]\cos^2(\pi - \alpha_{TCSC})[\omega\tan\{\omega(\pi - \alpha_{TCSC})\} - \tan(\pi - \alpha_{TCSC})] \quad (4)$$

$$\text{Where } \omega = \sqrt{\frac{x_C}{x_L}} \quad (5)$$

The value of x_L and x_C are use to designate the fundamental harmonic reactance [4] [6] of TCSC's inductor and capacitor.

So that Net admittance of TCSC will be given as

$$y_{TCSC} = \frac{1}{z_{TCSC}} = g_{TCSC} + jb_{TCSC} \quad (6)$$

Where TCSC's conductance and susceptance are g_{TCSC} and b_{TCSC}, respectively and here g_{TCSC} are Zero.

The flow of Real Power and Reactive Power [3] of power system within two buses are derived as

$$P_i = v_i \sum_{j=1}^{nb} v_j(g_{busij}cos\delta_{ij} + b_{busij}sin\delta_{ij}) \quad (7)$$

$$Q_i = v_i \sum_{j=1}^{nb} v_j(g_{busij}sin\delta_{ij} - b_{busij}cos\delta_{ij}) \quad (8)$$

The linearised equation of newton's law [9] for the compensated transmission lines are specified as in equation. Where reactance of TCSC is taken as the State Variable parameter [4] for matrix

$$\begin{bmatrix} \Delta P_i \\ \Delta P_j \\ \Delta Q_i \\ \Delta Q_j \\ \Delta P_{ij} \end{bmatrix} = \begin{bmatrix} \frac{\partial P_i}{\partial \delta_i} & \frac{\partial P_i}{\partial \delta_j} & \frac{\partial P_i}{\partial V_i} & \frac{\partial P_i}{\partial V_j} & \frac{\partial P_i}{\partial X_C} \\ \frac{\partial P_j}{\partial \delta_i} & \frac{\partial P_j}{\partial \delta_j} & \frac{\partial P_j}{\partial V_i} & \frac{\partial P_j}{\partial V_j} & \frac{\partial P_j}{\partial X_C} \\ \frac{\partial Q_i}{\partial \delta_i} & \frac{\partial Q_i}{\partial \delta_j} & \frac{\partial Q_i}{\partial V_i} & \frac{\partial Q_i}{\partial V_j} & \frac{\partial Q_i}{\partial X_C} \\ \frac{\partial Q_j}{\partial \delta_i} & \frac{\partial Q_j}{\partial \delta_j} & \frac{\partial Q_j}{\partial V_i} & \frac{\partial Q_j}{\partial V_j} & \frac{\partial Q_j}{\partial X_C} \\ \frac{\partial P_{ij}}{\partial \delta_i} & \frac{\partial P_{ij}}{\partial \delta_j} & \frac{\partial P_{ij}}{\partial V_i} & \frac{\partial P_{ij}}{\partial V_j} & \frac{\partial P_{ij}}{\partial X_C} \end{bmatrix} \begin{bmatrix} \Delta\delta_i \\ \Delta\delta_j \\ \Delta V_i \\ \Delta V_j \\ \Delta X_C \end{bmatrix} \quad (9)$$

Where $\Delta P_{ij} = P_{ijreg} - P_{ijcal.}$, is the inequality in flow of real power in the series reactance.

$\Delta X_C = \Delta X_C^i - \Delta X_C^{i-1}$, is the incremental alteration in series reactance .

The state variable of TCSC i.e. X_C , of the series controller, is updated at the ending of every iterative step till the end of an iteration

$$\Delta X_C^i = \Delta X_C^{i-1} + \left(\frac{\Delta X_C}{X_C}\right)^i \Delta X_C^{i-1} \quad (10)$$

The load flow program implemented on Matlab/Simulation as the changes takes place in Xc.

B. Static Synchronous Compensator (STATCOM)

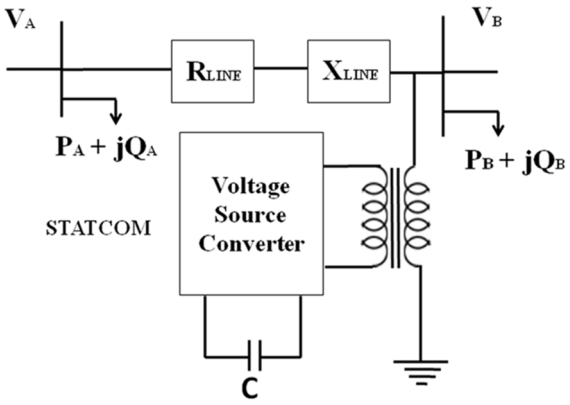


Fig.2. STATCOM

The Static Synchronous Compensator is known as STATCOM, is a power electronic controllable device. The Statcom is a Shunt FACTS Device [5] and sometime used as a voltage source converter with the help of capacitor. So that it may help in controlling of system power factor, voltage profile, steady state stability [5], flow of Power either Real or Reactive Power [7], System Reliability, congestion mitigation of transmission lines, etc. The STATCOM is provide a bidirectional flow of power among a transmission line. The voltage of STATCOM may be represented by $V_{st}\angle\theta_{st}$ and connected with a Bus Voltage of $V_B\angle\theta_B$. Practically STATCOM [8] [10] consists of the busbar, transformer, shunt impedance and system grounding. The devices are signify by:-

$$V_B\angle\theta_B = V_A\angle\theta_A - (R_{LINE} + j\,X_{LINE})I_L\angle\alpha_A - (R + jX)I_{St}\angle(\alpha_{St} + \frac{\pi}{2}) \quad (11)$$

$$E_{St} = V_{St}\angle\delta_{St} = V_{St}\,(\,Cos\,\delta_{St} + j\,Sin\,\delta_{St}\,) \quad (12)$$

$$Y_{St} = 1/(\,Z_{St}\angle\theta_{St}) = G_{St} + j\,B_{St} \quad (13)$$

So that

$$S_{St} = P_{St} + j\,Q_{St} = I_{St}^*\,V_{St} = V_{St}\,Y_{St}^*\,(\,V_{St}^* - V_B^*\,) \quad (14)$$

The following equations are obtained after the realisation of above equations into real & imaginary parts at bus 'a' is as follows:-

$$P_{St} = V_{St}^2\,G_{St} + V_{St}\,V_B[\,G_{St}\,Cos\,(\delta_{St} - \theta_B) + B_{St}\,Sin\,(\delta_{St} - \theta_B)] \quad (15)$$

$$Q_{St} = -V_{St}^2\,G_{St} + V_{St}\,V_B[\,G_{St}\,Cos\,(\,\theta_B - \delta_{St}) + B_{St}\,Sin\,(\theta_B - \delta_{St})] \quad (16)$$

$$P_B = V_B^2 G_{St} + V_{St} V_B [G_{St} \cos(\delta_{St} - \theta_B) + B_{St} \sin(\delta_{St} - \theta_B)] \quad (17)$$

$$Q_B = -V_B^2 G_{St} + V_{St} V_B [G_{St} \cos(\theta_B - \delta_{St}) + B_{St} \sin(\theta_B - \delta_{St})] \quad (18)$$

$$\begin{bmatrix} \Delta P_B \\ \Delta Q_B \\ \Delta P_{St} \\ \Delta Q_{St} \end{bmatrix} = \begin{bmatrix} \frac{\partial P_B}{\partial \theta_B} & \frac{\partial P_B}{\partial V_B} V_B & \frac{\partial P_B}{\partial \delta_{St}} & \frac{\partial P_B}{\partial V_{St}} V_{St} \\ \frac{\partial Q_B}{\partial \theta_B} & \frac{\partial Q_B}{\partial V_B} V_B & \frac{\partial Q_B}{\partial \delta_{St}} & \frac{\partial Q_B}{\partial V_{St}} V_{St} \\ \frac{\partial P_{St}}{\partial \theta_B} & \frac{\partial P_{St}}{\partial V_B} V_B & \frac{\partial P_{St}}{\partial \delta_{St}} & \frac{\partial P_{St}}{\partial V_{St}} V_{St} \\ \frac{\partial Q_{St}}{\partial \theta_B} & \frac{\partial Q_{St}}{\partial V_B} V_B & \frac{\partial Q_{St}}{\partial \delta_{St}} & \frac{\partial Q_{St}}{\partial V_{St}} V_{St} \end{bmatrix} \begin{bmatrix} \Delta \theta_B \\ \frac{\Delta V_B}{V_B} \\ \Delta \delta_{St} \\ \frac{\Delta V_{St}}{V_{St}} \end{bmatrix} \quad (19)$$

The normal operation parameter of IEEE 14 Bus system [11] obtained, where system operate without any additional device implementation for enhancement of performance of power system [7]. The parameter obtained related to Voltage magnitude, Real Power and Reactive Power on each buses of a system as per-unit basis, are as in Table 1:-

TABLE 1. BASE VALUES OF IEEE 14 BUS SYSTEM

Bus No.	Bus Voltage (V)(pu)	P (W) (pu)	Q (VAR) (pu)
1	1.7012	0.0123	0.0803
2	1.4328	- 0.015	0.0153
3	1.1047	0.0141	0.0057
4	1.0110	- 0.002	- 0.005
5	0.9979	- 0.049	- 0.044
6	1.3807	0.0096	0.0734
7	1.0881	- 0.001	- 0.003
8	1.6916	- 0.002	- 0.032
9	0.9584	0.0172	0.0114
10	0.9464	0.0000	0.0000
11	0.9696	- 0.003	0.006
12	0.9696	0.002	0.0067
13	0.9659	0.0127	0.0165
14	0.9439	- 0.000	0.001

III. SIMULATION RESULT AND DISCUSSION

The implementation of TCSC and STATCOM performed on IEEE 14 Bus System [7] as shown in Figure 3. The table 1 indicate Voltage profile and Reactive Power of power system during normal loading conditions. The parameter on buses indicate that the operational values of parameters are lying below its specific value and need some arrangement to strengthen these parameters. The congestion developed on such buses are mitigating with the help of dynamic controller in place of a fixed controller. So for that purpose, this paper proposed FACTS devices to mitigate such problem of parameter fall below a specific values of power system.

The STATCOM and TCSC are implemented on such specific buses where congestion shows by the system. The implementation of an individual device on such buses and evaluate their impact on the system. The Table 1 shows that the voltage profile of System will be 0.9696 pu, 0.9659 pu & 0.943873 pu on Bus 12, 13 & 14 respectively. So such buses are working on under loaded conditions and needed a device to upgrade the effective parameters of buses during normal as well as loading conditions.

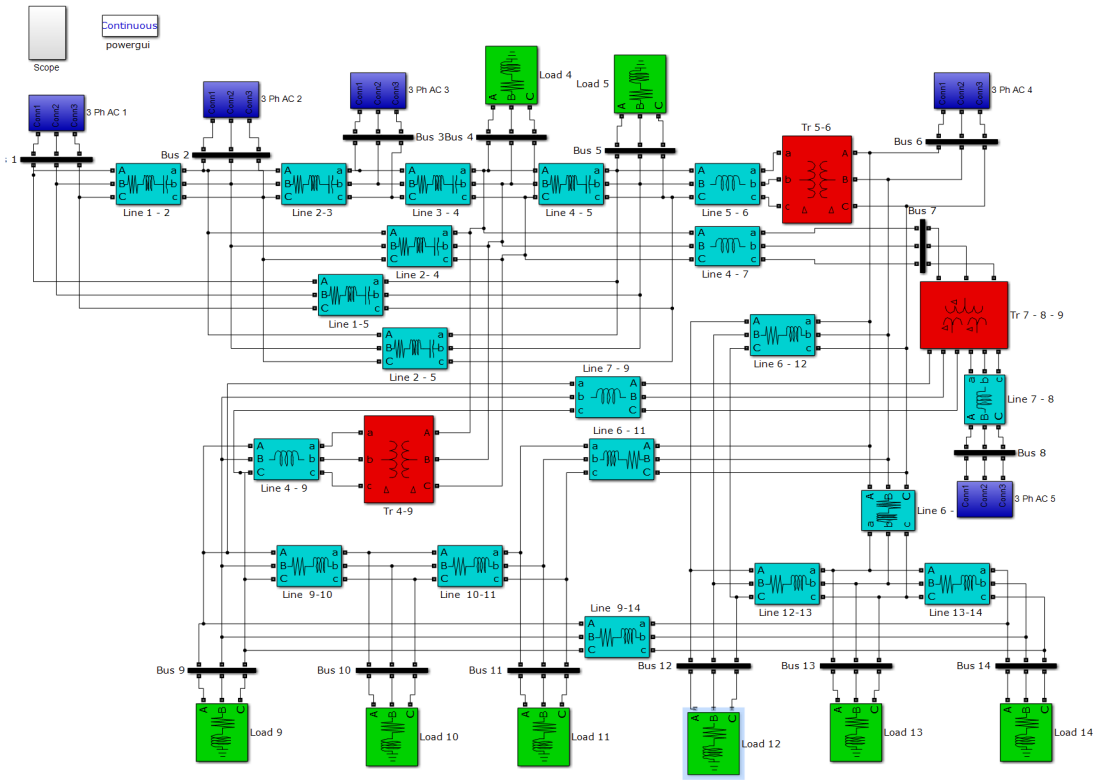


Fig.3. IEEE 14 BUS SYSTEM

The implementation of TCSC at 12 – 13 will improve it to 1.0 pu & 1.0495 pu and the STATCOM improved to 1.15 pu & 1.049 pu. The FACTS device at 13 – 14 will tend to 1.0 pu & 1.28 pu by TCSC and 1.016 pu & 1.237 pu by STATCOM. The effect of FACTS device on Voltage profile is shown in Table 2 & 3 while their graphical view are shown in Figure 4 - 9.

TABLE 2. EFFECT OF TCSC IMPLEMENTATION

Bus No.	Voltage with TCSC		
	Without	12-13	13-14
1	1.70E+00	1.70E+00	1.70E+00
2	1.43E+00	1.43E+00	1.43E+00
3	1.10E+00	1.10E+00	1.10E+00
4	1.01E+00	1.01E+00	1.01E+00
5	9.98E-01	9.98E-01	9.97E-01
6	1.38E+00	1.38E+00	1.38E+00
7	1.09E+00	1.09E+00	1.10E+00
8	1.69E+00	1.69E+00	1.69E+00
9	9.58E-01	9.65E-01	1.02E+00
10	9.46E-01	9.51E-01	9.88E-01
11	9.70E-01	9.71E-01	9.78E-01
12	9.70E-01	1.00E+00	9.80E-01
13	9.66E-01	1.05E+00	1.00E+00
14	9.44E-01	9.85E-01	1.28E+00

TABLE 3. EFFECT OF STATCOM IMPLEMENTATION

Bus No.	Voltage with STATCOM		
	Without	12-13	13-14
1	1.70E+00	1.70E+00	1.70E+00
2	1.43E+00	1.43E+00	1.43E+00
3	1.10E+00	1.10E+00	1.10E+00
4	1.01E+00	1.01E+00	1.01E+00
5	9.98E-01	9.98E-01	9.99E-01
6	1.38E+00	1.38E+00	1.38E+00
7	1.09E+00	1.09E+00	1.10E+00
8	1.69E+00	1.69E+00	1.69E+00
9	9.58E-01	9.62E-01	1.00E+00
10	9.46E-01	9.51E-01	9.81E-01
11	9.70E-01	9.73E-01	9.79E-01
12	9.70E-01	1.15E+00	1.02E+00
13	9.66E-01	1.05E+00	1.07E+00
14	9.44E-01	9.75E-01	1.24E+00

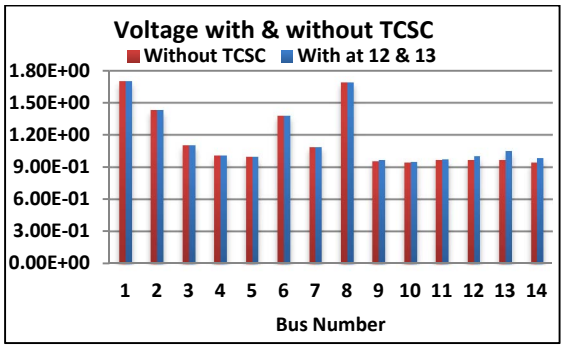


Fig.4. Performance of Voltage with TCSC at 12 & 13

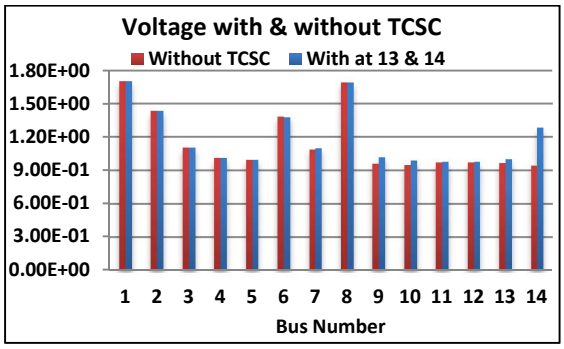


Fig.5. Performance of Voltage with TCSC at 13 & 14

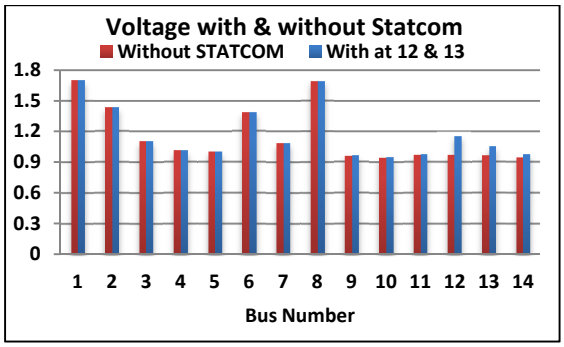


Fig.6. Performance of Voltage with STAT at 12 & 13

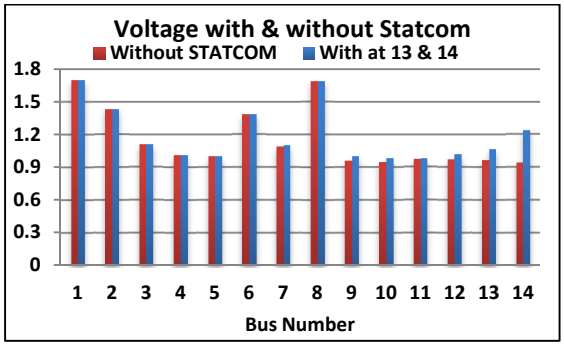
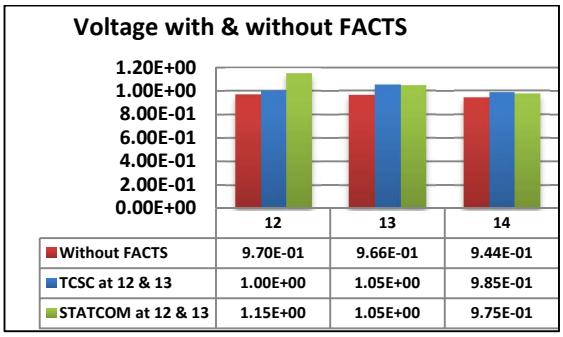
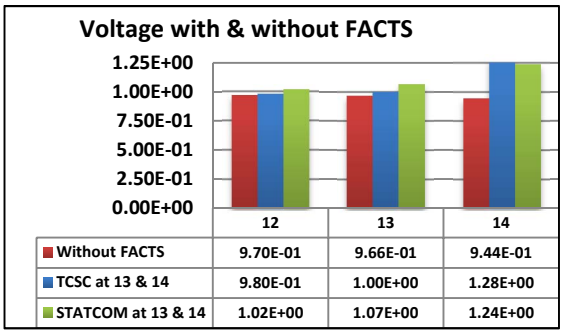


Fig.7. Performance of Voltage with STAT at 13 & 14



	12	13	14
Without FACTS	9.70E-01	9.66E-01	9.44E-01
TCSC at 12 & 13	1.00E+00	1.05E+00	9.85E-01
STATCOM at 12 & 13	1.15E+00	1.05E+00	9.75E-01

Fig.8. Performance of Voltage with FACTS at 12 & 13



	12	13	14
Without FACTS	9.70E-01	9.66E-01	9.44E-01
TCSC at 13 & 14	9.80E-01	1.00E+00	1.28E+00
STATCOM at 13 & 14	1.02E+00	1.07E+00	1.24E+00

Fig.9. Performance of Voltage with FACTS at 13 & 14

The STATCOM and TCSC are implemented on specific buses where overloading shows by a power system i.e. Table 1. The implementation of an individual device on such buses and evaluate their performance on the IEEE 14 Bus system. The Table 1 shows that the Reactive power profile of System will be 6.71e-03 pu, 1.65e-02 pu & 3.92e-04 pu on Bus 12, 13 & 14 respectively. So such buses are showing that they are working on under loaded conditions and needed a device to improve the performance of buses near to normal conditions. The implementation of TCSC at 12 – 13 will improve it to 7.16e-03 pu & 3.26e-02 pu and STATCOM improved to 1.23e-02 pu & 2.36e-02 pu. The FACTS device at 13 – 14 will tend to 1.78e-02 pu & 5.11e-04 pu by TCSC and 2.39e-02 pu & 5.05e-03 pu by STATCOM. The effect of FACTS devices on Reactive power profile is shown in Table 4 & 5 while their graphical view are shown in Figure. 10 – 15.

TABLE 4. EFFECT OF TCSC IMPLEMENTATION

Bus No.	Reactive Power with TCSC		
	Without	12-13	13-14
1	8.03E-02	8.03E-02	8.05E-02
2	1.53E-02	1.54E-02	1.58E-02
3	5.69E-03	5.75E-03	5.95E-03
4	-5.10E-03	-5.10E-03	-5.12E-03
5	-4.47E-02	-4.47E-02	-4.47E-02
6	7.34E-02	7.45E-02	7.53E-02
7	-2.66E-03	-2.60E-03	-2.49E-03
8	-3.22E-02	-3.24E-02	-3.34E-02
9	1.14E-02	1.16E-02	1.30E-02
10	8.56E-06	8.64E-06	9.34E-06
11	5.70E-03	5.71E-03	5.80E-03
12	6.71E-03	7.16E-03	6.86E-03
13	1.65E-02	3.26E-02	1.78E-02
14	3.92E-04	4.27E-04	5.11E-04

TABLE 5. EFFECT OF STATCOM IMPLEMENTATION

Bus No.	Reactive Power with STATCOM		
	Without	12-13	13-14
1	8.03E-02	8.04E-02	8.05E-02
2	1.53E-02	1.55E-02	1.58E-02
3	5.69E-03	5.77E-03	5.94E-03
4	-5.10E-03	-5.08E-03	-5.03E-03
5	-4.47E-02	-4.45E-02	-4.43E-02
6	7.34E-02	8.37E-02	8.09E-02
7	-2.66E-03	-2.53E-03	-2.06E-03
8	-3.22E-02	-3.21E-02	-3.21E-02
9	1.14E-02	1.17E-02	1.34E-02
10	8.56E-06	8.74E-06	9.66E-06
11	5.70E-03	5.74E-03	5.87E-03
12	6.71E-03	1.23E-02	7.57E-03
13	1.65E-02	2.36E-02	2.39E-02
14	3.92E-04	4.58E-04	5.05E-03

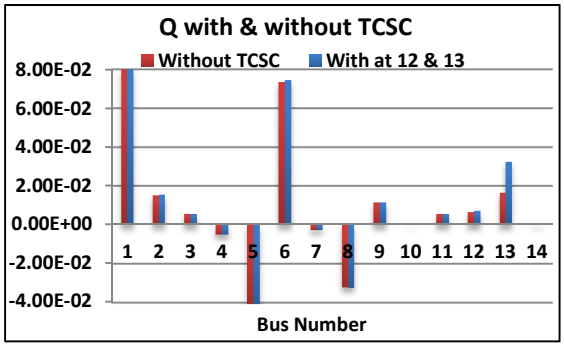


Fig.10. Performance of Q with TCSC at 12 & 13

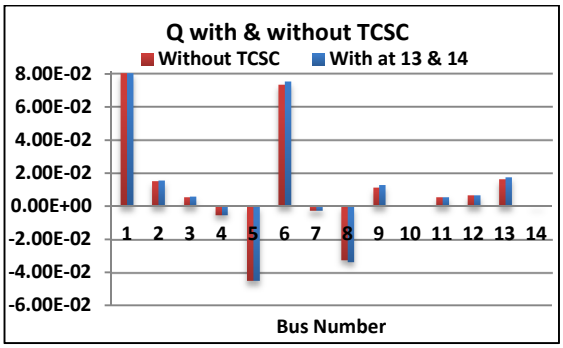


Fig.11. Performance of Q with TCSC at 13 & 14

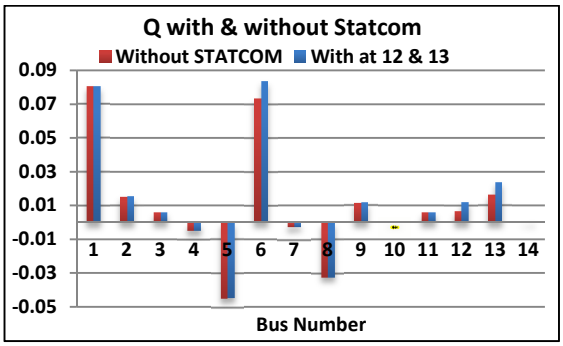


Fig.12. Performance of Q with STAT at 12 & 13

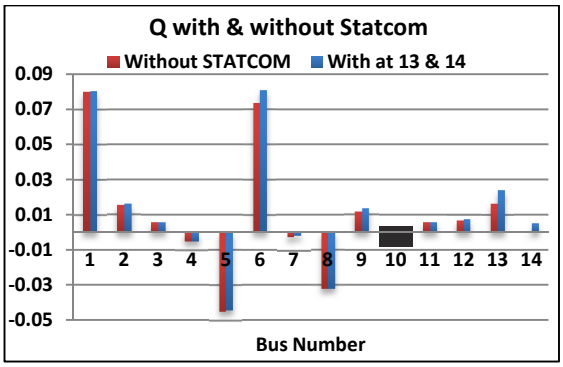


Fig.13. Performance of Q with STAT at 13 & 14

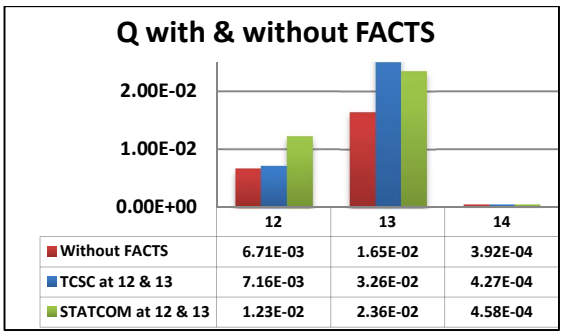


Fig.14. Performance of Q with FACTS at 12 & 13

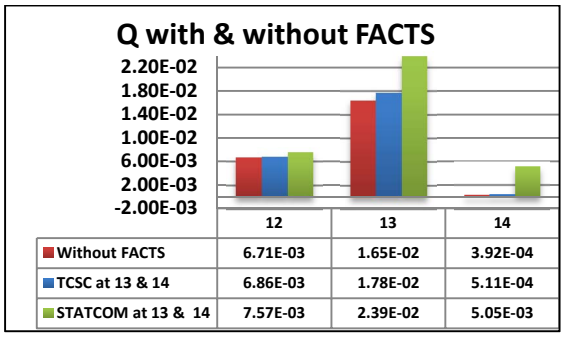


Fig.15. Performance of Q with FACTS at 13 & 14

The Figure 8, 9, 14 and 15 are shows the impact of FACTS devices on the performance of Power System under loading conditions. The Parameter variations are identified on a comparison of Table 1 with Table 2, 3, 4 & 5 with FACTS devices allocation on a IEEE 14 Bus system.

IV. CONCLUSION

In this paper, the power system congestion is discussed with the implementation of FACTS devices. The study is carried out to identified the location and the effect of Series connected FACTS device i.e. TCSC and Shunt connected FACTS device i.e. STATCOM on a power system. The parameter studied are Voltage Profile and flow of Reactive Power on IEEE 14 Bus System. The FACTS device are dynamic in nature and fast to compensate the parameter as per their requirement for the system. They are efficient in delivering or absorbing of either the reactive power of system or system voltage of bus. The main objective of the operation is to maintain all the relevant parameters to a specific value and system become more stable during loading conditions. The result obtained before and after implementation of FACTS Devices will show that entire system are more stable and reliable after FACTS devices implementations. The STATCOM is more effective device as compare with the TCSC implementation on the same type of system.

REFERENCES

[1] Abdelaziz A. Y., Sharkawy M. A. El, Attia M. A., Saadany E. F. El, “Optimal location of series FACTS to improve the performance of power system with wind penetration”, 978-1-4799-6415-4/14/$31.00 ©2014 IEEE, pp. 1 – 5.

[2] Gad M, Shinde P, Kulkarni SU (2012) : Optimal location of TCSC by sensitivity methods. International Journal on Computer Engineering Resources Vol. 2, Issue 6, pp : 162–168.

[3] Ghaeth Fandi, Zdenek Muller, Libor Straka, Jan Svec, “FACTS Devices Influence on Power Losses in Transmission Systems”, 978-1-4799-3807-0/14/$31.00 ©2014 IEEE, pp. 1 – 5.

[4] Hingorani NG, Gyugyi L (2012) : Understanding FACTS. IEEE Press, New York.

[5] Jain Vipin and Kumar Narendra (2015): 'Designing of supplementary controller for STATCOM for Mitigation of Oscillations in Power Systems', Journal of Engineering, Science & Management Education (Published by: National Institute of Technical Teachers Training & Research, Bhopal & funded by: Ministry of Human Resource Development, Govt. of India), August 2015,Vol.8, Issue 2, pp : 124-133.

[6] Kundur, P.(1994): 'Power System Stability And Control' (McGraw-Hill Press, New York, 1994).

[7] Kumar Jitender, Rai J.N. and Hasan Naimul, “Use of Phasor Measurement Unit (PMU) for Large Scale Power System State Estimation”, 2012 IEEE Fifth India International Conference on Power Electronics, IICPE2012, Delhi Technological University, New Delhi, India, ISSN : 2160-3162, PRINT ISBN : 978-1-4673-0931-8, Dec 6 - 8, 2012, 978-1-4673-0934-9/12/$31.00 ©2012 IEEE, pp.1 – 5.

[8] Malathy P. and Shunmugalatha A. : 'Loadability Analysis During Single Contingency With FACTS Devices Using Differential Evolution', IEEE Transactions on Power Systems, 2016, 978-1-5090-0128-6/16/$31.00 ©2016. pp : 1 - 11.

[9] Manganuril Yogasree, Choudekar Pallavi, Abhishek, Asija Divya and Ruchira, “Optimal Location of TCSC using Sensitivity and Stability Indices for Reduction in Losses and Improving the V oltage Profile”, 1st IEEE International Conference on Power Electronics. Intelligent Control and Energy Systems (ICPEICES-2016), 978-1-4673-8587-9/16/$31.00 ©2016 IEEE, pp. 1 – 4.

[10] Sharma P.R., Kumar Ashok and Kumar Narendra (2007): 'Optimal Location of Shunt Connected FACTS Devices in a Series compensated Long Transmission Line', Turk. Journal of Electrical Engineering, 2007, Vol.15, No.3, pp : 321-328.

[11] Taher SA, Besharat H (2008) : Transmission congestion management by determining optimal location of FACTS devices in deregulated power systems. American Journal Application Science Vol. 5, Issue 3, pp : 242–247.

View publication stats

Acta Physiologiae Plantarum (2018) 40:190
https://doi.org/10.1007/s11738-018-2766-9

ORIGINAL ARTICLE

Effect of various culture conditions on shoot multiplication and GC–MS analysis of *Plumbago zeylanica* accessions for plumbagin production

Arpita Roy[1] · **Navneeta Bharadvaja**[1]

Received: 9 March 2018 / Revised: 14 September 2018 / Accepted: 9 October 2018
© Franciszek Górski Institute of Plant Physiology, Polish Academy of Sciences, Kraków 2018

Abstract

Plumbago zeylanica, a pharmaceutically important medicinal plant, contains a wide range of phytocompounds. Culture parameters like carbon source, nitrogen source, and culture media are essential for the development and growth of explants. In this investigation, the influence of various carbon sources (sucrose, glucose, and fructose at 3% concentration), nitrogen source (ammonium nitrate, sodium nitrate, and potassium nitrate) and plant tissue culture media (MS medium, Gamborg's B5 medium, White medium and Nitsch medium) on shoot multiplication of five different accessions was studied. Optimum growth of all five accessions was observed in MS media containing 3% sucrose and ammonium nitrate as a source of carbon and nitrogen. Out of five accessions, IC-524441 showed the highest shoot multiplication. Further, methanolic extracts of all accessions (grown in MS media containing 3% sucrose and ammonium nitrate as nitrogen source) were prepared and comparison of extracts in DPPH assay indicated that accession number IC-524441 was the most effective free radical scavenging agent. Total phenolic, flavonoid and tannin content ranges were from 20 to 70 µg/ml, 40 to 100 µg/ml and 55 to 120 µg/ml, respectively, and the highest amount was found in accession number IC-524441. Sucrose and ammonium nitrate content may be responsible for increased antioxidant activity, flavonoids content, phenolic content, and tannin content in accession number IC-524441. GC–MS of ethyl acetate extract of all five accessions of *P. zeylanica* was conducted (grown in MS media containing 3% sucrose and ammonium nitrate as nitrogen source). GC–MS analysis of the aerial part showed the presence of various phytocompounds, which include 1,4-naphthalenedione, 3-eicosene, 5-eicosene, phthalic acid, *o*-anisic acid, thioctic acid, 1-octadecene, 5-*t*-butyl-cycloheptene, 2-benzoyl-1,2-dihydro-1-isoquinolinecarbonitrile, octadecanal, silane, 3-methoxy-2-methyl-2-(1-phenyl-ethylamino)-propionic acid, and 1-nonadecene. Accession number IC-524441 contains the highest amount of plumbagin, i.e. 14.19 ± 0.5 µg/ml as compared to the others.

Keywords Accessions · Carbon source · Antioxidant activity · Phenolic content · GC–MS analysis

Abbreviations

ANOVA	Analysis of variance
BAP	6-Benzylaminopurine
DPPH	2,2-Diphenyl-1-picrylhydrazyl
F–C	Folin–Ciocalteu
LOD	Limit of detection
LOQ	Limit of quantification
MS	Murashige and Skoog medium
NBPGR	National Bureau of Plant Genetic Resources
SD	Standard deviation

Communicated by P. Wojtaszek.

✉ Navneeta Bharadvaja
navneetab@dce.ac.in

[1] Plant Biotechnology Laboratory, Department of Biotechnology, Delhi Technological University, New Delhi 42, India

Introduction

Plumbago zeylanica (Chitrak) is a perennial herbaceous plant which belongs to *Plumbaginaceae* family and found in Bengal and Southern India region (Sharma and Singh 2015). It is widely utilized for the treatment of skin diseases, diarrhoea, piles, rheumatism and leprosy (Roy and Bharadvaja 2017). It contains a wide variety of secondary metabolites like steroids, glycosides, saponins, flavonoids, triterpenoids, alkaloids, coumarins, tannins, phenolic compounds, carbohydrates, fats, proteins, and naphthoquinones (Ming et al. 2011) which are responsible for its medicinal

Published online: 19 October 2018

activities. It possesses antibacterial, antifungal (Uma Devi et al. 1999) anticancer (Roy et al. 2018), antiplasmodal, antitumour, stimulatory activity, anti-inflammatory, antihyperglycaemic, central nervous system and hepatoprotective activity (Kumar et al. 2009). Previously identified and isolated compounds from this plant include plumbagin, chitranone, elliptinone, seselin, suberosin, etc (Lin et al. 2003; Jetty et al. 2010; Roy and Bharadvaja 2017). Due to its wide range of medicinal activities, this plant is overexploited. Medicines currently available in market from this plant are Chitrak Capsules (Dr Wakde's Organic), Chitrak Haritaki (Dabur) and Chitrak powder (Herbal hills) (Chitrak products source 2018). Conventional method of propagation using seed is not a reliable approach because of poor seed setting and germination. In vitro propagation provides an alternative method of conservation and rapid growth of elite genotype which is independent of seasonal changes as well as production of virus-free plant species. The advantages of in vitro propagation are higher growth and multiplication rate. The growth and multiplication of shoots are influenced by various parameters (Israeli et al. 1996), one of which is a carbon source addition in culture media which is necessary for the growth of the plants, because the autotrophic ability of plants is limited in in vitro conditions. Addition of carbon source provides energy and also maintains the osmotic potential of cells (Yaseen et al. 2013). Different media and variation in nitrogen sources are also important parameters that affect growth and the biochemical composition of tissue-cultured plants. Till now, there have been no studies regarding the effect of various carbon sources, nitrogen source and media in different accessions of *P. zeylanica*. However, the influence of various sources of carbon (glucose, sucrose and maltose) at various concentrations (1%, 3%, 5%, 7%, and 9%) on the growth of hairy root of *Arnica montana* L. has been reported by Petrova et al. (2015). Similarly, Sridhar and Naidu (2011) also reported the impact of fructose, glucose, maltose and sucrose on *Solanum nigrum* growth. Mohamed (2011) reported the effect of different nitrogen sources (Nitrophoska, urea, ammonium sulphate nitrate and ammonium sulphate) in *Zea mays*. Roy et al. (2016) reported the influence of various media (MS, B5 and Nitsch) on the multiplication of shoot in *Centella asiatica*. Variation in carbon source also influences the amount of phytochemicals present in different accessions (Bruni and Sacchetti 2009; Petrova et al. 2015). The quantity of pharmaceutically important compound depends upon the geographical conditions of that accession (Bruni and Sacchetti 2009; Kundu et al. 2016). Identification of potential accessions of medicinal plants poses a significant challenge, because various factors such as genetic composition of the plant, climate variations, age of the plant or harvesting period and specific part of plants harvested for processing influence the phytocompounds (Bruni and Sacchetti 2009). The objective of this study is to investigate a suitable carbon source, nitrogen source and culture media that promote shoot multiplication and growth of different accessions. Variation of phytochemical (antioxidant activity, total phenolic, flavonoid and tannin contents) amount among accessions was also investigated. Further, ethyl acetate leaf extracts of all accessions were analysed using GC–MS (gas chromatography–mass spectrometry). This study helps to identify the best source of carbon, nitrogen and culture media for the highest growth and presence of various compounds in all the five accessions.

Materials and methods

Collection of plant material

Plumbago zeylanica accessions—IC-421418, IC-524441, IC-439212, IC-398891 and IC-539866—were collected from NBPGR, New Delhi. These accessions were further maintained in MS media (0.2 mg/l BAP) in our Plant Tissue Culture Laboratory, DTU.

Culture medium and parameters

For shoot culture establishment, semisolid MS basal medium ($CaCl_2 \cdot 2H_2O$, $CuSO_4 \cdot 5H_2O$, $CoCl_2 \cdot 6H_2O$, KNO_3, KH_2PO_4, $MgSO_4 \cdot 7H_2O$, H_3BO_3, KI, $Na_2MoO_4 \cdot 2H_2O$, glycine, $FeSO_4 \cdot 7H_2O$, Na_2EDTA, NH_4NO_3, $MnSO_4 \cdot H_2O$, $ZnSO_4 \cdot 7H_2O$, nicotinic acid, meso-inosito, pyridoxine hydrochloride), sucrose and agar were purchased from Himedia Pvt. Ltd., Mumbai, India. For various experiments, MS media were prepared from stock solutions (major, minor, vitamins) by varying the carbon source (sucrose, glucose and fructose at 3% concentration), nitrogen source (ammonium nitrate, sodium nitrate and potassium nitrate) and culture media (MS medium, Gamborg's B5 medium, White medium and Nitsch medium); 0.8% agar was used for all the cultures. For all the experiments, the medium was supplemented with 1 mg/l BAP. Nodal explants from each accession were aseptically transferred to media and 25 ± 2 °C was used for incubation with 16 h photoperiod using 36 W fluorescent lamps (Philips, Kolkata, India).

Statistical analysis

Results were expressed as mean ± SD with three replicates each containing three explants. The effect of various treatments on shoot multiplication was compared to detect significant differences among the treatments using ANOVA at 5% probability level.

Phytochemical studies

Methanolic extract preparation

Fresh in vitro-grown plant material MS media containing 3% sucrose and ammonium nitrate as nitrogen source was dried at room temperature and ground with mortar and pestle. 1.0 gm of the plant material was macerated in 10 ml methanol for 48 h and the plant extract was filtered and stored at 4 °C for further investigation.

Antioxidant activity

DPPH assay was done to estimate the antioxidant activity (Brand-Williams et al. 1995). Stock solutions of plant extract (1 mg/ml) were prepared in methanol and from this 50 µg/ml was used. Similarly, 1 mM DPPH stock solution was prepared in methanol and from this 0.06 mM DPPH was used. 100 µl plant extracts + 3.9 ml of DPPH solution was mixed by vortexing for 30 s and left in the dark for 30 min incubation. Methanol and DPPH solution were used as blank and negative control. Sample absorbance was determined at 517 nm.

Antioxidant activities were calculated as:

$$\text{Antioxidant activity } (\%) = [(A_{\text{control}} - A_{\text{sample}})/A_{\text{control}}] \times 100.$$

Total flavanoid estimation

Aluminium chloride ($AlCl_3$) colorimetric assay was done to estimate the flavanoids (Lamaison and Carnat 1990). 1 ml extract + 4 ml DW (distilled water) + 300 µl $NaNO_2$ (5%) was added. Five minutes later, 300 µl $AlCl_3$ (10%) + 2 ml NaOH (1M) + 2.4 ml DW was added. Different concentrations of standard quercetin (100, 80, 60, 40 and 20 µg/ml) were prepared. Sample absorbance and quercetin standards were determined using UV–visible spectrophotometer at 510 nm.

Total phenolic content estimation

Total phenolic content was estimated using the F–C method (Singleton and Rossi 1965). 200 µl plant extract + 1.5 ml F–C (10%) reagent was mixed and kept in dark for 5 min. After this, 1.5 ml $NaCO_3$ (5%) was added and mixed properly. The final mixture was kept in dark for 30 min. The calibration curve of gallic acid standard solutions (200, 100, 50 and 25 mg/l) was prepared for quantification of total phenolic content. Absorbance was recorded at 750 nm.

Total tannin estimation

The F–C method was used to determine the tannin content (Singleton et al. 1999). 100 µl plant extract + 0.5 ml F–C (10%) reagent + 1 ml $NaCO_3$ (35%) + 8.4 ml DW were added in a test tube. The entire reaction mixture was mixed well and kept for 30 min. Various concentrations of standard gallic acid (100, 80, 60, 40 and 20 µg/ml) were used. The absorbance of samples and gallic acid was recorded at 725 nm.

GC–MS analysis

Ethyl acetate extract preparation Fresh in vitro-grown plant material (MS media with 3% sucrose) was dried at room temperature and then ground with mortar and pestle. 1.0 gm of plant material was then macerated in 10 ml of ethyl acetate for 48 h. Plant extracts were filtered with 0.45-µm filter.

Plumbagin estimation and instrumentation For quantitative analysis of plumbagin, 10 mg of plumbagin standard (Sigma Aldrich) was dissolved in 10 ml of ethyl acetate, i.e. astock solution of (1 mg/ml) was prepared. From the stock solutions, further dilutions were prepared by diluting the required solution volume and 1 µl solution was injected into the system.

GC–MS analysis was conducted using an Agilent GC (7890B GC) system. The analytical column was 5MS (5% phenyl, 95% methyl siloxane) capillary column. The injection port and detector port temperature were 270 °C and 280 °C. 1 ml/min flow rate and 1:20 split was used (Rajakrishnan et al. 2017). The sample injection volume was 1 µl and direct injection mode was used. Compound identification was carried out using the database of NISTII library (McLafferty and Stauffer 1989).

Validation Validation of the method was done using parameters such as linearity, correlation coefficient (r), LOD and LOQ.

Linearity Linearity study was done by serially diluting the stock solution of the standard (1 mg/ml) to a given concentration range as 50, 100, 200, 300 and 500 µg/ml. The calibration curve was made after triplicate analysis of plumbagin standard by plotting the peak area against concentration (µg/ml). The correlation coefficient (r) was determined by calculating $\sqrt{R^2}$.

LOD is the minimum amount of analyte present in a sample which can be detected but not necessarily quantified.

LOQ is the minimum amount of analyte present in a sample which can be quantified.

LOD and LOQ studies were done to detect and quantify the limits of the method to check the presence of any impurities using the following equations:

LOD = 3.3 σ/S and LOQ = 10 σ/S, where S is the slope of the curve and σ is SD.

Results and discussion

Effect of culture conditions on shoot multiplication

Effect of different carbon sources

Carbon source is essential for in vitro culture conditions; addition of carbon source in culture medium enhances plantlet growth. In this study, 3% of different carbon sources were used. Shoot initiation started after 3 weeks of their inoculation on MS media containing sucrose as a carbon source. Comparative analysis showed that the highest shoot length (4.2 ± 0.54 cm) and maximum shoot number (5.2 ± 0.44) were observed in accession number IC-524441 when 3% sucrose was used as carbon source in media (Table 1). Accession number IC-539866 showed the minimum number of shoots (2.2 ± 0.44) and minimum length of shoot (3.0 ± 0.29 cm). Glucose was the second most productive carbon source for multiplication of shoots. At 3% glucose level, accession number IC-524441 showed maximum shoot number (3.2 ± 0.44) and shoot length (3.4 ± 0.27 cm) (Table 1). Fructose showed least growth in all accessions. The requirement of carbon source depends on the developmental stage of the culture and there can be variations based on carbon source type (Thompson and Thorpe 1987). In in vitro conditions, plants are not completely autotrophic in nature and, therefore, there is requirement of carbon source in culture media which supports proliferation of shoots, induction of roots and overall plant growth. Various types of carbon sources, i.e. non-reducing and reducing sugars, are utilized in the medium. The morphogenetic potential of in vitro cultures can be influenced by manipulating the carbon source type and concentration. Sucrose is a most widely utilized source of carbon and energy; it is a disaccharide, highly water-soluble substance which works as a molecule transporter and is able to transport through the plasma membrane (Petrova et al. 2015). A study reported that media containing sucrose generated more number of leaves as compared to media containing fructose and glucose in banana plants (Buah et al. 2000). They also recorded that at higher temperature, fructose released a toxic substance, i.e. 5-hydroxymethyl-2-furaldehyde which increased hyperhydricity as well as declined the water potential in leaves, causing leaf enlargement. Gubis et al. (2005) also reported that media containing 30 g/l of sucrose generated healthy plantlets in *Lycopersicon esculentum* as compared to other types and concentrations of carbon source. Baskaran and Jayabalan (2005) reported that the relative growth rate of plantlet was best found in medium containing 30 g/l of glucose in *Eclipta alba*. Sridhar and Naidu (2011) reported that the highest shoot number was obtained at 4% fructose and maximum shoot length at 4% sucrose in case of *S. nigrum*. Petrova et al. (2015) reported that the carbon source influences the growth of *A. montana* and optimum growth was found at 3% sucrose. Similarly, Kundu et al. (2016) also reported that 3% sucrose along with 1.5 mg/l BAP in a nutrient media gives the highest multiplication of shoots in case of *C. asiatica*.

Table 1 Effect of different carbon sources on the number and length of shoots in five different accessions of *Plumbago zeylanica* after 8 weeks of inoculation

Experiment detail	Accession number	Number of shoots (M ± SD)	Length of shoots (cm) (M ± SD)
Sucrose	524441	5.2 ± 0.44	4.2 ± 0.54
	398891	2.8 ± 0.45	3.4 ± 0.76
	439212	2.4 ± 0.54	3.2 ± 0.44
	539866	2.2 ± 0.44	3.0 ± 0.29
	421418	4.6 ± 0.54	3.5 ± 0.41
Glucose	524441	3.2 ± 0.44	3.4 ± 0.27
	398891	2.2 ± 0.45	3.2 ± 0.44
	439212	1.6 ± 0.54	2.8 ± 0.22
	539866	1.8 ± 0.45	3.3 ± 0.44
	421418	2.4 ± 0.54	3.1 ± 0.38
Fructose	524441	2.1 ± 0.45	2.7 ± 1.27
	398891	1.8 ± 0.45	2.3 ± 1.13
	439212	1.7 ± 0.54	2.5 ± 0.72
	539866	1.4 ± 0.54	3.6 ± 0.68
	421418	1.2 ± 0.45	3.3 ± 0.65

Values are expressed as mean ± standard deviation (M ± SD)

Effect of different nitrogen sources

Nitrogen is a crucial element for growth. Nitrogen availability and the form in which it is present are of great importance. Nitrate is considered to be an essential form of nitrogen for tissue culture. To identify the best source of nitrate, different sources of nitrogen, i.e. ammonium nitrate, sodium nitrate and potassium nitrate, were tested on accessions of *P. zeylanica*. Data on shoot multiplication are presented in Table 2. Shoot initiation started 3 weeks after inoculation on MS media containing ammonium nitrate as a nitrogen source. Comparative analysis showed that the highest shoot length (3.179 ± 0.21 cm) and maximum shoot number (4.1 ± 0.56) were observed in accession number IC-524441 when ammonium nitrate was added to the media. Accession number IC-539866 showed the minimum number of shoots (2.2 ± 0.42) and minimum shoot length (4.111 ± 0.39 cm). Sodium nitrate was the second most efficient nitrogen source for multiplication of shoots and potassium nitrate showed least growth in all accessions. In plant tissue culture medium, nitrate ions, ammonium salt, amino acids and complex organic compounds supply nitrogen. Nitrate ions and ammonium salts are good sources of nitrogen because they can be absorbed and metabolized easily by cells, which further leads to seed breaking, root branching and dormancy of bud and stops apical dominance. Hence, nitrogen in the form of ammonium ion and nitrate ion is a dominant mineral nutrient in most plant tissue culture formulations. Optimization of these compounds stimulates regeneration in recalcitrant cultivars (Benson 2000). Mohamed (2011) reported the effect of different nitrogen sources in *Zea mays* and found that ammonium sulphate nitrate increased the growth as compared to other nitrogen sources. Similarly, Chandravanshi et al. (2014) recorded maximum multiplication of shoots in *P. zeylanica* in MS media containing ammonium nitrate and 13.3 µM BAP and 135.74 µM adenine sulphate. Several studies have shown that tissue culture growth is possible on a medium having ammonium as the sole source of nitrogen. Since ammonium nitrate has been banned due to its explosive nature (New York explosion chelsea 2016), sodium nitrate can be utilized as a substitute in culture media to supply the nitrogen source.

Effect of different culture media

Various factors influence optimum shoot multiplication. Variations in external factors such as culture media are an important factor which may influence the multiplication of shoots (Staba 1980). The effect of various media on shoot number in accessions of *P. zeylanica* (IC-524441, 398891, 439212, 539866, and 421418) is furnished in Table 3. Plant height (3.839 ± 0.7 cm) and maximum shoot number (6.0 ± 0.47) was higher in MS media for accession number IC-524441 followed by Nitsch, B5 and White media, respectively. Accession number IC-524441 was the best among all accessions, followed by 398891, 421418, 539866 and 439212 (Table 3). Chandravanshi et al. (2014) also reported maximum shoot multiplication in MS media supplemented by 13.3 µM BAP and 135.74 µM adenine sulphate. Vijay et al. (2016) reported that 4.44 µM BAP gave maximum shoot multiplication in *P. zeylanica*. Similarly, Roy et al. (2016) reported that MS media containing 1 mg/l BAP showed the highest shoot multiplication in *C. asiatica*.

Table 2 Effect of different nitrate sources on the number and length of shoots in five different accessions of *Plumbago zeylanica* after 8 weeks of inoculation

Experiment detail	Accession number	Number of shoots (M ± SD)	Length of shoots (cm) (M ± SD)
NH_4NO_3	524441	4.1 ± 0.56	3.179 ± 0.21
	398891	3.3 ± 0.67	3.759 ± 0.41
	439212	2.3 ± 0.48	4.188 ± 0.05
	539866	2.2 ± 0.42	4.111 ± 0.39
	421418	3.4 ± 0.51	3.719 ± 0.80
$NaNO_3$	524441	3.0 ± 0.47	3.033 ± 0.36
	398891	1.8 ± 0.42	2.883 ± 0.23
	439212	2.2 ± 0.42	3.194 ± 0.13
	539866	1.9 ± 0.57	4.083 ± 0.66
	421418	2.3 ± 0.48	3.8 ± 0.78
KNO_3	524441	2.8 ± 0.42	2.744 ± 0.35
	398891	2.1 ± 0.31	2.963 ± 0.25
	439212	2.2 ± 0.42	3.35 ± 0.18
	539866	1.7 ± 0.48	3.766 ± 0.28
	421418	2.7 ± 0.48	2.633 ± 0.24

Values are expressed as mean ± standard error (M ± SE)

Table 3 Effect of different media on number and length of shoots in five different accessions of *Plumbago zeylanica* after 8 weeks of inoculation

Experiment detail	Accession number	Number of shoots (M ± SD)	Length of shoots (cm) (M ± SD)
White media	524441	3.2 ± 0.42	3.244 ± 0.108
	398891	2.9 ± 0.56	2.822 ± 0.2
	439212	2.6 ± 0.51	2.486 ± 0.22
	539866	2.0 ± 0.47	3.588 ± 0.73
	421418	3.0 ± 0.47	2.664 ± 1.37
Nitsch media	524441	3.3 ± 0.48	3.688 ± 0.55
	398891	2.5 ± 0.52	3.969 ± 0.5
	439212	2.7 ± 0.57	1.9 ± 0.18
	539866	2.1 ± 0.56	2.572 ± 0.25
	421418	3.2 ± 0.42	2.836 ± 0.14
MS media	524441	6.0 ± 0.47	3.839 ± 0.7
	398891	5.2 ± 0.42	3.341 ± 0.13
	439212	4.9 ± 0.56	2.928 ± 0.26
	539866	3.9 ± 0.56	3.236 ± 0.64
	421418	5.3 ± 0.48	3.694 ± 0.22
B5 media	524441	4.7 ± 0.48	2.922 ± 0.49
	398891	2.7 ± 0.48	3.48 ± 0.2
	439212	1.6 ± 0.51	3.933 ± 1.65
	539866	1.2 ± 0.53	2.966 ± 1.12
	421418	3.2 ± ± 0.42	2.5 ± 0.52

Values are expressed as mean ± standard error (M ± SE)

Phytochemical studies

Methanol was used for the extraction process as it shows a broad range of solubility properties for low molecular and relatively polar substances which include antioxidant active compounds.

In the present study, DPPH free radical scavenging effect of different extracts ranged from 18 to 185 µg/ml and were in the order of IC-524441 > IC-421418 > IC-398891 > IC-439219 > IC-539866 (Table 4). IC-524441 exhibited the highest scavenging activity (185 ± 0.22 µg/ml), while lowest scavenging activity was observed in IC-539866 (18 ± 0.44 µg/ml). This may be due to the presence of high polyphenolic compounds which possess scavenging free radicals, hydroxyl and superoxide radical by single electron transfer (Czapecka et al. 2005).

Total phenolic content of samples ranged from 20 to 70 µg/ml (Table 4). Among all the accessions, IC-524441 showed the highest phenolic content (70 ± 0.54 µg/ml), while the lowest was observed in IC-539866 (20 ± 0.45 µg/ml). Considerable differences between the results of phenolic content may be due to various factors associated with the accessions such as environmental factors, location and temperature. Polyphenols are important dietary antioxidants due to their radical scavenging activity. Various in vitro investigations have reported the antioxidant role of polyphenols which protects against many diseases (Matkowski et al. 2008). These compounds have a vital role in growth and reproduction, and provide protection against harmful predators and pathogens. A study reported that methanolic extract of *P. zeylanica* leaves have 28.25 ± 0.001 µg/ml of phenolic content (Sharma et al. 2014) (Fig. 1).

Table 4 Phytochemical studies in different accessions of *Plumbago zeylanica*

Accession number	Antioxidant activity (µg/ml)	Total phenolic content (µg/ml)	Total flavanoid content (µg/ml)	Total tannin content (µg/ml)
524441	185.18 ± 0.22	70 ± 0.54	100 ± 0.54	120 ± 0.55
421418	120.81 ± 0.45	65 ± 0.66	69 ± 0.5	105 ± 0.71
398891	66.67 ± 0.54	50 ± 0.45	53 ± 0.4	75 ± 0.8
439212	85.18 ± 0.54	35 ± 0.54	49 ± 0.5	69 ± 0.55
539866	18.51 ± 0.44	20 ± 0.45	40 ± 0.66	55 ± 0.8

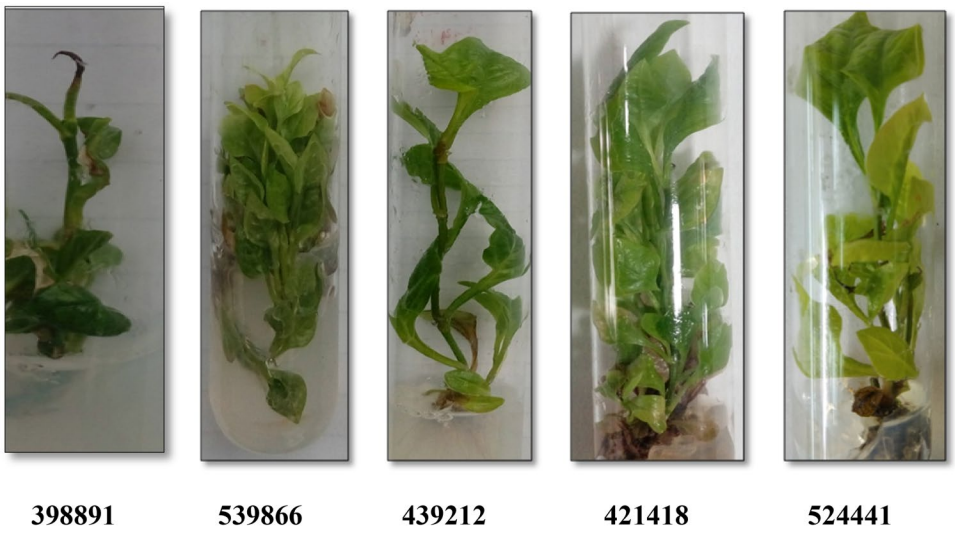


Fig. 1 In vitro culture of five different accessions of *Plumbago zeylanica* in MS medium with sucrose as carbon source and ammonium nitrate as nitrogen source

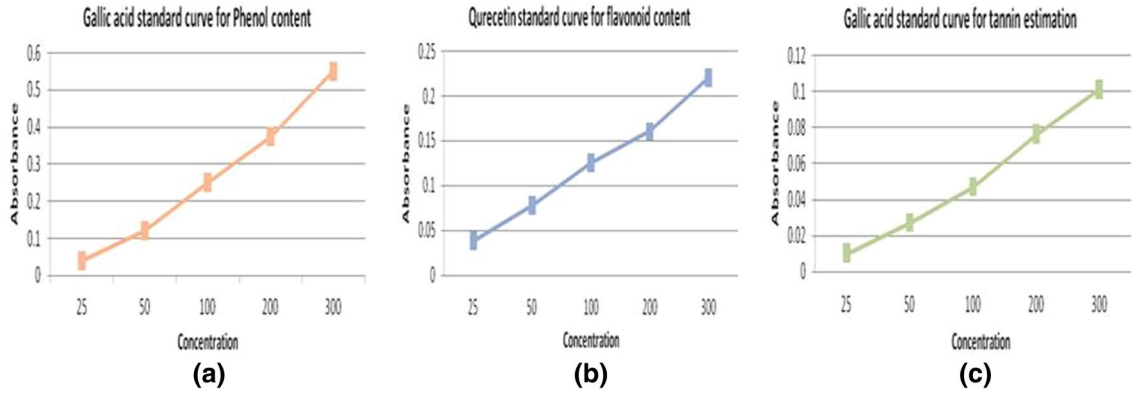


Fig. 2 **a** Standard curve of gallic acid for phenol estimation. **b** Standard curve of qurecetin for flavonoid estimation. **c** Standard curve of gallic acid for tannin estimation

The total flavonoid content varied between 40 and 100 µg/ml (Table 4; Fig. 2b). Among all the accessions, IC-524441 showed the highest amount of flavonoid content (100 ± 0.8 µg/ml), while the lowest content was observed in IC-539866 (40 ± 0.7 µg/ml). Flavonoids include flavanols, flavones and condensed tannins. Antioxidant activity of these compounds depend on the presence of free OH group, especially 3-OH. Flavonoids are polyphenolic compounds which possess in vitro antioxidant activity and act as in vivo antioxidants (Shimoi et al. 1996). Flavonoids (water-soluble compound) are potential scavenging molecules and possess strong anticancer activity (Salah et al. 1995). Flavonoids suppress reactive oxygen formation and increase antioxidant defences (Agati et al. 2012). A study reported the presence of flavonoid content in stem and leaves of *P. zeylanica*, i.e. 72.3 µg/ml and 32.433 µg/ml (Sharma et al. 2014).

Total tannin content was determined using gallic acid as standard; the maximum amount was found in accession number IC-524441, 120 ± 0.55 µg/ml, and the minimum in accession number 539866, 55 ± 0.8 µg/ml. (Table 4; Fig. 2c). Tannins are widespread in nature and probably present in all plants. These are amorphous, astringent substances and water-soluble phenolic compounds (Elgailani and Ishak 2014). Tannins show several biological activities such as protection against oxidative stress and degenerative diseases. Tannins are able to heal wounds and inflamed mucous membranes. Tannins possess high antioxidant, antimicrobial, free radical scavenging and antiulcerogenic activities; moreover in heart mitochondria, they show potent lipid peroxidation inhibition (Shad et al. 2012).

Recently, much consideration has been given to natural phytocompounds and their medical advantages. Plants are major source of antioxidants and are able to produce various secondary metabolites that have therapeutic potential. Medicinal plants contain phenolic compounds such as flavonoids, tannins and phenolic acids. The difference in antioxidant activity among accessions was due to the variation in phenolic, flavonoid and tannin content which may

Table 5 List of phytochemicals identified by GC–MS in ethyl acetate extract of the aerial part of different accessions of *P. zeylanica*

Peak	Retention time	Area percentage	Compound	Nature of compound	Molecular formula
Accession number 398891					
1	13.981	37.817	1,4-Naphthalenedione	Naphthalene-derived organic compound	$C_{10}H_6O_2$
2	16.681	5.281	*o*-Anisic acid	Carboxylic acid	$C_8H_8O_3$
3	17.039	7.612	5-Eicosene	Acyclic alkenes	$C_{20}H_{40}$
4	17.950	6.377	Cyclopropaneoctanal	Aldehyde	C_3H_6
5	18.633	9.002	Phthalic acid	Aromatic dicarboxylic acid	$C_6H_4(COOH)_2$
6	18.769	3.463	Thioctic acid	Saturated fatty acid	$C_8H_{14}O_2S_2$
7	19.381	5.479	2-Benzoyl-1,2-dihydro-1-isoquinoli necarbonitrile	–	–
8	20.396	17.713	Phthalic acid	Aromatic dicarboxylic acid	$C_6H_4(COOH)_2$
9	20.803	7.257	1-Nonadecene	Long-chain hydrocarbon and an alkene	$C_{19}H_{38}$
Accession number 539866					
1	13.974	73.804	1,4-Naphthalenedione	Naphthalene-derived organic compound	$C_{10}H_6O_2$
2	17.038	3.429	1-Octadecene	Long-chain hydrocarbon and an alkene	$C_{18}H_{36}$
3	17.950	3.885	5-*t*-Butyl-cycloheptene	Cycloalkene	$C_{11}H_{20}$
4	18.632	4.656	Phthalic acid	Aromatic dicarboxylic acid	$C_6H_4(COOH)_2$
5	18.769	1.909	Octadecanal	Long chain fatty aldehyde	$C_{18}H_{36}O$
6	19.832	2.004	3-Methoxy-2-methyl-2-(1-phenyl-ethylamino)-propionic acid	Carboxylic acid	–
7	20.396	10.273	Phthalic acid	Aromatic dicarboxylic acid	$C_6H_4(COOH)_2$
Accession number 439212					
1	13.974	39.207	1,4-Naphthalenedione	Naphthalene-derived organic compound	$C_{10}H_6O_2$
2	17.046	3.016	5-Eicosene	Acyclic alkenes	$C_{20}H_{40}$
3	17.955	2.841	Silane	Saturated chemical compounds	SiH_4
4	18.638	5.325	5-(3-Methylbutyl)-2-pyridinecarboxylic acid	Carboxylic acid	–
5	19.386	11.002	2-Benzoyl-1,2-dihydro-1-isoquinoli necarbonitrile	–	–
6	20.402	36.322	Phthalic acid	Aromatic dicarboxylic acid	$C_6H_4(COOH)_2$
7	20.810	2.286	3-Eicosene	Acyclic alkenes	$C_{20}H_{40}$
Accession number 524441					
1	14.010	74.529	1,4-Naphthalenedione	Naphthalene-derived organic compound	$C_{10}H_6O_2$
2	17.048	2.149	3-Eicosene	Acyclic alkenes	$C_{20}H_{40}$
3	17.954	2.180	Phytol	Acyclic diterpene alcohol	$C_{20}H_{40}O$
4	18.638	4.086	Di-sec-butyl phthalate 1,2-Benzenedicarboxylic acid	Esters of phthalic acid, carboxylic acid	$C_{16}H_{22}O_4$, $C_8H_6O_4$
5	19.387	3.504	Bicyclo[3.3.0]octan-3-one, glycine	Ethyl pentyl ketone , amino acid	$C_{10}H_{14}O$, $C_2H_5NO_2$
6	20.400	10.606	Phthalic acid	Aromatic dicarboxylic acid	$C_6H_4(COOH)_2$
7	20.805	2.948	5-Eicosene, 1-nonadecene	Acyclic alkenes, long-chain hydrocarbon and an alkene	$C_{20}H_{40}$, $C_{19}H_{38}$
Accession number 421418					
1	13.952	52.839	1,4-Naphthalenedione	Naphthalene-derived organic compound	$C_{10}H_6O_2$
2	16.681	2.816	*o*-Anisic acid	Carboxylic acid	$C_8H_8O_3$
3	17.043	2.916	Cetene	Oily hydrocarbon	$C_{16}H_{32}$
4	17.945	4.582	Phytol	Acyclic diterpene alcohol	$C_{20}H_{40}O$
5	18.629	7.683	Phthalic acid	Aromatic dicarboxylic acid	$C_6H_4(COOH)_2$
6	18.768	1.563	3,7,11-Trimethyl-2,4-dodecadiene	Alkadienes	$C_{15}H_{28}$
7	19.377	5.703	Bicyclo[3.3.0]octan-3-one	Ethyl pentyl ketone	$C_{10}H_{14}O$
8	20.394	19.247	Phthalic acid	Aromatic dicarboxylic acid	$C_6H_4(COOH)_2$
9	20.800	2.651	1-Eicosene	Acyclic alkenes	$C_{20}H_{40}$

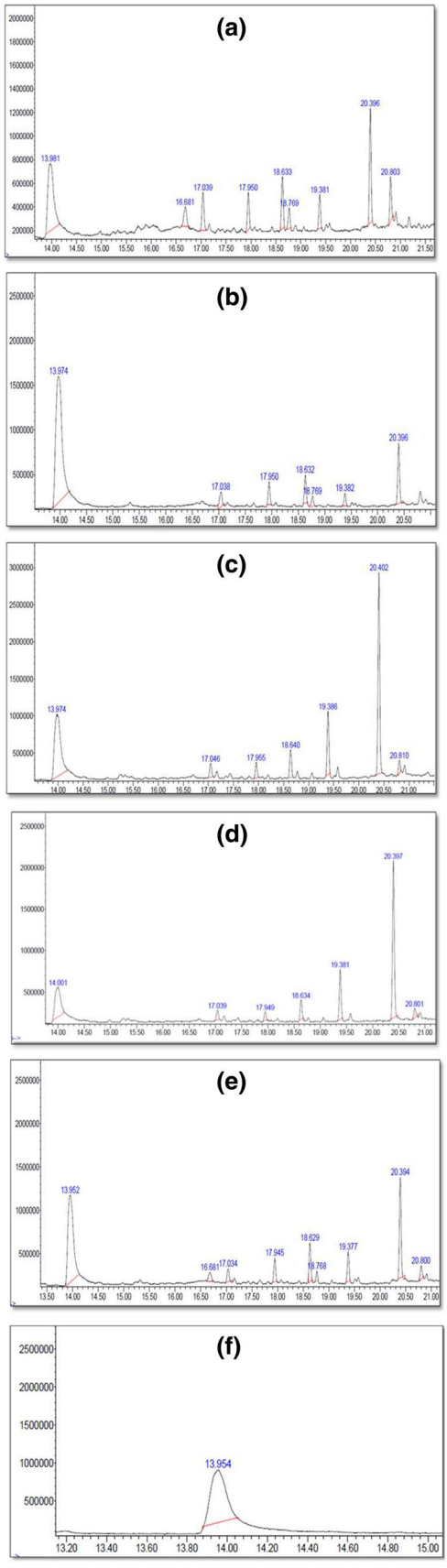


Fig. 3 GC–MS chromatogram of ethyl acetate extracts of *Plumbago zeylanica* accessions: **a** 439212, **b** 524441, **c** 398891, **d** 539866, **e** 421418 and **f** plumbagin standard

be influenced by the carbon and nitrogen source. Accession number IC-524441 showed maximum growth and possesses maximum content of total phenolic, flavonoid, tannin and antioxidant activity when grown in MS media containing sucrose and ammonium nitrate as carbon and nitrogen source. Similar observations were noticed on other four accessions. This may be because sugars induce some signals which affect metabolism, growth, development and gene expression in plants and also affect phytocompound production (Petrova et al. 2015).

GC–MS analysis

GC–MS of different accessions of *P. zeylanica* (Table 5; Fig. 3) revealed the presence of 1,4-naphthalenedione, 3-eicosene, 5-eicosene, phthalic acid, *o*-anisic acid, thioctic acid, 1-octadecene, 5-*t*-butyl-cycloheptene, 2-benzoyl-1,2-dihydro-1-isoquinoli necarbonitrile, octadecanal, silane, 3-methoxy-2-methyl-2-(1-phenyl-ethylamino)-propionic acid and 1-nonadecene. Compound identification was carried out using the NISTII library. In accession number IC-524441, 1,4-naphthalenedione was present in higher amount, whereas in accession number IC-398891, phthalic acid was in higher amount. Rajakrishnan et al. (2017) reported the presence of 1,4-naphthalenedione in roots of *P. zeylanica*. Similarly, Sharma et al. (2015) reported the presence of tetradecanoic acid, 2-pentadecanone, 1-heptadecene, tetracontane, isophytol, etc. in *P. zeylanica*.

Linearity for plumbagin was determined by plotting calibration graph of peak area against the standard concentrations (Fig. 4). Calibration curve shows that plumbagin had a range between 50 and 500 µg/ml. Linear regression equation for plumbagin is $y = 17{,}753x + 148{,}266$, $R^2 = 0.9986$, where x is standard plumbagin concentration and y is peak area. LOQ was 0.101 µg/ml and LOD was 0.033 µg/ml. A linear relationship was obtained for plumbagin concentration (50–500 µg/ml) and the correlation coefficient (r) was 0.99.

Plumbagin estimation Plumbagin concentration in the plant extracts was obtained by comparing samples with standard plumbagin (Sigma-Aldrich) solution. On GC analysis of ethyl acetate extract, chromatogram showed peak at retention time of 13.939–14.001. Results showed that accession numbers IC-524441, IC-421418, IC-398891, IC-439219 and IC-539866 contained 14.19 ± 0.5 µg/ml,

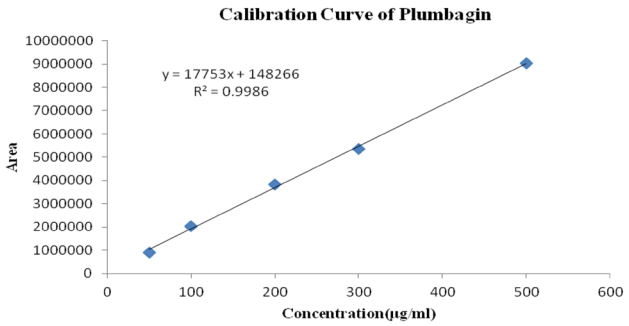


Fig. 4 Calibration curve of plumbagin

10.68 ± 0.5 µg/ml, 7.82 ± 0.5 µg/ml, 6.71 ± 0.5 µg/ml and 5.07 ± 0.5 µg/ml of plumbagin respectively (Fig. 3). The data clearly revealed that accession number IC-524441 contained the highest amount of plumbagin. This is the first report on plumbagin estimation from various in vitro-grown accessions using GC–MS.

Conclusion

Plumbagin is the most important phytocompound of *P. zeylanica*. Plant growth and production of secondary metabolite is affected by interactions between genotype and different environmental conditions. Hence, screening and selection of particular accession for increase in plumbagin production is essential. In the present study, various contributory factors like carbon, nitrogen source and plant tissue culture media were used for the assessment of the best carbon source, nitrogen source and media for higher growth. It was found that all the three parameters used in this study affected plant growth. MS media containing sucrose and ammonium nitrate promote in vitro growth of *P. zeylanica* accessions. It is interesting that total flavonoid content, phenolic content and tannin content correlate well with the results of DPPH test, and findings of this study support the fact that some medicinal plants commonly consumed in India are promising sources of potential antioxidants. The highest amount of phytochemical presence could be due to the influence of sucrose and ammonium nitrate which results in the highest shoot multiplication rate. Gas chromatography of ethyl acetate fraction revealed the presence of 7–9 peaks in five different accessions of *P. zeylanica*. Accession number IC-524441 showed the highest amount of plumbagin content. This is the first study on the effect of various culture parameters on growth of different accessions and plumbagin content. This information can be utilized for more comprehensive studies on influences of various other parameters for the production of important phytocompounds present in *P. zeylanica*.

Author contribution statement AR: designed and performed all the experiments and wrote the manuscript with support from Dr NB. NB: supervised Arpita Roy for conducting the experiments successfully.

Acknowledgements This study was supported by our department. We are thankful to Dr. Neelam Sharma (NBPGR) and Dr. Girish Mishra (DU) for providing accessions and GC–MS facility.

References

Agati G, Azzarello E, Pollastri S, Tattini M (2012) Flavonoids as antioxidants in plants: location and functional significance. Plant Sci 196:67–76

Amin MEMH (2011) Effect of different nitrogen sources on growth, yield and quality of fodder maize (*Zea mays* L.). J Saudi Soc Agric Sci 10(1):17–23

Baskaran P, Jayabalan N (2005) Role of basal medium, carbon source and growth regulator in micropropagation of *Eclipta alba*—a valuable medicinal herb. KMITL Sci J 5:469–682

Benson EE (2000) In vitro plant recalcitrance, an introduction. In Vitro Cell Dev Biol Plant 36:141–148

Brand-Williams W, Cuvelier ME, Berset C (1995) Use of free radical method to evaluate antioxidant activity. Lebens Wiss Technol 28:25–30

Bruni R, Sacchetti G (2009) Factors affecting polyphenol biosynthesis in wild and field grown St. John's Wort (*Hypericum perforatum* L. Hypericaceae/Guttiferae. Molecules 14:682–725

Buah JN, Kawamitsu Y, Yonemori S, Hayashi M (2000) Effect of various carbon sources and their combination on in vitro growth and photosyntesis of banana plantlets. Plant Prod Sci 3:392–397

Chitrak products source (2018) https://www.google.co.in/search?q=chitrak+products&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjhx NGnbn ZAhVL G5QKH cO-D1YQ_AUICigB&biw=1440&bih=794. Accessed 09 Mar 2018

Chandravanshi M, Sahu Y, Agrawal A, Raja W (2014) In vitro micropropagation of important commorcial medicinal plant: *Plumbago zeylanica*. Adv Biol Res 8(3):139–142

Czapecka E, Mareczek A, Leja M (2005) Antioxidant activity of fresh and dry herbs of ome *Lamiaceae* species. Food Chem 93:223–226

Elgailani IEH, Ishak CY (2014) Determination of tannins of three common Acacia species of Sudan. Adv Chem 2014:5. https://doi.org/10.1155/2014/192708

Gubis J, Lajchova Z, Klcova L (2005) The effect of carbon source on plant regeneration in tomato. Hort Sci 32:6–8

Israeli Y, Lahav E, Reuveni O (1996) In vitro culture of bananas and plantains. In: Gowen S (ed) Bananas and plantains. Chapman and Hall, London, pp 1–612

Jetty A, Subhakar C, Rajagopal D, Jetty M, Subramanyam M, Murthy MM (2010) Antimicrobial activities of neo- and 1-epineoisoshinanolones from *Plumbago zeylanica* roots. Pharm Biol 48(9):1007–1011

Kumar R, Kumar S, Patra A, Jayalakshmi S (2009) Hepatoprotective activity of aerial parts of *Plumbago zeylanica* linn against carbon tetrachloride-induced hepatotoxicity in rats. Int J Pharm Pharm Sci 1:171–175

Kundu K, Roy A, Saxena G, Kumar L, Bharadvaja N (2016) Effect of different carbon sources and elicitors on shoot multiplication in accessions of *Centella asiatica*. Med Aromat Plants 5:251

Lamaison JL, Carnat A (1990) Content of principal flawonoids of the flowers and leaves of *Crataegus monogyna* Jacq. and *Crataegus laeviagata*. Pharm Acta Helv 65:315–320

Lin LC, Yang LL, Chou CJ (2003) Cytotoxic naphthoquinones and plumbagic acid glucosides from *Plumbago zeylanica*. Phytochemistry 62:619–622

Matkowski A, Tasarz P, Szypuła E (2008) Antioxidant activity of herb extracts from five medicinal plants from Lamiaceae, subfamily Lamioideae. J Med Plants Res 2(11):321–330

McLafferty FW, Stauffer DB (1989) Wiley/NBS registry of mass spectral data. Wiley, New York

Ming Y, Wang J, Yang J, Liu W (2011) Chemical constituents of *Plumbago Zeylanica* L. Adv Mater Res 308:1662–1664

New York explosion chelsea (2016) https://www.nytimes.com/2016/09/19/nyregion/new-york-explosion-chelsea.html. Accessed 09 Mar 2018

Petrova M, Zayova E, Dincheva I, Badjakov I, Vlahova M (2015) Influence of carbon sources on growth and GC–MS based metabolite profiling of *Arnica montana* L. Hairy roots, Turkish. J Biol 39:469–478

Rajakrishnan R, Lekshmi R, Benil PB, Thomas J, AlFarhan AH, Rakesh V, Khalaf S (2017) Phytochemical evaluation of roots of *Plumbago zeylanica* L. and assessment of its potential as a nephroprotective agent. Saudi J Biol Sci 24(4):760–766

Roy A, Bharadvaja N (2017) A review on pharmaceutically important medical plant: *Plumbago zeylanica*. J Ayu Herb Med 3(4):225–228

Roy A, Bharadvaja N (2018) Biotechnological approaches for the production of pharmaceutically important compound: plumbagin. Curr Pharm Biotechnol 19(5):372–381

Roy A, Kundu K, Saxena G, Kumar L, Bharadvaja N (2016) Effect of different media and growth hormones on shoot multiplication of in vitro grown *Centella asiatica* accessions. Adv Tech Biol Med 4:172

Roy A, Jauhari N, Bharadvaja N (2018) Medicinal plants as a potential source of chemopreventive agents. Anticancer plants: natural products and biotechnological implements. Springer, Singapore

Salah N, Miler NJ, Pagange G, Tijburg L, Bolwell GP, Rice E et al (1995) Polyphenolic flavonoids as scavenger of aqueous phase radicals as chain breaking antioxidant. Arch Biochem Biophys 2:339–346

Shad MA, Nawaz H, Rehman T, Ahmad HB, Hussain M (2012) Optimization of extraction efficiency of tannins from *Cichorium intybus* L.: application of response surface methodology. J Med Plants Res 6(28):4467–4474

Sharma A, Singh N (2015) A multifarious potent herb: *Plumbago zeylanica*—a mini review. Int J Recent Sci Res 6:4825–4829

Sharma I, Mathur M, Singh GP, Rishi A (2014) Quantitative estimation of phenolic and flavonoid content and antioxidant activity of various extracts of different parts of *Plumbago zeylanica Linn*. Int J Drug Dev Res 6(2):136–140

Sharma I, Mathur M, Singh GP (2015) Gas chromatography–mass spectrometry analysis and phytochemical screening of methanolic leaf extract of *Plumbago zeylanica, Linn*. Int J Pharm Sci Rev Res 33(1):315–320

Shimoi K, Masuda S, Shen B, Furugori M, Kinz N (1996) Radioprotective effects of antioxidative plant flavonoidsin mice. Mutat Res Fund Mol 350:153–161

Singleton V, Rossi JA (1965) Colorimetry of total phenolics with phosphomolybdic—phosphotungstic acid reagents. Am J Enol Vitic 16:144–158

Singleton VL, Orthofer R, Lamuela-Raventos RM (1999) Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin–Ciocalteau reagent. Methods Enzymol 299:152–178

Sridhar TM, Naidu CV (2011) Effect of different carbon sources on in vitro shoot regeneration of Solanum nigrum (Linn.) - an important antiulcer medicinal plant. J Phytol 3:78–82

Staba EJ (1980) Plant tissue culture as a source of biochemicals. CRC Press, Florida

Thompson M, Thorpe T (1987) Metabolic and non-metabolic roles of carbohydrates. In: Bonga JM, Durzan DJ (eds) Cell and tissue culture in forestry. Martinus Nijhoff Publishers, Dordrecht, pp 89–112

Uma Devi P, Soloman FE, Sharda AC (1999) Plumbagin a plant naphthoquinone with radiomodifying properties. Pharm Biol 37:231–236

Vijay R, Shukla J, Saxena R (2016) Micropropagation of *Plumbago zeylanica* L.: an important medicinal plant of India. Int J Pure App Biosci 4(6):59–167

Yaseen M, Ahmad T, Sablok G, Standardi A, Hafiz IA (2013) Review: role of carbon sources for in vitro plant growth and development. Mol Biol Rep 40(4):2837–2849

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/328421027

Enhancement of 1.54 μm emission in Ce3+–Er3+ codoped Ca4Si2O7F2 phosphor

Article *in* Journal of Alloys and Compounds · October 2018

DOI: 10.1016/j.jallcom.2018.10.198

CITATIONS
0

READS
52

1 author:

Rupesh Talewar
K L University
18 PUBLICATIONS **32** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Development of Quantum Cutting Near Infra-red Emitting (NIR) Phosphors For c-Si Solar Cell View project

Development of novel Mn4+ doped phosphors. View project

All content following this page was uploaded by Rupesh Talewar on 22 October 2018.

The user has requested enhancement of the downloaded file.

Journal of Alloys and Compounds 775 (2019) 810–817

Contents lists available at ScienceDirect

Journal of Alloys and Compounds

journal homepage: http://www.elsevier.com/locate/jalcom

Enhancement of 1.54 μm emission in Ce^{3+}-Er^{3+} codoped $Ca_4Si_2O_7F_2$ phosphor

Rupesh A. Talewar [a, **], Sk. Mahamuda [a, *], Arpita Vyas [b], A.S. Rao [c], S.V. Moharil [d]

[a] Department of Physics, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur (Dt.), 522502, AP, India
[b] Physics Department, Shri Ramdeobaba College of Engineering and Management, Katol Road, Nagpur, 440013, India
[c] Department of Applied Physics, Delhi Technological University, Delhi, 110042, India
[d] Department of Physics, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur, 440010, India

ARTICLE INFO

Article history:
Received 14 July 2018
Received in revised form
14 October 2018
Accepted 16 October 2018
Available online 17 October 2018

Keywords:
Near-infrared (NIR) emitting phosphor
Sensitization
Photoluminescence
Ce^{3+}-Er^{3+}
Energy transfer
$Ca_4Si_2O_7F_2$ phosphor

ABSTRACT

A novel near-infrared (NIR) emitting phosphor, $Ca_4Si_2O_7F_2$:Ce^{3+},Er^{3+} was synthesized by conventional solid state reaction method and characterized with X-ray diffraction, photoluminescence emission, photoluminescence excitation spectra and fluorescence decay measurements. In $Ca_4Si_2O_7F_2$:Ce^{3+},Er^{3+} phosphors, intense NIR emission at 1540 nm which is assigned to the characteristic $^4I_{13/2} \rightarrow {}^4I_{15/2}$ transition of Er^{3+} along with the broadband emission of Ce^{3+} at 490 nm was observed under 380 nm excitation. The luminescence spectra, both in visible (VIS) and NIR regions, and lifetime curves of Ce^{3+} have been measured to prove energy transfer (ET) from Ce^{3+} to Er^{3+}. Efficient energy transfer from Ce^{3+} to Er^{3+} in $Ca_4Si_2O_7F_2$ is observed and leads to about 10 times enhancement of the Er^{3+}-1540 nm emission when excited at 380 nm. The critical distance between Ce^{3+} and Er^{3+} ET in $Ca_4Si_2O_7F_2$ host was calculated, as 14.81 Å. The Inokuti-Hirayama (I-H) model is applied in the analysis of the non-exponential fluorescence decay curves. From I-H curve fitting, it is inferred that the electric dipole-dipole interaction is the main process responsible for the energy transfer as well as for the large enhancement of 1540 nm emission. The results indicate that sensitization of Er^{3+} is possible via 4f-5d transition of Ce^{3+} in the $Ca_4Si_2O_7F_2$ host.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Over the years, the research on rare-earth based near-infrared (NIR) emission is gaining a great deal of interest based on their practical applications like lasers [1], optical telecommunication [2], solar spectral converters [3], optical sensors [4], biomedical optical imaging [5,6], etc. For NIR emission, trivalent rare earth ions like Nd^{3+}, Yb^{3+} and Er^{3+} are most favorable activators because of their favorable energy level schemes. Among these rare earth ions, Er^{3+} is especially attractive because of its wide range of emission in visible, NIR and Mid-NIR [7]. The Er^{3+} optical absorption spectrum is comprised of a number of transitions, including eight which lie in the visible to the near-infrared spectral region. Optical pumping in any of them will cause the Er^{3+} ion to emit at 1.54 μm, a wavelength which happens to coincide with the "third" telecommunications window in silica fiber [8,9]. Especially, it is of significant practical importance in the field of optical communication, remote probing, for medicine and biophysics. Also, the emission of erbium at 2.7 μm in fluoride glasses constitutes a very promising system to construct all solid state lasers emitting near 3 μm to be applied as surgical instruments [10,11]. Erbium as a dopant in solid-state materials is further interesting for the construction of upconversion lasers for possible applications in the field of optical data storage (550 nm) [12]. Recently, Chen et al. proposed efficient near-infrared three-photon quantum cutting excited by visible light in Er^{3+} ions [13].

Generation of Er^{3+} NIR emission around 1.54 μm corresponds to the transition between the excited ($^4I_{13/2}$) and the ground ($^4I_{15/2}$) energy levels of this ion with the ground state electronic configuration $4f^{11}$ [14]. However, the absorption of Er^{3+} in near-ultraviolet and visible region is very weak and narrow due to the forbidden nature of the f-f transitions. A small absorption cross-section, typically less than 1×10^{-20} cm^2, and the losses due to re-absorption and upconversion result in low absorption coefficients for materials doped with Er^{3+} alone [15]. The low intra-4f-

* Corresponding author.
** Corresponding author.
E-mail addresses: talewarrupesh@gmail.com (R.A. Talewar), mahamuda.ss@gmail.com, mahamuda.ss@kluniversity.in (Sk. Mahamuda).

https://doi.org/10.1016/j.jallcom.2018.10.198
0925-8388/© 2018 Elsevier B.V. All rights reserved.

configurational absorption cross section limits the pump efficiency of Er^{3+} ions and therefore the potential application of this ion is limited. In order to solve this inherent problem associated with Er^{3+}, sensitization might be an effective pathway to increase the intra-4f-configurational absorption and to make more efficient 1.54 μm luminescence generation [16]. Sensitization from rare earth ions may be appropriate for this purpose. Sensitizers like Eu^{2+} and Ce^{3+} are more suitable compared to other rare earth ions because of their 4f→5d allowed transitions which is in contrast to forbidden f-f transitions [17]. An efficient energy transfer from (Eu^{2+}/Ce^{3+}) sensitizer to (Er^{3+}) activator ions would allow one to obtain a more efficient phosphor. Energy transfers from Eu^{2+}/Ce^{3+} as the sensitizer ions to Tb^{3+}, Nd^{3+}, Yb^{3+} trivalent rare earth ions have already been reported [18–22]. In recent years, there are few reports on energy transfer from Ce^{3+} to Er^{3+} co-doped crystals and glasses [23–26]. However, the emphasis was to increase the non radiative decay rate from $^4I_{11/2} \rightarrow {}^4I_{13/2}$ via energy transfer through $^2F_{5/2} \rightarrow {}^2F_{7/2}$ transition of Ce^{3+} ions which may result in the enhancement of Er^{3+} emission.

According to the Förster-Dexter theory [27,28], the criterion for an efficient energy transfer to take place is the spectral overlap between a sensitizer emission band and an acceptor absorption band. As Er^{3+} ions have a number of energy levels in the UV-VIS and NIR part of the spectrum, the sensitizer with an emission coinciding with one of these can be suitable for an efficient sensitization [29]. Ce^{3+} ion may seem to have the appropriate feature as a sensitizer since the dipole allowed inter-configuration 4f→5d transitions are involved in the processes of absorption and energy transfer. The strengths of these transitions exceed by several orders of magnitude those of parity forbidden intra-configuration transitions [30]. But the positions of the absorption and emission bands of Ce^{3+} strongly depend on the host matrix [31]. Therefore the choice of host matrix is crucial because efficient energy transfer between two species can occur when the emission band of the donor overlaps the excitation band of the acceptor. For example, in $Ca_2BO_3Cl:Eu^{2+}$, Nd^{3+}, Eu^{2+} ion sensitizes the $^4F_{3/2} \rightarrow {}^4I_j$ (j = 9/2, 11/2 and 13/2) NIR emission of Nd^{3+} due to overlap of 5d-4f emission of Eu^{2+} with $^4G_{7/2}$, $^2H_{11/2}$, $^2G_{7/2}$ levels of Nd^{3+} [32].

With the aim of preparing NIR emitting phosphor and sensitization from 4f-5d allowed transition, we choose Ce^{3+} as a sensitizer for Er^{3+}. For this, we prepared $Ca_4Si_2O_7F_2$ co-doped with Ce^{3+}-Er^{3+} through two-step solid-state reaction. This phosphor exhibits a great potential to act as an efficient converter of UV to near infrared. $Ca_4Si_2O_7F_2$ can be a suitable host as it is easy to prepare and stable at room temperature. $Ca_4Si_2O_7F_2$ was first reported to crystallize in the monoclinic belonging to space group $P2_1/c$ by Smirnova et al., in 1955 [33]. Thereafter, a number of reports appeared focussing on rare earth doped $Ca_4Si_2O_7F_2$. Jia et al. discussed the electronic structure and photoluminescence properties of Eu^{2+}-activated $Ca_4Si_2O_7F_2$ [34]. The photoluminescence properties of phosphors $Ca_4Si_2O_7F_2:Ce^{3+}$ having tunable emission are reported by Xie et al. who suggested that these have potential application in near-UV based w-LEDs [35]. Huang et al. investigated color tunability of $Ca_4Si_2O_7F_2:Eu^{2+},Mn^{2+}$ phosphors for similar applications [36]. Results of Lü et al. indicated that $Ca_4Si_2O_7F_2:Ce^{3+},Mn^{2+}$ may serve as a potential color-tunable NUV phosphor for white-light LED devices [37]. Though a number of reports appeared on rare earth doped $Ca_4Si_2O_7F_2$ phosphor in recent years, the emphasis was on visible region only; however, there were no reports on NIR emission in rare earth doped $Ca_4Si_2O_7F_2$. To the best of our knowledge, energy transfer from Ce^{3+}-Er^{3+} codoped $Ca_4Si_2O_7F_2$ via 4f to 5d transition of Ce^{3+} ions has not yet been reported. Such results are reported here. This phosphor may have potential applications in lasers, optical telecommunication, optical sensors, biomedical optical imaging, or as a spectral conversion for Ge based solar energy, etc.

2. Experimental

$Ca_4Si_2O_7F_2:Ce^{3+}$ and $Ca_4Si_2O_7F_2:Ce^{3+},Er^{3+}$ powder samples with different doping concentrations were prepared by conventional solid-state reaction technique. All the starting materials viz. $CaCO_3$, SiO_2, $Ce_2(CO_3)_3$, Er_2O_3 and NH_4HF_2 were of Analytical Reagent grade. The powders were thoroughly mixed in an agate mortar by grinding and the mixture was transferred into crucibles, pre-calcined at 600 °C for 3 h in the air and finally sintered at 1000 °C for 15 h under reducing atmosphere, followed by an additional grinding.

The crystalline phases of the synthesized samples were identified by x-ray diffraction (XRD) patterns recorded on XPERT-PRO diffractometer using Cu K_α (λ = 1.54059 Å) radiation. The UV-VIS and NIR photoluminescence (PL) and photoluminescence excitation (PLE) spectra were measured by a Hitachi-F7000 fluorescence spectrophotometer and Photon Technology International QM-51 NIR spectrophotometer. Luminescence decay curves were measured by using a time-correlated single photon counting (TCSPC) instrument from IBH Horiba-JY (United Kingdom) with 374 nm LED as the excitation source. All the measurements were carried out at room temperature.

3. Results and discussion

Fig. 1 shows the XRD patterns of $Ca_4Si_2O_7F_2$ sintered at 1000 °C for 15 h under reducing atmosphere. The XRD patterns of the sample reveal that all the peaks of the sample can be identified with the pure crystalline monoclinic $Ca_4Si_2O_7F_2$ (ICDD 41-1474) belonging to space group $P2_1/c$, and the lattice parameter values are $a = 10.909$ Å, $b = 10.547$ Å, $c = 7.539$ Å. The XRD graph shows that there are no additional diffraction peaks corresponding to any impurity phases, suggesting that a pure crystalline compound was obtained. In order to further investigate the phase purity of the as-synthesized samples, the Rietveld refinement was performed for $Ca_4Si_2O_7F_2$ as a representative. The initial structural model was constructed with crystallographic data previously reported for $Ca_4Si_2O_7F_2$ (JCPDS 41–1474) [38]. The Rietveld refined XRD pattern of $Ca_4Si_2O_7F_2$ host is shown in Fig. 2. The refined crystallographic parameters were well satisfied with the reflection conditions, and good fits were obtained with $\chi^2 = 1.38$, validating the phase purity of the as-prepared sample. Table 1 demonstrates the atomic

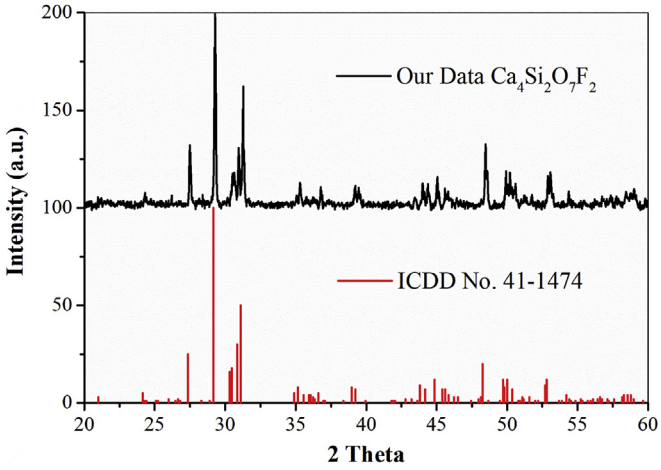


Fig. 1. XRD pattern of the $Ca_4Si_2O_7F_2$ and the reference data of ICDD 41-1474.

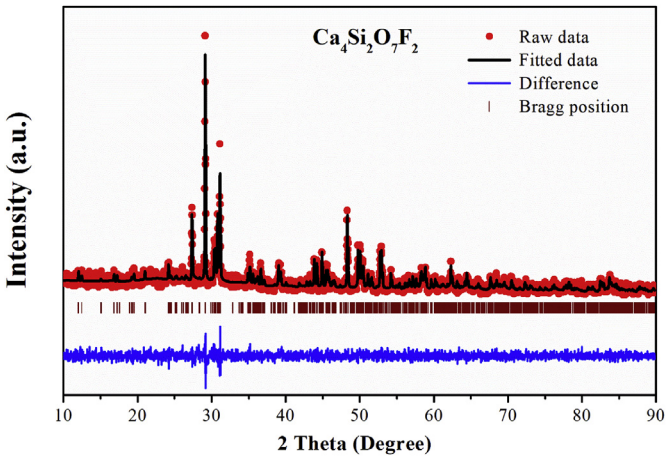


Fig. 2. Rietveld refined XRD pattern of $Ca_4Si_2O_7F_2$ host.

Table 1
Atomic coordinates, cation and anion occupancies of $Ca_4Si_2O_7F_2$ compound.

Atom	x	y	z	Site occupancy factor (SOF)
Ca1	−0.03454	0.40466	0.69298	1.00
Ca2	0.67237	0.12644	0.58415	0.97
Ca3	0.83258	0.63721	0.92780	0.94
Ca4	0.47504	0.41163	0.68942	1.00
Si1	0.28432	0.30768	0.36837	1.00
Si2	−0.15931	0.30882	0.37871	1.00
F1	0.57131	0.50393	0.88544	1.00
F2	0.99000	0.50222	0.90673	0.93
O1	0.73447	0.27744	0.74645	1.00
O2	0.06289	0.27500	0.40525	0.91
O3	−0.12651	0.27104	0.52718	1.00
O4	0.74019	0.78238	0.74748	1.00
O5	0.40044	0.26194	0.50969	1.00
O6	0.22255	0.53132	0.65842	0.96
O7	−0.28419	0.53996	0.67359	1.00

coordinates, cation and anion occupancies of $Ca_4Si_2O_7F_2$ compound. After refinement, crystal structures of $Ca_4Si_2O_7F_2$ compound is generated using output.cif file by Vesta software as displayed in Fig. 3. Fig. 3 shows the crystal structure of $Ca_4Si_2O_7F_2$, emphasizing the coordination environment of Ca^{2+} and Si^{4+} atoms. The crystal structure of cuspidine ($Ca_4Si_2O_7F_2$) was first determined by Smirnova et al., in 1955 [33]. According to their work, the crystal structure of $Ca_4Si_2O_7F_2$ is built up from edge-sharing octahedral CaO_6 chains parallel to a-axis. In their study, they considered that all calcium atoms were six coordinated. However, this consideration was not correct and later it was corrected by Saburi et al. [39]. Saburi et al. findings show that four different crystallographic sites are available for the divalent Ca^{2+} ions. The polyhedra formed by these four different crystallographic sites form the columns and these columns are linked to construct the three-dimensional network. Ca^{2+} crystallographic sites are directly coordinated with the O and F atoms in different O/F ratio. In $Ca_4Si_2O_7F_2$, Ca^{2+} ions have several different coordination environment: 1) Ca(1) atom is eight-coordinated by five O and three F atoms [$Ca(1)O_5F_3$]. 2) Ca(2) atom is surrounded by four O and three F atoms [$Ca(2)O_4F_3$]. 3) Ca(3) atom is six-coordinated by five O atoms and one F atom [$Ca(3)O_5F$]. 4) Ca(4) atoms are seven-coordinated by six O atoms and one F atom [$Ca(4)O_6F$]. Hence, when Ce^{3+} ions are introduced into the $Ca_4Si_2O_7F_2$ host, they may enter different Ca^{2+} sites owing to the effective ionic radii of cations Ca^{2+}, 1.12 Å [$Ca(1)O_5F_3$]; 1.06 Å [$Ca(2)O_4F_3$]; 1.00 Å [$Ca(3)O_5F$] and 1.06 Å [$Ca(4)O_6F$]; Ce^{3+}, 1.07 Å; Er^{3+}, 0.94 Å [40].

Generally, when the trivalent rare earth ions, such as Ce^{3+} and Er^{3+}, substitute for the bivalent alkaline earth, such as Ca^{2+}, in the compound, there has the charge unbalance which may result in the formation of defects in the phosphor. Therefore, the incorporation of alkali metal ions such as Li, Na or K might offset the charge unbalance generated by Ce^{3+} and Er^{3+} ions substitution for Ca^{2+} ion, reduced the lattice distort, and affected the luminescent properties [35,41]. However, for the case of Ce^{3+}/Er^{3+} co-doped $Ca_4Si_2O_7F_2$, the charge loss in the lattice $Ca_4Si_2O_7F_2$ is probably compensated by Ca^{2+} vacancies. Moreover, in the present work, the research focuses on the Ce^{3+}-Er^{3+} energy transfer in $Ca_4Si_2O_7F_2$ [42]. Hence, the effect of charge compensation on the luminescent properties of the phosphors is no any longer discussed.

The photoluminescence excitation and emission spectra of Ce^{3+} doped $Ca_4Si_2O_7F_2$ with Ce^{3+} concentration varying from 0.5 mol% to 2.5 mol% [$Ca_{4-x}Ce_xSi_2O_7F_2$ (x = 0.02, 0.04, 0.06, 0.08 and 0.10)] are shown in Fig. 4. Corresponding to the 465 nm emission, a broad excitation band centered at 325 nm with a hump at 365 nm can be observed. These broad excitation bands are assigned to the electronic transitions from the 4f level ($^2F_{5/2}$) ground state to the different crystal-field splitting of 5d excited states of Ce^{3+} ions [43]. Under 365 nm excitation, the emission spectrum exhibits an intense and broad asymmetrical band in the wavelength range of 370–630 nm with a peak at about 465 nm. This broadband emission of Ce^{3+} at 465 nm arises due to parity allowed transitions from

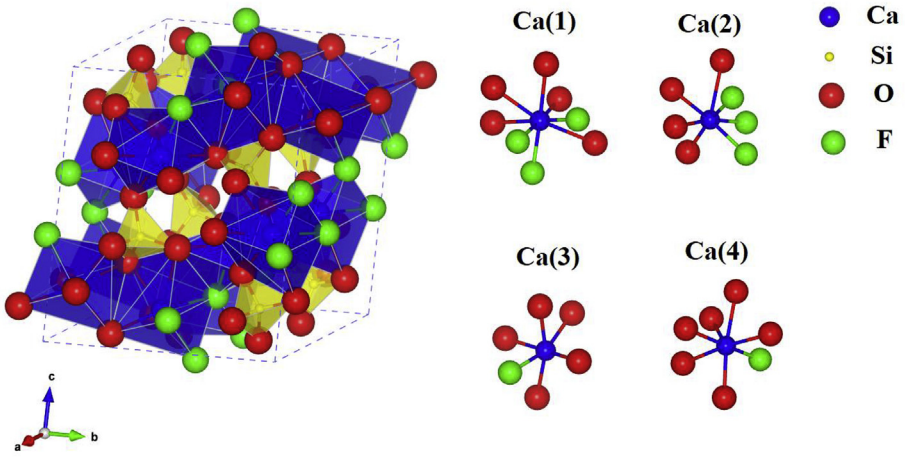


Fig. 3. Crystal structure of $Ca_4Si_2O_7F_2$, emphasizing the coordination environment of Ca^{2+} and Si^{4+} atom.

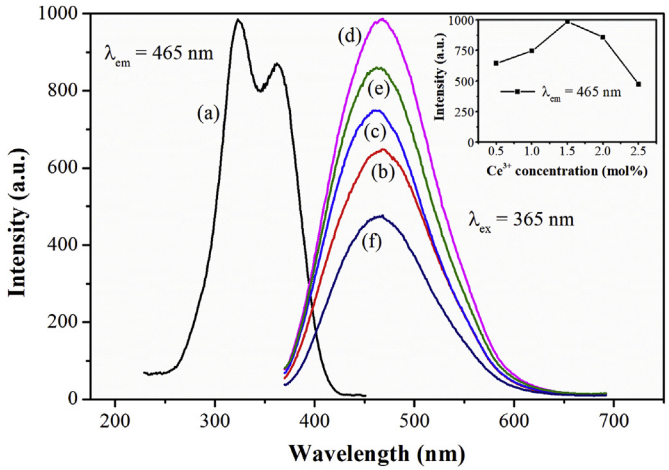


Fig. 4. PLE (a, $\lambda_{em} = 465$ nm) and PL (b-f, $\lambda_{ex} = 365$ nm) spectra of [$Ca_{4-x}Ce_xSi_2O_7F_2$ (x = 0.02, 0.04, 0.06, 0.08 and 0.10)] phosphors. Inset shows variation of emission intensity with activator concentration.

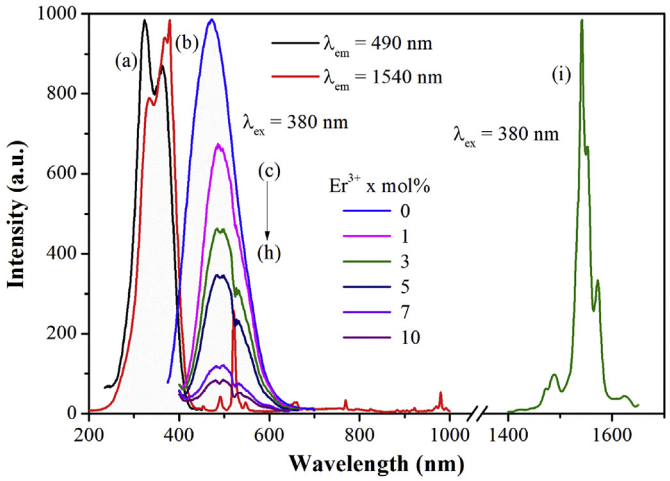


Fig. 6. PLE (a, $\lambda_{em} = 490$ nm and b, 1540 nm) and PL (c-i, $\lambda_{ex} = 380$ nm) spectra of the $Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.00, 0.04, 0.12, 0.20, 0.28 and 0.40) samples.

the lowest component of 5d excited state to $^2F_{5/2}$ ground state of Ce^{3+} ions [31]. The broad and asymmetric emission band of $Ca_4Si_2O_7F_2$: Ce^{3+} may be assigned to the transitions of Ce^{3+} ions occupying four different Ca^{2+} sites in the host lattice. The emission spectra of the Ce^{3+} doped $Ca_4Si_2O_7F_2$ phosphor can be decomposed by Gaussian deconvolution into four Gaussian profiles with peaks centered at 418 nm (23923 cm^{-1}), 464 nm (21551 cm^{-1}), 508 nm (19685 cm^{-1}) and 540 nm (18518 cm^{-1}) as shown in Fig. 5(b). These peaks can be ascribed to four different emission sites, which could be identified as the four different coordination environments of the Ca^{2+} ion sites occupied by Ce^{3+} ions. The emission and excitation spectra of Ce^{3+} doped $Ca_4Si_2O_7F_2$ phosphor are in good agreement with the reported value of $Ca_4Si_2O_7F_2$:Ce^{3+} [44]. The effect of doping concentration of Ce^{3+} (0.5–2.5 mol%) [$Ca_{4-x}Ce_xSi_2O_7F_2$ (x = 0.02, 0.04, 0.06, 0.08 and 0.10)] on the emission intensity of $Ca_4Si_2O_7F_2$ phosphors was also investigated. The PL spectrum also reveals that as the concentration of Ce^{3+} increases, the emission intensity increases and reaches a maximum at 1.5 mol% [$Ca_{3.94}Ce_{0.06}Si_2O_7F_2$], beyond this the concentration quenching is observed. This concentration quenching may be due to energy transfer process among Ce^{3+} ions.

Fig. 6 shows the PLE and PL spectra of Ce^{3+}-Er^{3+} co-doped $Ca_4Si_2O_7F_2$ with constant Ce^{3+} concentration (1.5 mol%) and different Er^{3+} concentrations (1–10 mol%) [$Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.00, 0.04, 0.12, 0.20, 0.28 and 0.40)]. Under 380 nm excitation, the phosphor exhibits broad asymmetric emission centered at 490 nm and intense sharp emission at 1540 nm. The former broad emission is assigned to the allowed 5d→4f transition of Ce^{3+} ions and the later is the characteristic of Er^{3+} emission arising due to $^4I_{13/2} \rightarrow {}^4I_{15/2}$ transition [45,46]. Er^{3+} emission in $Ca_4Si_2O_7F_2$:Ce^{3+}, Er^{3+} consists of several peaks in the infrared region with a sharper peak located at 1540 nm and other peaks at 1472 nm, 1488 nm, 1552 nm, 1572 and 1624 nm (Fig. 6 curve i). The reason is that Stark sub-levels are observed in the $^4I_{15/2}$ ground state and $^4I_{13/2}$ excited state of Er^{3+} [47,48]. The excitation spectrum (Fig. 6 curve a & b) was measured by monitoring the Ce^{3+} emission at 490 nm and characteristic Er^{3+} emission at 1540 nm. Corresponding to the $^4I_{13/2} \rightarrow {}^4I_{15/2}$ transitions of Er^{3+} at 1540 nm, a broad absorption band in the UV region with centered at 380 nm with a hump at about 335 nm were observed. In addition to the UV band, several lines in the wavelength region 400–1000 nm also can be observed. It is interesting to note that the excitation spectrum in the UV region monitored for 1540 nm emission is almost the same as that for the 490 nm emission, and it is similar to the excitation spectrum for $Ca_4Si_2O_7F_2$:Ce^{3+} as shown in Fig. 4 (curve a). Therefore we can say that this broad absorption band in the UV region is ascribed to 4f→5d transition of Ce^{3+} ions. However, the lines observed in the wavelength region 400–1000 nm are due to various *f-f* transitions of Er^{3+} ions [49,50]. It is also seen that compared to Ce^{3+} singly doped $Ca_4Si_2O_7F_2$ phosphor, the excitation and emission band for Ce^{3+}-Er^{3+} codoped samples corresponds to Ce^{3+} ions is red-shifted by about 15 nm. The red-shift in the excitation and emission spectra may be due to perturbation of Ce^{3+} energy levels by Er^{3+} [43,51]. As reported by Dorenbos, the red-shift of Ce^{3+} excitation and emission wavelength occurs when the smaller ions are introduced to the host, which can be attributed to larger crystal field splitting [43]. Introduction of smaller ions in the host will destroy the crystal symmetry and balance, i.e. changing the crystal field, leading to a decrease of 5d configuration levels. This trend is increased with the decreasing ions radius. In the Ce^{3+}/Er^{3+} co-doped sample, the radius of Er^{3+} (0.94 Å) is much smaller than that of Ca^{2+} (1.06 Å). Therefore, the red-shift of Ce^{3+} f–d transition is expected. The similarity in shape of excitation spectra in the UV region monitored for Ce^{3+} emission at 490 and Er^{3+} emission at 1540 nm can be considered as a direct evidence of ET from Ce^{3+} to Er^{3+} in

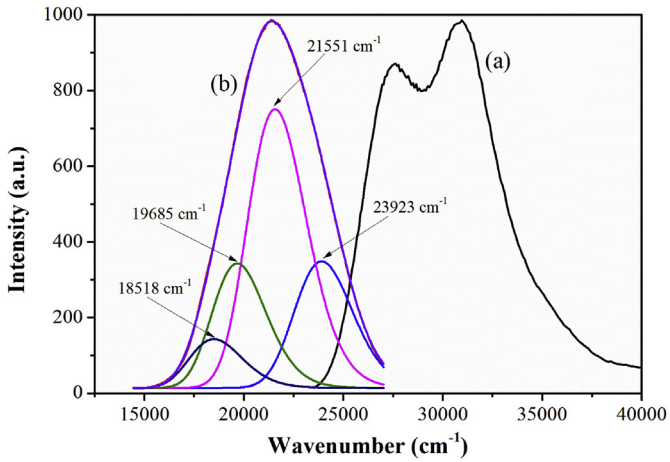


Fig. 5. PLE (a, $\lambda_{em} = 465$ nm) and PL (b, $\lambda_{ex} = 365$ nm) spectra of $Ca_{3.94}Ce_{0.06}Si_2O_7F_2$ phosphors.

$Ca_4Si_2O_7F_2$ host.

In order to investigate energy transfer from Ce^{3+} to Er^{3+} in $Ca_4Si_2O_7F_2$ phosphor, Er^{3+} singly doped $Ca_4Si_2O_7F_2$ phosphor is also prepared. The PLE and PL spectra of Er^{3+} singly doped and Ce^{3+}-Er^{3+} codoped $Ca_4Si_2O_7F_2$ phosphor is shown in Fig. 7. The excitation spectrum for the $^4I_{13/2} \rightarrow {}^4I_{15/2}$ transition of Er^{3+} at 1540 nm, (Fig. 7 curve b), contains several lines arising from f-f transitions of Er^{3+} ions. It is seen that the broad excitation band in the UV region is not observed, indicating that the broadband is associated with Ce^{3+} in $Ca_4Si_2O_7F_2$:Ce^{3+},Er^{3+} phosphor. Under 380 nm excitation, very weak emission (Fig. 7 curve d) is obtained at 1540 nm which is attributed to the characteristic $^4I_{13/2} \rightarrow {}^4I_{15/2}$ transition of Er^{3+}. For $Ca_4Si_2O_7F_2$:Ce^{3+},Er^{3+}, the excitation intensity in the range 250–430 nm is increased several folds as compared to $Ca_4Si_2O_7F_2$:Er^{3+} (Fig. 7 curve b). Emission intensity also shows a similar increase when the excitation wavelength is 380 nm (Fig. 7 curve c). According to the Förster-Dexter theory [27,28], the criterion for an energy transfer to take place is the spectral overlap between a donor emission band and an acceptor absorption band. Inset of Fig. 7 shows the photoluminescence emission spectra of Ce^{3+} with different concentrations of Er^{3+} and excitation spectra of Er^{3+} singly doped $Ca_4Si_2O_7F_2$ phosphor in the range 400–700 nm. It can be seen that the Ce^{3+} emission spectra show a dip at around 490 nm and 521 nm, where Er^{3+} has absorption bands corresponding to $^4I_{15/2} \rightarrow {}^4F_{3/2}$ (490 nm), $^4I_{15/2} \rightarrow ({}^2H,{}^4G)_{11/2}$ (521 nm) transitions. Moreover, the emission spectrum of Ce^{3+} corresponding to 5d→4f transitions shows a clear overlap with the absorption bands of Er^{3+} at $^4I_{15/2} \rightarrow {}^4F_{3/2}$ (436 nm), $^4I_{15/2} \rightarrow {}^4F_{5/2}$ (452 nm), $^4I_{15/2} \rightarrow {}^4F_{3/2}$ (490 nm), $^4I_{15/2} \rightarrow ({}^2H,{}^4G)_{11/2}$ (521 nm), $^4I_{15/2} \rightarrow {}^4S_{3/2}$ (546 nm) in the spectral range 400–700 nm. Also, the Ce^{3+} emission spectra show a dip at around 490 nm and 521 nm due to the absorption of Er^{3+} $^4I_{15/2} \rightarrow {}^4F_{3/2}$ (490 nm), $^4I_{15/2} \rightarrow ({}^2H,{}^4G)_{11/2}$ (521 nm) transitions. This overlap of Ce^{3+} emission spectrum and Er^{3+} absorption spectra assures the possibility of energy transfer from Ce^{3+} to Er^{3+}.

To further verify the occurrence of ET from Ce^{3+} to Er^{3+}, the dependence of visible emission at 490 nm and NIR emission at 1540 nm on the Er^{3+} doping concentration in the $Ca_4Si_2O_7F_2$:Ce^{3+},Er^{3+} samples upon excitation at 380 nm were studied and the results are shown in the inset of Fig. 8. It is observed

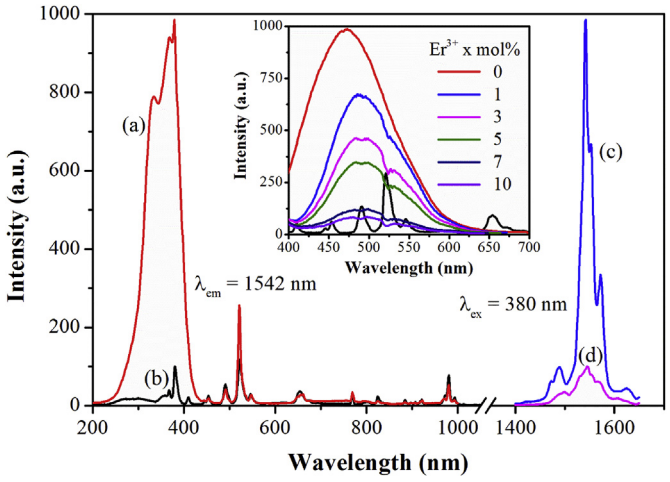


Fig. 7. PLE (a, $\lambda_{em} = 1540$ nm) and PL (c, $\lambda_{ex} = 380$ nm) spectra of the $Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.00, 0.04, 0.12, 0.20, 0.28 and 0.40) samples. PLE (b, $\lambda_{em} = 1540$ nm) and PL (d, $\lambda_{ex} = 380$ nm) spectra of the $Ca_{3.88}Er_{0.12}Si_2O_7F_2$ sample. Inset shows emission spectra of Ce^{3+} with different concentration of Er^{3+} (1–10 mol%) [$Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.00, 0.04, 0.12, 0.20, 0.28 and 0.40)] and excitation spectra of Er^{3+} (3 mol%) [$Ca_{3.88}Er_{0.12}Si_2O_7F_2$] singly doped $Ca_4Si_2O_7F_2$ phosphor.

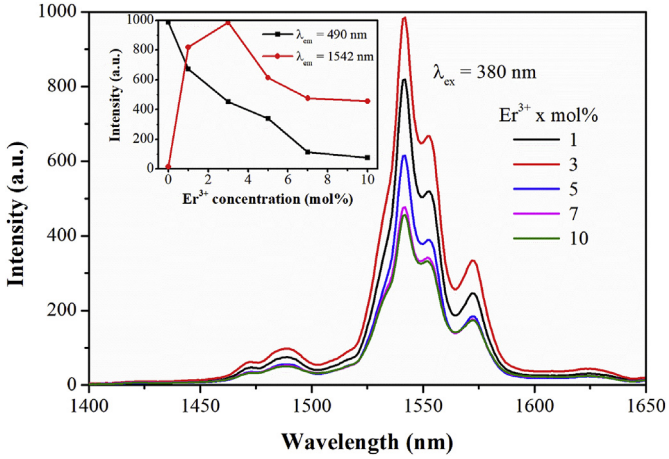


Fig. 8. PL ($\lambda_{ex} = 380$ nm) spectra of the $Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.04, 0.12, 0.20, 0.28 and 0.40) samples. Inset shows the dependence of Ce^{3+} (490 nm) and Er^{3+} (1540 nm) emission intensity on the doping concentration of Er^{3+} in $Ca_4Si_2O_7F_2$:Ce^{3+},Er^{3+}.

that, as the Er^{3+} content increases from 1 mol% to 10 mol% [$Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.04, 0.12, 0.20, 0.28 and 0.40)], the intensity of Ce^{3+} emission at 490 nm decreases monotonously, while the intensity of 1540 nm emission of Er^{3+} increases rapidly first and reaches a maximum at 3 mol% [$Ca_{3.82}Ce_{0.06}Er_{0.12}Si_2O_7F_2$], beyond which the concentration quenching is observed. The above phenomenon is a direct proof of ET from Ce^{3+} to Er^{3+}. It may be noted that visible and NIR emissions have been recorded on different instruments, hence the intensities are not comparable. Peak Heights of emission and excitation bands are normalized to 1000.

In order to get insight in the energy transfer phenomenon and to give a convincing evidence for the presence of ET from Ce^{3+} to Er^{3+}, PL decay curves of the Ce^{3+} ions in [$Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.00, 0.04, 0.12, 0.20, 0.28 and 0.40)] were measured with excitation at 374 nm and monitored the emission at 490 nm are shown in Fig. 9. For Ce^{3+} single doped [$Ca_{3.96}Ce_{0.06}Si_2O_7F_2$] and Er^{3+} (1–3 mol%) [$Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.04 and 0.12)] codoped phosphor, the decay curves were fitted using second-order exponential equation (1), whereas for higher Er^{3+} (5–10 mol%) [$Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.20, 0.28 and 0.40)] co-doped samples, third-order exponential equation (2) were used.

$$I = I_0 + a_1 * \exp(-t/\tau_1) + a_2 * \exp(-t/\tau_2) \quad (1)$$

$$I = I_0 + a_1 * \exp(-t/\tau_1) + a_2 * \exp(-t/\tau_2) + a_3 * \exp(-t/\tau_3) \quad (2)$$

where, I_0 is the intensity at t = 0, and I is intensity at time t, a_1, a_2 and a_3 are the amplitudes of decay constants and τ_1, τ_2 and τ_3 are the luminescence lifetimes for fast and slow channels of decay respectively.

It is interesting to note that for Ce^{3+} singly doped $Ca_4Si_2O_7F_2$ phosphor, the second-order exponential equation was well fitted with an average lifetime of 32.23 ns. The lifetime of the order of nanosecond is one of the characteristics of the electric-dipole allowed 5d→4f transition of Ce^{3+} [52]. The presence of bi-exponential decay is may be due to the energy transfer between Ce^{3+}-Ce^{3+} ions, as there are four Ca^{2+} sites in $Ca_4Si_2O_7F_2$ phosphor which can be occupied by Ce^{3+} ions. Moreover when the content of Er^{3+}, increases from 5 to 10 mol% [$Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.20, 0.28 and 0.40)], the decay curves gradually deviates from a bi-exponential to tri-exponential, implying the introduction of the

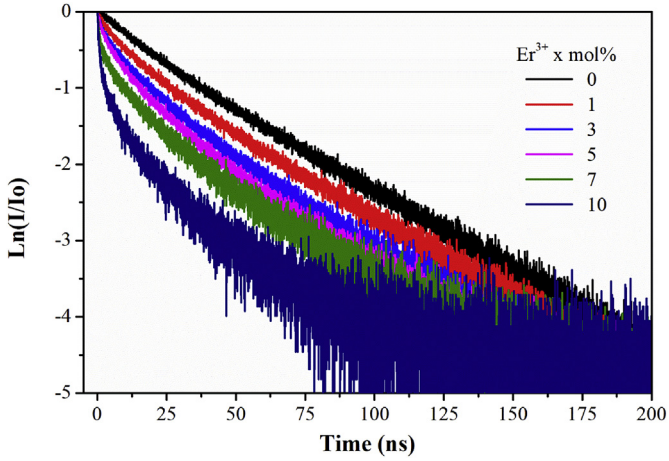


Fig. 9. Decay curves corresponding to Ce^{3+}: 5d-4f transition in $Ca_4Si_2O_7F_2$:Ce^{3+},Er^{3+} with different Er^{3+} concentrations under 374 nm excitation.

extra decay pathway. This decay shifting from bi-exponential to tri-exponential can be attributed to ET from Ce^{3+} ions to neighboring Er^{3+} ions. The average lifetime was calculated using the equation given in the literature [53]. The obtained average lifetimes measured from the decay curves are presented in Table 2. The energy transfer efficiency (η_{ETE}) from Ce^{3+} to Er^{3+} in the $Ca_4Si_2O_7F_2$ phosphor can be calculated by using the following equation [54,55].

$$\eta_{ETE}(\%) = \left(1 - \frac{\tau_x}{\tau_0}\right) x\ 100 \tag{3}$$

where η_{ETE} is the energy transfer efficiency expressed as a fraction, τ_0 is the lifetime of the sensitizer Ce^{3+} in the sample without Er^{3+} co-doping, and τ_x is the lifetime of Ce^{3+} along with Er^{3+} co-doping. The energy transfer efficiency values (η_{ETE}) calculated using the above equation are given in Table 2. From the data given in Table 2, it is conspicuous that the average lifetime values are decreasing with increasing of Er^{3+} ion whereas the energy transfer efficiency increases with the increase of Er^{3+} ion concentration gradually from 1 to 10 mol% [$Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.04, 0.12, 0.20, 0.28 and 0.40)].

In general, the exchange and electric multipolar interactions are two dominant mechanisms responsible for the energy transfer between a sensitizer and an activator. It is well known that if the energy transfer takes by the exchange mechanism, the critical distance between sensitizer and an activator should be shorter than 5 Å [56], indicating that this interaction is only important at very short distance between the energy donor and acceptor. Therefore, to investigate the Er^{3+}-Er^{3+} exchange interaction in $Ca_4Si_2O_7F_2$, the critical distance (R_c) for Er^{3+}-Er^{3+} energy transfer has been calculated according to the equation given by Blasse [57]:

$$R_c = 2\left[\frac{3V}{4\pi X_c N}\right]^{1/3} \tag{4}$$

where X_c is the critical concentration (0.03), N is the number of cation sites in the unit cell, and V is the volume of the unit cell. For $Ca_4Si_2O_7F_2$, N = 16 and V = 817.21 Å. The critical distance R_c is estimated to be about 14.81 Å. Therefore as the critical distance R_c = 14.81 Å, is much longer than 5 Å; hence, the exchange interaction mechanism plays no role in the energy transfer within the $Ca_4Si_2O_7F_2$:Ce^{3+},Er^{3+} phosphors.

In order to study the electric multipolar interaction between the sensitizer and activator, the Inokuti-Hirayama (I-H) model of the donor to acceptor energy transfer is applied in the analysis of the non-exponential fluorescent decays of Ce^{3+}-Er^{3+} co-doped $Ca_4Si_2O_7F_2$. The non-exponential decay curves of Ce^{3+}-Er^{3+} for Er^{3+} = 0 and 10 mol% [$Ca_{3.94-x}Ce_{0.06}Er_xSi_2O_7F_2$ (x = 0.00 and 0.40)] were used to analyse the electric multipolar interaction between Ce^{3+} and Er^{3+} ions. According to the Inokuti-Hirayama (I-H) [58], the non-exponential decay can be expressed as

$$I(t) = I_o \exp\left\{-\frac{t}{\tau_o} - Q\left(\frac{t}{\tau_o}\right)^{\frac{3}{s}}\right\} \tag{5}$$

where I_0 denotes the intensity at t = 0, τ_0 is the intrinsic decay time of the donors in the absence of acceptors. The s value depends on the mechanism of energy transfer and is 6, 8, and 10 for the electric dipole-dipole, electric dipole-quadrupole, and electric quadrupole-quadrupole interaction between luminescent centers, respectively. Q is the macroscopic parameter that determines the effect of multipolar interaction in decay property i.e the energy transfer parameter and it is defined as

$$Q = \frac{4\pi}{3}\Gamma\left(1 - \frac{3}{s}\right)NR_c{}^3 \tag{6}$$

where Γ is a gamma function, N is the rare earth ion concentration in cm^{-3} and R_c is the critical transfer distance defined as the donor-acceptor separation when the rate of energy transfer between a donor and acceptor ion is equal to the rate of intrinsic decay, ($1/\tau_0$). Fig. 10 shows the experimental data for Ce^{3+} singly [$Ca_{3.94}Ce_{0.06}Si_2O_7F_2$] doped and Ce^{3+}, Er^{3+} codoped [$Ca_{3.54}Ce_{0.06}Er_{0.40}Si_2O_7F_2$] fitted using equation (4). The linear fitting can be made to get the best value of s. As can be seen in the inset of Fig. 8, the concentration quenching begins at about 5 mol% [$Ca_{3.74}Ce_{0.06}Er_{0.20}Si_2O_7F_2$] of Er^{3+}. For lower concentrations of Er^{3+} ions, the decay curves are nearly in single exponential form as can be seen from Fig. 9. At higher concentrations, the decay curves deviate from exponential to non-exponential. For higher Er^{3+} concentrations i.e. for 10 mol% [$Ca_{3.54}Ce_{0.06}Er_{0.40}Si_2O_7F_2$], linear fit were obtained for s = 6. However, with s = 8 and 10, the I-H fitting deviates from the experimental decay curve. The good agreement of the I-H fitting for s = 6 suggests that the resonant type energy transfer caused by the electric dipole-dipole transfer is responsible for the non-

Table 2
Average lifetime (τ_{avg}) (ns) and energy transfer efficiency (η_{ETE}) (%) of Ce^{3+} emission in singly doped, and Ce^{3+}, Er^{3+} co-doped $Ca_4Si_2O_7F_2$ phosphor.

x mol% Er^{3+}	a_1	τ_1 (ns)	a_2	τ_2 (ns)	a_3	τ_3 (ns)	τ_{avg} (ns)	η_{ETE} (%)
0	0.2261	14.15	0.1954	39.68			32.23	0
1	0.1192	11.37	0.0564	41.59			30.52	5
3	0.4269	10.99	0.0524	46.33			23.05	28
5	0.0344	5.29	0.0346	29.91	0.6585	0.49	20.79	35
7	0.2524	11.63	0.0314	40.76	0.0981	1.49	19.85	38
10	0.0349	2.83	0.0165	20.19	0.7214	0.46	9.38	71

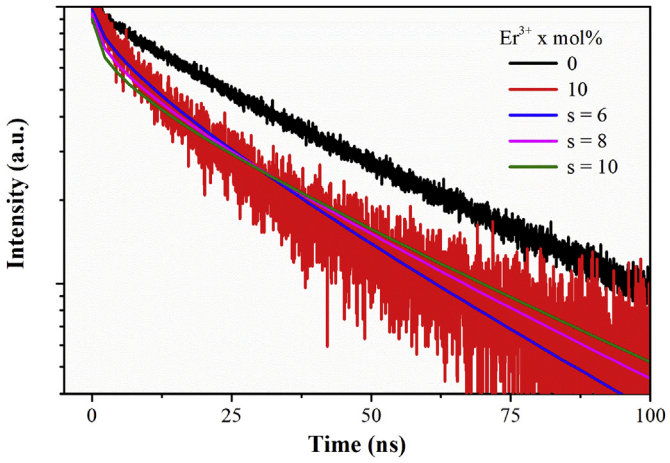


Fig. 10. Decay curve corresponding to Ce^{3+}: 5d-4f transition in $Ca_{3.96}Ce_{0.06}Si_2O_7F_2$ and $Ca_{3.54}Ce_{0.06}Er_{0.40}Si_2O_7F_2$ under 374 nm excitation. The solid lines correspond to the fits to the Inokuti-Hirayama Model for $s = 6$, 8 and 10.

exponential character of the fluorescence decay. Therefore, the large resonant rates at higher concentrations with an electric dipole-dipole interaction are the main processes responsible for the shortening of the fluorescence lifetime in the Ce^{3+}-Er^{3+} co-doped $Ca_4Si_2O_7F_2$ samples.

The schematic energy level diagram with transitions which may be involved in the energy transfer process from Ce^{3+} ion to Er^{3+} ions in $Ca_4Si_2O_7F_2$ and the possible energy transfer mechanisms are shown in Fig. 11. From this diagram, it can be seen that the 5d→4f transition of Ce^{3+} mismatches with the excited $^4F_{3/2}$, $^4F_{5/2}$, $^4F_{3/2}$, $(^2H,^4G)_{11/2}$ and $^4S_{3/2}$ level of Er^{3+} in the spectral range 400–630 nm. When excited at 230–430 nm, the energy could be efficiently absorbed by Ce^{3+} ions via 4f→5d transition. The electrons in the 5d excited levels of Ce^{3+} can either relax radiatively to its 4f ground state, producing broadband luminescence with the peak at about 490 nm, or transfer energy to Er^{3+}: [$^4F_{3/2}$ + $^4F_{5/2}$ + $^4F_{3/2}$ + $(^2H,^4G)_{11/2}$ + $^4S_{3/2}$ and other closely spaced levels] and populate these states in the region 400–630 nm. The excitation energy from 5d→4f state can be transferred to these levels through the resonance energy transfer process. Then, the activated electrons in these levels would relax non-radiatively to $^4I_{9/2}$, $^4I_{11/2}$ and $^4I_{13/2}$ levels. Electrons at $^4I_{13/2}$ level then decay to $^4I_{15/2}$ ground state with the NIR emission of 1540 nm.

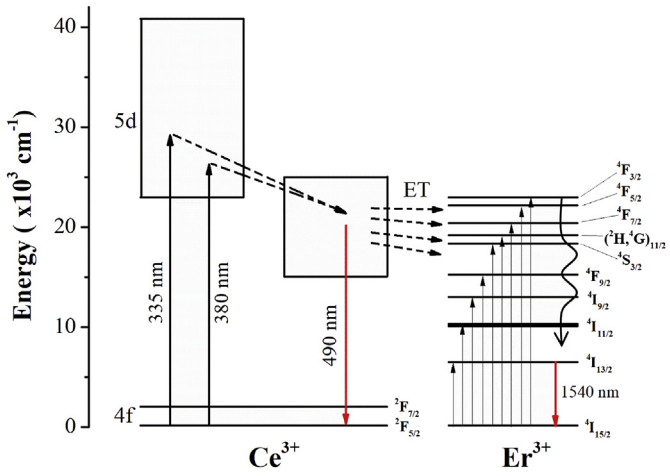


Fig. 11. Schematic energy level diagram of Ce^{3+} and Er^{3+} in the $Ca_4Si_2O_7F_2$ phosphor, showing the energy transfer mechanism for the NIR luminescence under 380 nm excitation.

4. Conclusions

In conclusion, a novel and efficient near-infrared (NIR) emitting phosphor, $Ca_4Si_2O_7F_2$:Ce^{3+},Er^{3+} was successfully synthesized by conventional two-step solid-state reaction. An efficient energy transfer from Ce^{3+} to Er^{3+} leads to the NIR emission enhancement. The excitation spectrum of Ce^{3+} in $Ca_4Si_2O_7F_2$ spans the range of 230–430 nm and efficient energy transfer from Ce^{3+} to Er^{3+} enables broadband sensitization of Er^{3+} in the Ce^{3+}-Er^{3+} co-doped phosphor. The Ce^{3+}→Er^{3+} energy transfer efficiency obtained from the Ce^{3+} lifetimes in the single doped and co-doped samples reaches up to 71%. Efficient energy transfer from Ce^{3+} to Er^{3+} in $Ca_4Si_2O_7F_2$ is observed and gives about 10 times enhancement on the Er^{3+}-1540 nm emission excited at 380 nm. The Inokuti-Hirayama (I-H) model of sensitizer to activator energy transfer is applied in the analysis of the non-exponential fluorescent decays of Ce^{3+}-Er^{3+} co-doped $Ca_4Si_2O_7F_2$. The dipole-dipole resonant type interaction is the most probable mechanism for the Ce^{3+}-Er^{3+} energy transfer in $Ca_4Si_2O_7F_2$. The interesting luminescence properties highlight the Ce^{3+}-Er^{3+} codoped $Ca_4Si_2O_7F_2$ system as useful for Er^{3+}-based optical and photonic devices.

Acknowledgement

This work was financially supported by the Department of Science and Technology, Government of India (File No.: ECR/2016/000376 and SR/S2/CMP-008/2012 dated 20/05/2013). The authors would like to acknowledge Dr. Anindya Datta, Professor at Department of Chemistry, Indian Institute of Technology Bombay, Powai, India and Dr. Sujit Kumar Ghosh, Professor at Department of Chemistry, Visvesvaraya National Institute of Technology (VNIT), Nagpur, India for valuable discussions and for providing Fluorescence Decay Measurements facilities.

References

[1] Suchinda Sattayaporn, Gerard Aka, Pascal Loiseau, Akio Ikesue, Yan Lin Aung, Optical spectroscopic properties, 0.946 and 1.074 μm laser performances of Nd^{3+}-doped Y_2O_3 transparent ceramics, J. Alloys Compd. 711 (2017) 446–454.
[2] Fangwei Qi, Feifei Huang, Tao Wang, Renguang Ye, Ruoshan Lei, Ying Tian, Junjie Zhang, Long Zhang, Shiqing Xu, Highly Er^{3+} doped fluorotellurite glass for 1.5 μm broadband amplification and 2.7 μm microchip laser applications, J. Lumin. 202 (2018) 132–135.
[3] Bryan M. van der Ende, Linda Aarts, Andries Meijerink, Lanthanide ions as spectral converters for solar cells, Phys. Chem. Chem. Phys. (2009) 11081–11095.
[4] Xiangfu Wang, Ye Wang, Jose Marques-Hueso, Xiaohong Yan, Improving optical temperature sensing performance of Er^{3+} doped Y_2O_3 microtubes via co-doping and controlling excitation power, Sci. Rep. 7 (2017) 758.
[5] Abdukader Abdukayum, Jia-Tong Chen, Qiang Zhao, Xiu-Ping Yan, Functional near infrared-emitting Cr^{3+}/Pr^{3+} Co-doped zinc gallogermanate persistent luminescent nanoparticles with superlong afterglow for in vivo targeted bioimaging, J. Am. Chem. Soc. 135 (2013) 14125–14133.
[6] Yuanzeng Min, Jinming Li, Fang Liu, Parasuraman Padmanabhan, K. Edwin, L. Yeow, Bengang Xing, Recent advance of biological molecular imaging based on lanthanide-doped upconversion-luminescent nanomaterials, Nanomaterials 4 (2014) 129–154.
[7] I. Sokólska, Spectroscopic characterisation of $LaGaO_3$:Er^{3+} crystals, Appl. Phys. B 71 (2000) 157–162.
[8] Madhab Pokhrel, G.A. Kumar, Sathravada Balaji, Radhaballabh Debnath, Dhiraj K. Sardar, Optical characterization of Er^{3+} and Yb^{3+} co-doped barium fluorotellurite glass, J. Lumin. 132 (2012) 1910–1916.
[9] Peixiong Zhang, Zhenqiang Chen, Yin Hang, Zhen Li, Hao Yin, Siqi Zhu, Shenhe Fu, Anming Li, Enhanced emission of the 1.50–1.67 μm fluorescence in Er^{3+}, Ce^{3+}–codoped $Lu_3Al_5O_{12}$ crystal, J. Alloys Compd. 696 (2017) 795–798.
[10] Muzhi Cai, Beier Zhou, Fengchao Wang, Ying Tian, Jiajia Zhou, Shiqing Xu, Junjie Zhang, Mid-infrared emission and quantitative analysis of energy transfer processes in Er^{3+} doped oxyfluogermanate glasses, J. Appl. Phys. 117 (2015), 243106.

[11] Peixiong Zhang, Zhenqiang Chen, Yin Hang, Zhen Li, Hao Yin, Siqi Zhu, Shenhe Fu, Anming Li, Enhanced 2.7 μm mid-infrared emissions of Er^{3+} via Pr^{3+} deactivation and Yb^{3+} sensitization in $LiNbO_3$ crystal, Optic Express 24 (2016) 25202–25210.

[12] Yanping Li, Jiahua Zhang, Xia Zhang, Yongshi Luo, Xinguang Ren, Haifeng Zhao, Xiaojun Wang, Lingdong Sun, Chunhua Yah, Near-infrared to visible upconversion in Er^{3+} and Yb^{3+} codoped Lu_2O_3 nanocrystals: enhanced red color upconversion and three-photon process in green color upconversion, J. Phys. Chem. C 113 (2009) 4413–4418.

[13] X.B. Chen, J.G. Wu, X. Xu, Y.Z. Zhang, N. Sawanobori, C.L. Zhang, Q.H. Pan, G.J. Salamo, Three-photon infrared quantum cutting from single species of rare-earth Er^{3+} ions in $Er_{0.3}Gd_{0.7}VO_4$ crystalline, Opt. Lett. 34 (2009) 887–889.

[14] Cao T.M. Dung, Le Van Hieu, Lam Quang Vinh, Tran T.T. Van, Remarkable enhancement of Er^{3+} emission at 1.54 μm in Er/Yb co-doped SiO_2-SnO_2 glass-ceramics, J. Alloys Compd. 757 (2018) 489–495.

[15] P. Le Boulanger, J.-L. Doualan, S. Girard, J. Margerie, R. Moncorgé, Excited-state absorption spectroscopy of -doped and phosphate glass, Phys. Rev. B 60 (1999) 11380–11390.

[16] Xue jun Gao, Wei Li, Xiaoliang Yang, Xiang liang Jin, Siguo Xiao, Near-infrared emission of Er^{3+} sensitized by Mn^{4+} in $Ca_{14}Zn_6Al_{10}O_{35}$ matrix, J. Phys. Chem. C 119 (2015) 28090–28098.

[17] P. Dorenbos, Relation between Eu^{2+} and Ce^{3+} f↔d-transition energies in inorganic compounds, J. Phys. Condens. Matter 15 (2003) 4797–4807.

[18] Arpita Vyas, C.P. Joshi, P.D. Sahare, S.V. Moharil, NIR emission in Ba_2SiO_4:Eu^{2+}, Nd^{3+} phosphors with near UV/violet excitation, J. Alloys Compd. 743 (2018) 789–794.

[19] R.A. Talewar, C.P. Joshi, S.V. Moharil, Sensitization of Nd^{3+} near infrared emission in Ca_2PO_4Cl host, J. Lumin. 197 (2018) 1–6.

[20] S.F. Zou, Z.L. Zhang, F. Zhang, Y.L. Mao, High efficient quantum cutting in Ce^{3+}/Yb^{3+} co-doped oxyfluoride glasses, J. Alloys Compd. 572 (2018) 110–112.

[21] Lege Wang, Zijian Yang, You-Fen Li, Ru Yang, Zhigang Dai, Shuqiao Hu, Liangliang Sun, Yaoyao Tong, Fluorescence resonance energy transfer of CaF_2: Eu^{2+}, Tb^{3+} applied to dye-sensitized solar cells, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 202 (2018) 76–80.

[22] Hang Lin, Daqin Chen, Yunlong Yu, Zhifa Shan, Ping Huang, Anping Yang, Yuansheng Wang, Broadband UV excitable near-infrared downconversion luminescence in Eu^{2+}/Yb^{3+}:CaF_2 nanocrystals embedded glass ceramics, J. Alloys Compd. 509 (2011) 3363–3366.

[23] Peixiong Zhang, Zhenqiang Chen, Yin Hang, Zhen Li, Hao Yin, Siqi Zhu, Shenhe Fu, Anming Li, Enhanced emission of the 1.50–1.67 μm fluorescence in Er^{3+}, Ce^{3+}–codoped $Lu_3Al_5O_{12}$ crystal, J. Alloys Compd. 696 (2017) 795–798.

[24] Xinyang Huang, Liang Jiang, Xiaoxia Li, Aiqun He, Manipulating upconversion emission of cubic $BaGdF_5$:Ce^{3+}/Er^{3+}/Yb^{3+} nanocrystals through controlling Ce^{3+} doping, J. Alloys Compd. 721 (2017) 374–382.

[25] Yaxun Zhou, Na Gai, Jun Wang, Comparative investigation on spectroscopic properties of Er^{3+} between Ce^{3+}-doped and B_2O_3-added bismuth glasses, J. Phys. Chem. Solid. 70 (2009) 261–265.

[26] Shichao Zheng, Yaxun Zhou, Dandan Yin, Xingchen Xu, Yawei Qi, Shengxi Peng, Improvement of 1.53μm band fluorescence and energy transfer in Er^{3+}/Ce^{3+} codoped tellurite glasses, J. Alloys Compd. 566 (2013) 90–97.

[27] T. Förster, Improvement of 1.53 μm band fluorescence and energy transfer in Er^{3+}/Ce^{3+} codoped tellurite glasses, Ann. Phys. 437 (1948) 55–75.

[28] D.L. Dexter, A theory of sensitized luminescence in solids, J. Chem. Phys. 21 (1953) 836–850.

[29] Jiajia Zhou, Yu Teng, Xiaofeng Liu, Song Ye, Xiaoqiu Xu, Zhijun Ma, Jianrong Qiu, Intense infrared emission of Er^{3+} in $Ca_8Mg(SiO_4)_4Cl_2$ phosphor from energy transfer of Eu^{2+} by broadband down-conversion, Optic Express 18 (2010) 21663–21668.

[30] Oscar L. Malta, Luís D. Carlos, Intensities of 4f-4f transitions in glass materials, Quim. Nova 26 (2003) 889–895.

[31] P. Dorenbos, The 5d level positions of the trivalent lanthanides in inorganic compounds, J. Lumin. 91 (2000) 155–176.

[32] R.A. Talewar, C.P. Joshi, S.V. Moharil, Near infrared emission and energy transfer in Eu^{2+}-Nd^{3+} co-doped Ca_2BO_3Cl, Opt. Mater. 55 (2016) 44–48.

[33] R.F. Smirnova, I.M. Rumanova, N.V. Belov, Crystal structure of cuspidine, Zap. Vses. Miner. O-va 84 (1955) 159–169.

[34] Yongchao Jia, Hui Qiao, Ning Guo, Yuhua Zheng, Mei Yang, Yeju Huang, Hongpeng You, Electronic structure and photoluminescence properties of Eu^{2+}-activated $Ca_4Si_2O_7F_2$, Opt. Mater. 33 (2011) 1803–1807.

[35] Mubiao Xie, Guoxian Zhu, Dongyu Li, Rongkai Pan, Xiaoping Zhou, Wei Xie, Synthesis and photoluminescence properties of tunable emission phosphors $Ca_4Si_2O_7F_2$:Ce^{3+}, Solid State Sci. 48 (2015) 193–196.

[36] Chien-Hao Huang, Ting-Shan Chan, Wei-Ren Liu, De-Yin Wang, Yi-Chen Chiu, Yao-Tsung Yeha, Teng-Ming Chen, Crystal structure of blue–white–yellow color-tunable $Ca_4Si_2O_7F_2$:Eu^{2+},Mn^{2+} phosphor and investigation of color tunability through energy transfer for single-phase white-light near-ultraviolet LEDs, J. Mater. Chem. 22 (2012) 20210–20216.

[37] Wei Lü, Yongshi Luo, Zhendong Hao, Xia Zhang, Xiaojun Wang, Jiahua Zhang, A new dual-emission phosphor $Ca_4Si_2O_7F_2$:Ce^{3+}, Mn^{2+} with energy transfer for near-UV LEDs, Mater. Lett. 77 (2012) 45–47.

[38] L. Keller, J. Rask, P. Buseck, Arizona State University, ICDD Grant-in-Aid, Tempe, Arizona, 1989.

[39] S. Saburi, A. Kawahara, C. Henmi, I. Kusachi, K. Kihara, The refinement of the crystal structure of cuspidine, Mineral. J. 8 (1977) 286–298.

[40] R.D. Shannon, Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides, Acta Crystallogr. A32 (1976) 751–767.

[41] Panlai Li, Zhijun Wang, Zhiping Yang, Qinglin Guo, Xu Li, Luminescent characteristics of $LiCaBO_3$:M (M=Eu^{3+}, Sm^{3+}, Tb^{3+}, Ce^{3+}, Dy^{3+}) phosphor for white LED, J. Lumin. 130 (2010) 222–225.

[42] Zhiguo Xia, Ru-Shi Liu, Tunable blue-green color emission and energy transfer of $Ca_2Al_3O_6F$:Ce^{3+},Tb^{3+} phosphors for near-UV white LEDs, J. Phys. Chem. C 116 (2012) 15604–15609.

[43] P. Dorenbos, Relating the energy of the configuration of in inorganic compounds with anion polarizability and cation electronegativity, Phys. Rev. B 65 (2002), 235110.

[44] Bo Cui, Qinghong Zhang, Hongzhi Wang, Yaogang Li, A single-phase full-color $Ca_4Si_2O_7F_2$:Ce^{3+}, Tb^{3+}, Sm^{3+} micron belt mat for white-light emitting diode application, Key Eng. Mater. 697 (2016) 727–732.

[45] R.A. Talewar, Sk. Mahamuda, A.S. Rao, C.P. Joshi, S.V. Moharil, Sensitization of Nd^{3+} by 4f-5d transition of Ce^{3+} in $Ba_2Y(BO_3)_2Cl$ phosphor for the prospective NIR applications, J. Lumin. 202 (2018) 1–6.

[46] Dhiraj K. Sardar, Charles C. Russell, Raylon M. Yow, John B. Gruber, Bahram Zandi, Edvard P. Kokanyan, Spectroscopic analysis of the Er^{3+} ($4f^{11}$) absorption intensities in $NaBi(WO_4)_2$, J. Appl. Phys. 95 (2004) 1180–1184.

[47] A.A. Kaminskii, S.E. Sarkisov, F. Below, H.J. Eichler, Spectroscopic and laser properties of Er^{3+}-doped monoclinic BaY2F8 single crystals, Opt. Quant. Electron. 22 (1990) S95–S105.

[48] V.K. Tikhomirov, D. Furniss, A.B. Seddon, M. Ferrari, R. Rolli, Er^{3+} doped ultra-transparent oxy-fluoride glass-ceramics for application in the 1.54 μm telecommunication window, J. Mater. Sci. Lett. 21 (2002) 293–295.

[49] Wenqin Luo, Jinsheng Liao, Renfu Li, Xueyuan Chen, Determination of Judd–Ofelt intensity parameters from the excitation spectra for rare-earth doped luminescent materials, Phys. Chem. Chem. Phys. 12 (2010) 3276–3282.

[50] Sreerenjini Chandra, Leonard Deepak Francis, John B. Gruber, Dhiraj K. Sardar, Synthesis, morphology, and optical characterization of nanocrystalline Er^{3+}: Y_2O_3, J. Phys. Chem. C 114 (2010) 874–880.

[51] G. Blasse, A. Bril, Investigation of some Ce^{3+}-activated phosphors, J. Chem. Phys. 47 (12) (1967) 5139–5145.

[52] Rui Zhang, Boyang Wang, Wenfeng Zhu, Cunjun Li, Hai Wang, Preparation and luminescent performances of transparent screen-printed Ce^{3+}:$Y_3Al_5O_{12}$ phosphors-in-glass thick films for remote white LEDs, J. Alloys Compd. 720 (2017) 340–344.

[53] Rupesh A. Talewar, Sk. Mahamuda, K. Swapna, M. Venkateswarlu, A.S. Rao, Spectroscopic studies of Sm^{3+} ions doped alkaline-earth chloro borate glasses for visible photonic applications, Mater. Res. Bull. 105 (2018) 45–54.

[54] Rupesh A. Talewar, Sk. Mahamuda, K. Swapna, A.S. Rao, Sensitization of Yb^{3+} by Nd^{3+} emission in alkaline-earth chloro borate glasses for laser and fiber amplifier applications, J. Alloys Compd. 771 (2019) 980–986.

[55] P.K. Tawalare, V.B. Bhatkar, R.A. Talewar, C.P. Joshi, S.V. Moharil, Host sensitized NIR emission in rare-earth doped $NaY(MoO_4)_2$ phosphors, J. Alloys Compd. 732 (2018) 64–69.

[56] G. Blasse, B.C. Grabmaier, Luminescent Materials, Springer-Verlag, Berlin, 1994, pp. 91–93.

[57] G. Blasse, Energy transfer in oxidic phosphors, Philips Res. Rep. 24 (1969) 131–144.

[58] M. Inokuti, F. Hirayama, Influence of energy transfer by the exchange mechanism on donor luminescence, J. Chem. Phys. 43 (1965) 1978–1989.

View publication stats

arXiv:1810.09361v1 [cs.MS] 22 Oct 2018

ensmallen: a flexible C++ library for efficient function optimization

Shikhar Bhardwaj
Delhi Technological University
Delhi, India 110042
shikhar_bt2k15@dtu.ac.in

Ryan R. Curtin
RelationalAI
Atlanta, GA, USA 30318
ryan@ratml.org

Marcus Edel
Free University of Berlin
Arnimallee 7, 14195 Berlin
marcus.edel@fu-berlin.de

Yannis Mentekidis
Independent Researcher
mentekid@gmail.com

Conrad Sanderson
Data61, CSIRO, Australia
University of Queensland, Australia
conradsand@ieee.org

Abstract

We present ensmallen, a fast and flexible C++ library for mathematical optimization of arbitrary user-supplied functions, which can be applied to many machine learning problems. Several types of optimizations are supported, including differentiable, separable, constrained, and categorical objective functions. The library provides many pre-built optimizers (including numerous variants of SGD and Quasi-Newton optimizers) as well as a flexible framework for implementing new optimizers and objective functions. Implementation of a new optimizer requires only one method and a new objective function requires typically one or two C++ functions. This can aid in the quick implementation and prototyping of new machine learning algorithms. Due to the use of C++ template metaprogramming, ensmallen is able to support compiler optimizations that provide fast runtimes. Empirical comparisons show that ensmallen is able to outperform other optimization frameworks (like Julia and SciPy), sometimes by large margins. The library is distributed under the BSD license and is ready for use in production environments.

1 Introduction

Mathematical optimization is the workhorse of virtually all machine learning algorithms. For a given objective function $f(\cdot)$ (which may have a special structure or constraints), almost all machine learning problems can be boiled down to the following optimization form:

$$\underset{x}{\operatorname{argmin}} f(x). \tag{1}$$

Optimization is often computationally intensive and may correspond to most of the time taken to train a machine learning model. For instance, the training of deep neural networks is dominated by the optimization of the model parameters on the data [31]. Even popular machine learning models such as logistic regression have training times mostly dominated by an optimization procedure [13].

The ubiquity of optimization in machine learning algorithms highlights the need for robust and flexible implementations of optimization algorithms. We present ensmallen, a C++ optimization toolkit that contains a wide variety of optimization techniques for many types of objective functions. Through the use of C++ template metaprogramming [2, 34], ensmallen is able to generate efficient code that can help with the demanding computational needs of many machine learning algorithms.

Although there are many existing machine learning optimization toolkits, few are able to take explicit advantage of metaprogramming based code optimizations, and few offer robust support for various types of objective functions. For instance, deep learning libraries like Caffe [11], PyTorch [24], and TensorFlow [1] each contain a variety of optimization techniques. However, these techniques

are limited to stochastic gradient descent (SGD) and SGD-like optimizers that operate on small batches of data points at a time. Other machine learning libraries, such as `scikit-learn` [25] contain optimization algorithms but not in a coherent or reusable framework. Many programming languages have higher-level packages for mathematical optimization. For example, `scipy.optimize` [12], is widely used in the Python community, and MATLAB's function optimization support has been available and used for many decades. However, these implementations are often unsuitable for modern machine learning tasks—for instance, computing the full gradient of the objective function may not be feasible because there are too many data points.

In this paper, we describe the functionality of `ensmallen` and the types of problems that it can be applied to. We discuss the mechanisms by which `ensmallen` is able to provide both computational efficiency and ease-of-use. We show a few examples that use the library, as well as empirical performance comparisons with other optimization libraries.

`ensmallen` is open-source software licensed under the 3-clause BSD license [27], allowing unencumbered use in both open-source and proprietary projects. It is available for download from `https://ensmallen.org`. Armadillo [30] is used for efficient linear algebra operations, with optional interface to GPUs via NVBLAS [23].

2 Types of Objective Functions

`ensmallen` provides a **set of optimizers** for optimizing **user-defined objective functions**. It is also easy to implement a new optimizer in the `ensmallen` framework. Overall, our goal is to provide an easy-to-use library that can solve the problem $\operatorname{argmin}_x f(x)$ for any function $f(x)$ that takes a vector or matrix input x. In most cases, $f(x)$ will have special structure; one example might be that $f(x)$ is differentiable. Therefore, the abstraction we have designed for `ensmallen` can optionally take advantage of this structure. For example, in addition to $f(x)$, a user can provide an implementation of $f'(x)$, which in turn allows first-order gradient-based optimizers to be used. This generally leads to significant speedups.

There are a number of other properties that `ensmallen` can use to accelerate computation. These are listed below:

- **arbitrary**: no assumptions can be made on $f(x)$
- **differentiable**: $f(x)$ has a computable gradient $f'(x)$
- **separable**: $f(x)$ is a sum of individual components: $f(x) = \sum_i f_i(x)$
- **categorical**: x contains elements that can only take discrete values
- **sparse**: the gradient $f'(x)$ or $f'_i(x)$ (for a separable function) is sparse
- **partially differentiable**: the separable gradient $f'_i(x)$ is also separable for a different axis j
- **bounded**: x is limited in the values that it can take

	unified framework	constraints	batches	arbitrary functions	arbitrary optimizers	sparse gradients	categorical
`ensmallen`	●	●	●	●	●	●	●
Shogun [33]	●	-	●	●	●	-	-
Vowpal Wabbit [16]	-	-	●	-	-	-	●
TensorFlow [1]	●	-	●	◐	-	◐	-
Caffe [11]	●	-	●	◐	◐	-	-
Keras [4]	●	-	●	◐	◐	-	-
scikit-learn [25]	◐	-	◐	◐	-	-	-
SciPy [12]	●	●	-	●	-	-	-
MATLAB [19]	●	●	-	●	-	-	-
Julia (`Optim.jl`) [20]	●	-	-	●	-	-	-

Table 1: Feature comparison: ●= provides feature, ◐= partially provides feature, - = does not provide feature. *unified framework* indicates if there is some kind of generic/unified optimization framework; *constraints* and *batches* indicate support for constrained problems and batches; *arbitrary functions* means arbitrary objective functions are easily implemented; *arbitrary optimizers* means arbitrary optimizers are easily implemented; *sparse gradient* indicates that the framework can natively take advantage of sparse gradients; and *categorical* refers to if support for categorical features exists.

Due to its straightforward abstraction framework, `ensmallen` is able to provide a large set of diverse optimization algorithms for objective functions with these properties. Below is a list of currently available optimizers:

- **SGD variants** [29]: Stochastic Gradient Descent (SGD), SGD with Restarts, Parallel SGD (Hogwild!) [22], Stochastic Coordinate Descent, AdaGrad [6], AdaDelta [35], RMSProp, SMORMS3, Adam [13], AdaMax
- **Quasi-Newton variants**: Limited-memory BFGS (L-BFGS) [36], incremental Quasi-Newton method [21], Augmented Lagrangian Method [10]
- **Genetic variants**: Conventional Neuro-evolution [3], Covariance Matrix Adaptation Evolution Strategy [9]
- **Other**: Conditional Gradient Descent, Frank-Wolfe algorithm [8], Simulated Annealing [14]

In `ensmallen`'s framework, if a user wants to optimize a differentiable objective function, they only need to provide implementations of $f(x)$ and $f'(x)$, and then they can use any of the gradient-based optimizers that `ensmallen` provides. Table 1 contrasts the classes of objective functions that can be handled by `ensmallen` and other popular frameworks and libraries.

Not every optimization algorithm provided by `ensmallen` can be used by every class of objective function; for instance, a gradient-based optimizer such as L-BFGS cannot operate on a non-differentiable objective function. As such, the best the library can attain is to maximize the flexibility available, so that a user can easily implement a function $f(x)$ and have it work with as many optimizers as possible.

To accomplish the flexibility, `ensmallen` makes heavy use of C++ template metaprogramming. When implementing an objective function to be optimized, a user may only need to implement a few methods; metaprogramming is then automatically used to check that the given functions match the requirements of the optimizer that is being used. When implementing an optimizer, we can assume that the given function to be optimized meets the required assumptions of the optimizers, and encode those requirements as compile-time checks (via `static_assert`).

For the most common case of a differentiable $f(x)$, the user only needs to implement two methods:

- `double Evaluate(`x`)`: given coordinates x, this function returns the value of $f(x)$.
- `void Gradient(`x`, `g`)`: given coordinates x and a reference to g, set $g = f'(x)$.

Alternatively, the user can simply implement a `EvaluateWithGradient()` function that computes both $f(x)$ and $f'(x)$ simultaneously, which is useful in cases where both the objective and gradient depend on similar computations.

The required API for separable differentiable objective functions (i.e. those that would use an optimizer like SGD) is very similar, except that `Evaluate()`, `Gradient()` and `EvaluateWithGradient()` should operate only on mini-batches, and utility methods `Shuffle()` and `NumFunctions()` must be added. The same pattern applies for other types of objective functions: only a few methods specific to class of objective function itself must be implemented and then any optimizer may be used.

3 Example: Learning Linear Regression Models

As an example of usage, consider the linear regression objective function[1]. Given a dataset $X \in \mathcal{R}^{n \times d}$ and associated responses $y \in \mathcal{R}^n$, the model of linear regression is to assume that $y_i = x_i\theta$ for each point and response (x_i, y_i). To fit this model $\theta \in \mathcal{R}^d$ to the data, we must find

$$\underset{\theta}{\operatorname{argmin}} f(\theta) = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - x_i\theta)^2 = \underset{\theta}{\operatorname{argmin}} \|y - X\theta\|_F^2. \tag{2}$$

The objective function $f(\theta)$ has the associated gradient

$$f'(\theta) = \sum_{i=1}^{n} -2x_i(y_i - x_i\theta) = -2X^T(y - X\theta). \tag{3}$$

We can implement these two functions in a class named `LinearRegressionFunction`, as shown in Fig. 1. This is the entire required implementation to optimize the linear regression model with any of the gradient-based optimizers in `ensmallen`.

[1] For simplicity, we ignore the bias term. It can be rederived by taking $x_i^* = (x_i, 1)$.

```
class LinearRegressionFunction {
 public:
  // Construct the LinearRegressionFunction with the given data
  LinearRegressionFunction(arma::mat& X_in, arma::vec& y_in) : X(X_in), y(y_in) {}

  // Compute the objective function
  double Evaluate(const arma::mat& theta) {
    return std::pow(arma::norm(y - X * theta), 2.0);
  }
  // Compute the gradient and store in 'gradient'
  void Gradient(const arma::mat& theta, arma::mat& gradient) {
    gradient = -2 * X.t() * (y - X * theta);
  }
 private:
  arma::mat& X; arma::vec& y;
};
```

Figure 1: Implementation of objective and gradient functions for linear regression, used by optimizers in `ensmallen`. The types `arma::mat` and `arma::vec` are matrix and vector types from Armadillo [30].

Given the user-defined `LinearRegressionFunction` class, the code snippet below shows how the L-BFGS optimizer can be used to find the best parameters θ:

```
LinearRegressionFunction lrf(X, y); // we assume X and y already hold data
ens::L_BFGS lbfgs; // create L-BFGS optimizer with default parameters

arma::vec theta(X.n_rows, arma::fill::randu); // random uniform initialization
lbfgs.Optimize(lrf, theta); // after this call, theta holds the solution
```

To use the small-batch SGD-like optimizers provided by `ensmallen`, only a slight variation on the signature of `Evaluate()` and `Gradient()` would be needed, plus the `Shuffle()` and `NumFunctions()` utility methods.

4 Automatic Metaprogramming for Ease of Use and Efficiency

When optimizing a given function $f(x)$, the computation of the objective function $f(x)$ and its derivative $f'(x)$ often involve the computation of identical intermediate results. Consider the linear regression objective function described in Section 3, $f(\theta) = \|y - X\theta\|_F^2$. For this objective function, the derivative $f'(\theta)$ has a related form of $-2X^T(y - X\theta)$. Both the objective function and the derivative depend on the computation of the vector term $(y - X\theta)$, which can be computationally expensive to compute if X is a large matrix. Existing optimization frameworks do not have an easy way to avoid this duplicate computation. In many cases, an optimization algorithm may need the values of both $f(\theta)$ and $f'(\theta)$ for a given θ.

Using template metaprogramming, `ensmallen` provides an easy (and optional) way for users to avoid this extra computational overhead. Instead of specifying individual `Evaluate()` and `Gradient()` functions, a user may simply write an `EvaluateWithGradient()` function that returns both the objective value and the gradient value for an input θ. As an example, for the `LinearRegressionFunction` class in Fig. 1, we can replace `Evaluate()` and `Gradient()` with an implementation of `EvaluateWithGradient()` that computes $(y - X\theta)$ only once:

```
double EvaluateWithGradient(const arma::mat& theta, arma::mat& gradient) {
  const arma::vec v = (y - X * theta); // Cache result
  gradient = -2 * X.t() * v;  // Store gradient in the provided matrix
  return arma::accu(v % v); // Take squared norm of v
}
```

Template metaprogramming techniques are automatically used to detect which methods exist, and a wrapper class will use suitable mix-ins in order to provide 'missing' functionality [32]. For instance, if `EvaluateWithGradient()` is not provided, a version will be automatically generated that calls both `Gradient()` and `Evaluate()` in turn. Similarly, if `Evaluate()` or `Gradient()` does not exist, then `EvaluateWithGradient()` is called, and the unnecessary part of the result will be discarded.

The use of template metaprogramming in this manner also allows for compiler optimizations that would not otherwise be possible (and that are often not possible in other frameworks). Firstly, because the objective function class itself is a template parameter, the compiler is able to avoid the overhead of a function pointer dereference, which would not be easily possible when using a language with virtual

	ensmallen	scipy	Optim.jl	samin
default	**0.004s**	1.069s	0.021s	3.173s
tuned		0.574s		3.122s

Table 2: Runtimes for 100000 iterations of simulated annealing with various frameworks on the simple Rosenbrock function. Julia code runs do not count compilation time. The *tuned* row indicates that the code was manually modified for speed.

inheritance. The compiler is also able to use inlining and any optimizations that may imply, including removing temporary values and dead code elimination. Further, if `ensmallen` automatically generates an `Evaluate()` or `Gradient()` method from a user-supplied `EvaluateWithGradient()` method, the compiler can in some cases recognize and remove the computation of unnecessary results. For instance, in an automatically generated `Evaluate()` method, the computation of the gradient from `EvaluateWithGradient()` can be avoided entirely.

Overall, the automatic code generation functionality in `ensmallen` reduces the requirements for users when they are implementing their own objective functions to be optimized, and allows users a way to provide more efficient implementations of their objective functions. This leads to quicker development, quicker results, and reduces the likelihood of bugs.

At the time of writing, the automatic code generation is implemented for the most commonly-used cases: full-batch and small-batch `Evaluate()`, `Gradient()`, and `EvaluateWithGradient()`. We aim to expand this support to other sets of methods for other types of objective functions.

5 Experiments

To demonstrate the benefits of the metaprogramming based code optimizations that `ensmallen` can exploit, we compare the performance of `ensmallen` with several other optimization frameworks, including some that use automatic differentiation.

For our first experiment, we consider the simple and popular Rosenbrock function [28]: $f([x_1, x_2]) = 100(x_2 - x_1^2)^2 + (1 - x_1^2)$. For the optimizer, we use simulated annealing [14], a gradient-free optimizer. Simulated annealing will call the objective function numerous times; for each simulation we limit the optimizer to 100k objective evaluations. Since the objective function is straightforward and is called many times, this can help us understand the overheads of various frameworks.

We compare four frameworks for this task: **(i)** `ensmallen`, **(ii)** `scipy.optimize.anneal` from SciPy 0.14.1 [12], **(iii)** simulated annealing implementation in `Optim.jl` with Julia 1.0.1 [20], **(iv)** `samin` in the `optim` package for Octave [7]. While another option here might be `simulannealbnd()` in the Global Optimization Toolkit for MATLAB, no license was available. We ran our code on a MacBook Pro i7 2018 with 16GB RAM running macOS 10.14 with clang 1000.10.44.2, Julia version 1.0.1, Python 2.7.15, and Octave 4.4.1.

Initially, we implemented these functions as simply as possible and ran them without any tuning. This reflects how a typical user might interact with a given framework. The results are shown in the first row of Table 2. Only Julia and `ensmallen` are compiled, and thus are able to avoid the function pointer dereference and take advantage of inlining and related optimizations.

However, both Python and Octave have routes for acceleration, such as Numba [15], MEX bindings and JIT compilation. We hand-optimized the Rosenbrock implementation using Numba, which required significant modification of the underlying `anneal.anneal()` function. These techniques did produce some speedup, as shown in the second row of Table 2. For Octave, a MEX binding did not produce a noticeable difference. We were unable to tune either `ensmallen` or `Optim.jl` to get any speedup, suggesting that novice users will easily be able to write efficient code in these cases.

Next, we consider the linear regression example described in Sec. 3. For this task we use the first-order L-BFGS optimizer [36]. Using the same four packages, we implement the linear regression objective and gradient. For `ensmallen` we implement a version with only `EvaluateWithGradient()`, denoted as `ensmallen-1`. We also implement a version with both `Evaluate()` and `Gradient()`: `ensmallen-2`. We also use automatic differentiation for Julia via the `ForwardDiff.jl` [26] package and for Python via the `Autograd` [18] package. For GNU Octave we use the `bfgsmin()` function.

algorithm	d: 100, n: 1k	d: 100, n: 10k	d: 100, n: 100k	d: 1k, n: 100k
`ensmallen-1`	**0.001s**	**0.009s**	**0.154s**	**2.215s**
`ensmallen-2`	0.002s	0.016s	0.182s	2.522s
`Optim.jl`	0.006s	0.030s	0.337s	4.271s
`scipy`	0.003s	0.017s	0.202s	2.729s
`bfgsmin`	0.071s	0.859s	23.220s	2859.81s
`ForwardDiff.jl`	0.497s	1.159s	4.996s	603.106s
`autograd`	0.007s	0.026s	0.210s	2.673s

Table 3: Runtimes for the linear regression function on various dataset sizes, with n indicating the number of samples, and d indicating the dimensionality of each sample. All Julia runs do not count compilation time.

Results for various data sizes are shown in Table 3. For each implementation, L-BFGS was allowed to run for only 10 iterations and never converged in fewer iterations. The datasets used for training are highly noisy random data with a slight linear pattern. Note that the exact data is not relevant for the experiments here, only its size. Runtimes are reported as the average across 10 runs.

The results indicate that `ensmallen` with `EvaluateWithGradient()` is the fastest approach. Furthermore, the use of `EvaluateWithGradient()` yields considerable speedup over the `ensmallen-2` implementation with both the objective and gradient implemented separately. In addition, although the automatic differentiation support makes it easier for users to write their code, we observe that the output of automatic differentiators is not as efficient, especially with `ForwardDiff.jl`. We expect this effect to be more pronounced with increasingly complex objective functions.

Lastly, we demonstrate the easy pluggability in `ensmallen` for using various optimizers on the same task. Using a version of `LinearRegressionFunction` from Sec. 3 adapted for separable differentiable optimizers, we run six optimizers with default parameters in just 8 lines of code, as shown in Fig. 2(a). Applying these optimizers to the `YearPredictionMSD` dataset from the UCI repository [17] yields the learning curves shown in Fig. 2(b). Any other optimizer for separable differentiable objective functions can be dropped into place in the same manner.

6 Conclusion

We have described `ensmallen`, a flexible C++ library for function optimization that provides an easy interface for the implementation and optimization of user-defined objective functions. Many types of functions can be optimized, including separable and constrained functions. The library provides many pre-built optimizers (including numerous variants of SGD and Quasi-Newton optimizers). The library internally exploits template metaprogramming to maximise opportunities for efficiency gains, as well as to make the implementation of objective functions easier by automatically generating missing methods.

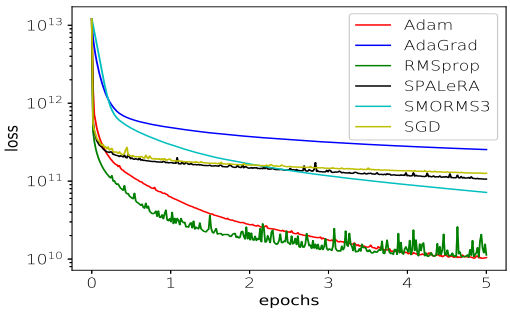
We aim to expand the library with further optimization techniques as the need arises. Since `ensmallen` is open source, external contributions to the codebase are welcome. For more information on the library, see the website and documentation at `https://ensmallen.org`. The library is already in use for function optimization in the `mlpack` machine learning toolkit [5].

Acknowledgements. We would like to thank the many contributors to `ensmallen`, who are listed on the associated website.

```
// X and y are data
LinearRegressionFunction lrf(X, y);

using namespace ens;
StandardSGD<>().Optimize(lrf, sgdModel);
Adam().Optimize(lrf, adamModel);
AdaGrad().Optimize(lrf, adagradModel);
SMORMS3().Optimize(lrf, smorms3Model);
SPALeRASGD().Optimize(lrf, spaleraModel);
RMSProp().Optimize(lrf, rmspropModel);
```

(a) Code.



(b) Learning curves.

Figure 2: Example usage of six `ensmallen` optimizers to optimize a linear regression function on the `YearPredictionMSD` dataset [17] for 5 epochs of training. The optimizers can be tuned for better performance.

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv:1603.04467*, 2016.

[2] D. Abrahams and A. Gurtovoy. *C++ Template Metaprogramming: Concepts, Tools, and Techniques from Boost and Beyond*. Addison-Wesley Professional, 2004.

[3] R. K. Belew, J. McInerney, and N. N. Schraudolph. Evolving networks: Using the genetic algorithm with connectionist learning, 1990. CSE Technical Report #CS90-174, University of California at San Diego.

[4] F. Chollet. Keras. https://github.com/fchollet/keras, 2015.

[5] R. R. Curtin, M. Edel, M. Lozhnikov, Y. Mentekidis, S. Ghaisas, and S. Zhang. mlpack 3: a fast, flexible machine learning library. *Journal of Open Source Software*, 3:726, 2018.

[6] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159, 2011.

[7] J. W. Eaton, D. Bateman, S. Hauberg, and R. Wehbring. *GNU Octave version 4.4.0 manual: a high-level interactive language for numerical computations*, 2018.

[8] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956.

[9] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[10] M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4(5):303–320, 1969.

[11] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *ACM International Conference on Multimedia*, pages 675–678, 2014.

[12] E. Jones, T. Oliphant, and P. Peterson. SciPy: open source scientific tools for Python, 2014. http://www.scipy.org/.

[13] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

[14] S. Kirkpatrick, C. G. Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[15] S. K. Lam, A. Pitrou, and S. Seibert. Numba: A LLVM-based python JIT compiler. In *Second Workshop on the LLVM Compiler Infrastructure in HPC*, page 7, 2015.

[16] J. Langford, L. Li, and A. Strehl. Vowpal wabbit open source project. Technical report, Yahoo!, 2007.

[17] M. Lichman. UCI Machine Learning Repository, 2013. http://archive.ics.uci.edu/ml.

[18] D. Maclaurin, D. Duvenaud, and R. P. Adams. Autograd: Effortless gradients in numpy. In *AutoML Workshop at ICML*, 2015.

[19] Mathworks. Matlab optimization toolbox, 2017. The MathWorks, Natick, MA, USA.

[20] P. K. Mogensen and A. N. Riseth. Optim: A mathematical optimization package for Julia. *Journal of Open Source Software*, 3(24):615, 2018.

[21] A. Mokhtari, M. Eisen, and A. Ribeiro. IQN: An incremental quasi-newton method with local superlinear convergence rate. *SIAM Journal on Optimization*, 28(2):1670–1698, 2018.

[22] F. Niu, B. Recht, C. Ré, and S. J. Wright. HOGWILD!: A lock-free approach to parallelizing stochastic gradient descent. In *Advances in Neural Information Processing Systems*, 2011.

[23] NVIDIA. NVBLAS Library. http://docs.nvidia.com/cuda/nvblas, 2015.

[24] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in PyTorch. *Autodiff Workshop at NIPS*, 2017.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[26] J. Revels, M. Lubin, and T. Papamarkou. Forward-mode automatic differentiation in Julia. *arXiv:1607.07892*, 2016.

[27] L. Rosen. *Open Source Licensing: Software Freedom and Intellectual Property Law*. Prentice Hall, 2004.

[28] H. H. Rosenbrock. An automatic method for finding the greatest or least value of a function. *The Computer Journal*, 3:175–184, 1960.

[29] S. Ruder. An overview of gradient descent optimization algorithms. *arXiv:1609.04747*, 2016.

[30] C. Sanderson and R. Curtin. Armadillo: a template-based C++ library for linear algebra. *Journal of Open Source Software*, 1:26, 2016.

[31] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[32] Y. Smaragdakis and D. Batory. Mixin-based programming in C++. In *International Symposium on Generative and Component-Based Software Engineering, Lecture Notes in Computer Science*, volume 2177, pages 164–178. Springer, 2000.

[33] S. Sonnenburg, G. Rätsch, S. Henschel, C. Widmer, J. Behr, A. Zien, F. de Bona, A. Binder, C. Gehl, and V. Franc. The SHOGUN machine learning toolbox. *Journal of Machine Learning Research*, 11:1799–1802, 2010.

[34] D. Vandevoorde and N. M. Josuttis. *C++ Templates: The Complete Guide*. Addison-Wesley, 2nd edition, 2017.

[35] M. D. Zeiler. ADADELTA: An adaptive learning rate method. *arXiv:1212.5701*, 2012.

[36] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560, 1997.

Materials Research Express

ACCEPTED MANUSCRIPT

Experimental investigation on the mechanical properties and water absorption behavior of randomly oriented short pineapple/coir fiber-reinforced hybrid epoxy composites

To cite this article before publication: Mohit Mittal *et al* 2018 *Mater. Res. Express* in press https://doi.org/10.1088/2053-1591/aae944

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2018 IOP Publishing Ltd**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.
As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.107 on 25/10/2018 at 10:58

Experimental Investigation on the Mechanical Properties and Water Absorption Behavior of Randomly Oriented Short Pineapple/Coir Fiber-Reinforced Hybrid Epoxy Composites

Mohit Mittal [a], Rajiv Chaudhary [a, *]

[a] *Department of Mechanical, Production, and Industrial Engineering, Delhi Technological University, Delhi, 110042, India*

[*] Corresponding author. *E-mail address*: **rch_dce@rediffmail.com**; Tel. +91-9868094086.

Abstract

To elevate the progressive utilization of natural fiber reinforced composites in structural applications, it is extremely required to heighten the properties of NFRC's by the exploitation of hybridization tool. In most of the cases, the mechanical properties of hybrid composites reinforced with natural fibers are far better than the pure single fiber reinforced composites and comparable to the synthetic fiber reinforced composites. Therefore in this research work, an experimental investigation has been carried out to study the mechanical properties and water absorption behavior of hybridized PALF/COIR fiber reinforced epoxy matrix composites. The biocomposite sheets consist of randomly oriented and intimately mixed short fibers (20 mm in length) were fabricated by employing the hand lay-up technique at 11 levels of COIR fiber loading (0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 vol.%) with fixed total fiber content (40 vol.%). The mechanical properties (tensile strength, tensile modulus, flexural strength, flexural modulus, and impact strength) and water uptake behavior with water diffusion mechanism of various formulated composites have been studied according to ASTM standard. The total of four samples for each composite specimen was tested and their average values were reported. The results showed the improvement in tensile and flexural properties (strength and modulus) of COIR/Epoxy composite by the hybridization with pineapple leaf fiber, up to 50 vol.%. The hybrid composite reinforced with equal volume content of pineapple leaf fiber (PALF) and coconut husk fiber (COIR) possess the highest tensile, flexural, and impact strength. Water absorption test revealed the decrement in sorption affinity with the increase of coir fiber content. The hybrid composite (P50-C50) absorbs 62% and 32% less water than that of pure PALF/Epoxy and COIR/Epoxy composites.

Accepted Manuscript

Keywords: pineapple leaf fiber, coir fiber, epoxy matrix, mechanical properties, water absorption, polymer composite

1. Introduction

Recent trends showed that the industries (automotive, packaging, building and construction, railway, and marine) are looking for the recyclable and bio-compatible materials having the properties of lightweight, low price, renewability, easy availability, easy to process into simple and complex shape, and satisfactory mechanical properties. To execute the demand of industries; researchers and technologists are putting their valuable efforts to consume the advantages of agricultural wastes and agro-industrial residues for the development of a high-performance recyclable material. In current days, the natural fiber reinforced composites (NFRC's) are enjoying and received a considerable attention from both industries and academicians in comparison to the traditional organic (carbon and aramid) and inorganic (glass and alumina) synthetic fiber reinforced composite materials. This was mainly attributed to the recyclability, high specific strength and modulus, lower energy requirements for processing, and excellent acoustic, electrical, and thermal insulating properties. Among all the natural cellulosic fibers (hemp, flax, sisal, coir, banana, bagasse, pineapple leaf fiber etc.), the coconut husk fiber (COIR) and pineapple leaf fiber (PALF) have high potential to effectively reinforce polymer matrices for the development of a high strength biocomposite material. The coir and pineapple leaf fiber can obtain from the husk of coconut fruit (*Cocos nucifera*) and from the leaves of pineapple fruit (*Ananas-comosus*) respectively. The inherent properties of coir and pineapple leaf fibers are mentioned in Table 1. Coconut coir fiber contains more amount of lignin as compared to other typically used natural fibers which result it has good resistance to salt water, microbial, and fungus attack. In comparison to coconut husk fiber (coir), the pineapple leaf fiber is a smooth, flexible, thin, and long fiber which contains less amount of lignin content. Pineapple natural fiber can be employed as an electrical, thermal, and acoustic insulator; for feedstock, and bio-fuel production [1]. According to the statistical database (2016) of "Food and Agriculture Organization," the total production of coconut and pineapple fruit in the whole world is 59 and 25.8 millions of tonnes respectively (Table 2) [2]. Therefore, the challenging job is to reduce their waste disposal in landfill areas and to utilize them in an efficient manner. From Table 1, we can conclude that the pineapple leaf fiber contains a large amount of cellulose (>65%) with a low spiral angle (14^0) as compared to the coir fiber, which results the inherent strength of pineapple leaf fiber is better than coir fiber. Whereas the impact energy absorbing power of

Accepted Manuscript

COIR is relatively higher than PALF. So to combine the desirable properties synergistically, incorporate both types of fiber in the same polymer matrix which results in a multifunctional composite material. Such types of multi fibers reinforced polymeric material are commonly called "Hybrid composites". Hybridization of two or more types of reinforcement in a single polymer matrix leads to the evolution in properties of single fiber reinforced composites. Hybrid composites possess high strength with improved fracture toughness, lightweight with reduced cost, and better creep & fatigue behavior. A. Shahzad confirmed the improvement in impact and fatigue strength of hemp fiber-polyester composite by hybridization with glass fiber [3]. M. Haq et al. reported that the novel multifunctional hybrid biocomposites, made by the incorporation of hemp fiber in a blend of UPE with EMS and nanoclay, have a good balance between stiffness and toughness which results these types of hybrid materials can be used in structural and transportation applications [4]. M.M. Davoodi et al. reported that the hybrid kenaf/glass epoxy composite shows comparable mechanical properties (tensile strength, flexural strength, Young's modulus, and flexural modulus) to the glass mat thermoplastic (GMT) which indicates its significant potential to select in an automotive structural application [5]. M.S. Sreekala et al. studied the effect of oil palm empty bunch fiber loading on physical and mechanical properties of hybrid OPEFB/Glass/Phenol formaldehyde composites and the results showed the improvement in impact strength with a considerably lower density of hybrid fiber reinforced composite [6]. R. Burgueno et al. reported that the biocomposite sheet made by the reinforcement of industrial hemp with chopped glass, woven jute, and unidirectional carbon fabrics in polyester resin shows excellent strength, stiffness, and dimensional stability which results in an economic and environment-friendly biocomposite material for structural application [7]. N. Ranganathan et al. investigated the influence of viscose fiber content on fracture toughness and fatigue life of Jute/Viscose reinforced polypropylene hybrid composites. The study shows that the fracture toughness (K_{IC}) and fatigue life were improved significantly with the addition of 10 wt.% viscose fibers [8]. M. Kaiser et al. observed that the hybridization of natural fiber-PLA composites with 3 wt.% nanoclay yields the improvement in impact strength by more than 50% [9]. V.S. Chevali et al. declared the diminution of tensile, flexural, and impact strength of the ABS polymer due to the hybridization by sunflower hull fiber and dried grain fibers [10]. S. Panthapulakkal et al. stated that the mechanical, thermal, and water sorption resistance properties of Hemp fiber/PP composites were improved with the incorporation of glass fiber [11]. A. Atiqah et al. have developed and characterized the kenaf/glass/UPE hybrid composites with varying glass fiber content and the results showed that the 15% composition

Accepted Manuscript

(15/15 v/v) of treated kenaf-glass hybrid fiber composite shows maximum values of tensile, flexural, and impact properties [12]. M. Jawaid et al. have studied the physical and chemical properties of OPEFB/Jute reinforced Epoxy hybrid composites and found the improvement in dimensional and chemical stability with less void content due to the increase of jute fiber volume fraction [13]. Sathishkumar et al. studied the mechanical properties of Snake grass/Coir and Snake grass/Banana fiber reinforced polyester based hybrid composites and found that the tensile strength and modulus of SG/B fiber hybrid composite are higher than SG and SG/C composites, while the flexural strength is maximum for SG/C hybrid composite [14]. P.N Khanam et al. investigated the effect of fiber lengths (1, 2, and 3 cm) and alkaline treatment on the mechanical properties of Coir/Silk fiber reinforced polyester based hybrid composites and found that the composites reinforced with treated fiber having length 2 cm have maximum tensile, flexural, and compressive strength [15]. R.K. Misra et al. studied the influence of coir fiber volume fraction on the mechanical properties of Jute-Coir reinforced epoxy based hybrid composites. They observed that the tensile and flexural strength increases with the increase of coir fiber content from 0% to 50% but beyond the addition of coir results the decrement in tensile and flexural properties [16]. S.K Saw et al. studied the water absorption, thickness swelling, and mechanical properties of Jute/Coir fiber hybrid composites and observed the improvement in dimensional stability and mechanical properties of pure coir reinforced epoxy novolac resin composite by the hybridization of coir fiber with jute fiber [17]. C.A Kakou et al. studied the effect of coir fiber loading on mechanical properties of Coir/oil palm fiber reinforced polyethylene hybrid composites and conclude that the tensile and flexural properties (strength and modulus) of the composite were increased with the increase of coir fiber content from 0% to 25% [18]. R. Siakeng et al. examined the physical properties of COIR/PALF reinforced hybrid PLA composites and the results showed the increase of density, water absorption and thickness swelling with the increase of coir fiber content [19]. S. Siddika et al. investigated the effect of fiber loading (5, 10, 15, and 20 wt.%) with fixed fiber ratio (Jute: Coir = 1:1) on mechanical properties of Jute-Coir fiber reinforced hybrid polypropylene composites. The results showed the decrease of tensile strength and the increase of tensile modulus, flexural strength, flexural modulus, impact strength, and hardness with the increase of fiber content from 5wt.% to 20 wt.% [20]. V.K Bhagat et al. studied the effect of coir fiber loading on physical properties of Coir/Glass reinforced epoxy based hybrid composites and observed that the density and hardness of composites decreases while tensile strength, tensile modulus, and flexural strength increases with the increase of coir fiber content from 5 wt.% to 10 wt.% [21]. O.O. Daramola et al. have studied the

Accepted Manuscript

mechanical properties and water absorption behavior of soil retted banana fiber reinforced polyester composites and found that the tensile properties (strength and modulus) and water sorption increases with increase in fiber content from 10 wt.% to 40 wt.% whereas the maximum flexural strength and modulus were obtained at 20 wt.% fiber content [22]. G. Das et al. investigated the mechanical properties of coir fiber reinforced composites hybridized with Al_2O_3 particulates. The results showed that the composites consist of 15 wt.% coir fiber content (12 mm length) and 10 wt.% Al_2O_3 particulates has maximum tensile, flexural, and impact strength as compared to others [23]. E. Osman et al. studied the water sorption behavior and its effect on the flexural properties of Kenaf/UPE composites and found that the percentage of water absorption increases with the increase of kenaf fiber content and temperature. Furthermore, the flexural properties (strength and modulus) of composites decrease significantly after the exposure to moisture [24]. K.G Girisha et al. stated that the percentage of moisture absorption increases with the increase of total fiber content in sisal/coir fiber reinforced epoxy hybrid composites [25]. IS Aji et al. reported that the composite reinforced with a fiber ratio of 1:1 (Kenaf and PALF) possess maximum tensile strength and modulus whereas K3P7 composite (Kenaf: PALF = 3:7) absorbs least amount of moisture as compared to other samples of different fiber fraction [26]. B. Bakri et al. reported that the composite contains 15:15 relative volume fraction of coir and agave fiber showed the higher tensile strength, tensile modulus, and flexural modulus than that of other relative fiber volume content (10:20 and 20:10) in an epoxy thermoset resin [27]. L. Uma Devi et al. stated that the water absorption of intimately mixed and GPG (Glass fiber as skin and PALF as a core) reinforced hybrid polyester composites increases with the increase of PALF content and testing temperature [28]. R. Yahaya et al. reported the decrement in density and dimensional stability of Kenaf-Kevlar reinforced epoxy composites with the increase of kenaf fiber content from 22 wt.% to 68%, keeping the fiber to matrix ratio 30:70 (v/v) [29]. T. Rihayat et al. reported that the tensile and flexural strength of pineapple leaf fiber and coir fiber composites increased due to the hybrid effect of bamboo, PALF, and COIR fiber [30].

From the above cited comprehensive works, it can be concluded that the hybridization of one type of fiber with another type (may be synthetic or natural fiber) results in the better physical, chemical, thermal, and mechanical properties compared to the single type of fiber reinforced composites. It can also be noticed that the work on development and characterization of hybrid composites made by the reinforcement of pineapple leaf fiber and coir fibers in an epoxy thermoset polymer have not been addressed so far. Therefore, in the

Accepted Manuscript

present communication, we have manufactured epoxy-based hybrid composites reinforced with short randomly oriented pineapple leaf fiber (PALF) and coconut husk fiber (COIR). Moreover, this work mainly deals with the effect of COIR fiber relative volume content on the mechanical properties and water absorption behavior of PALF/COIR reinforced hybrid epoxy composites. To accomplish this objective, the tensile, flexural, impact, and water absorption tests had been performed on fabricated biocomposite materials as per ASTM standard.

Table 1. Physical properties of pineapple leaf and coir fiber

Physical properties	Pineapple leaf fiber	Coir fiber
Density (g/cm^3)	0.98	1.2
Cellulose (%)	70-82	32-43
Hemicellulose (%)	18.8	0.15-0.25
Lignin (%)	5-12.7	40-45
Pectin (%)	1.1	3-4
Moisture content (%)	11.8	8
Microfibrillar angle (deg)	14	30-49
Diameter (μm)	20-80	100-460
Tensile strength (MPa)	413-1627	131-220
Young's modulus (GPa)	34.5-82.5	4-6
Elongation at break (%)	1.6	15-40

Table 2. Total annual production of pineapple leaf and coir fiber

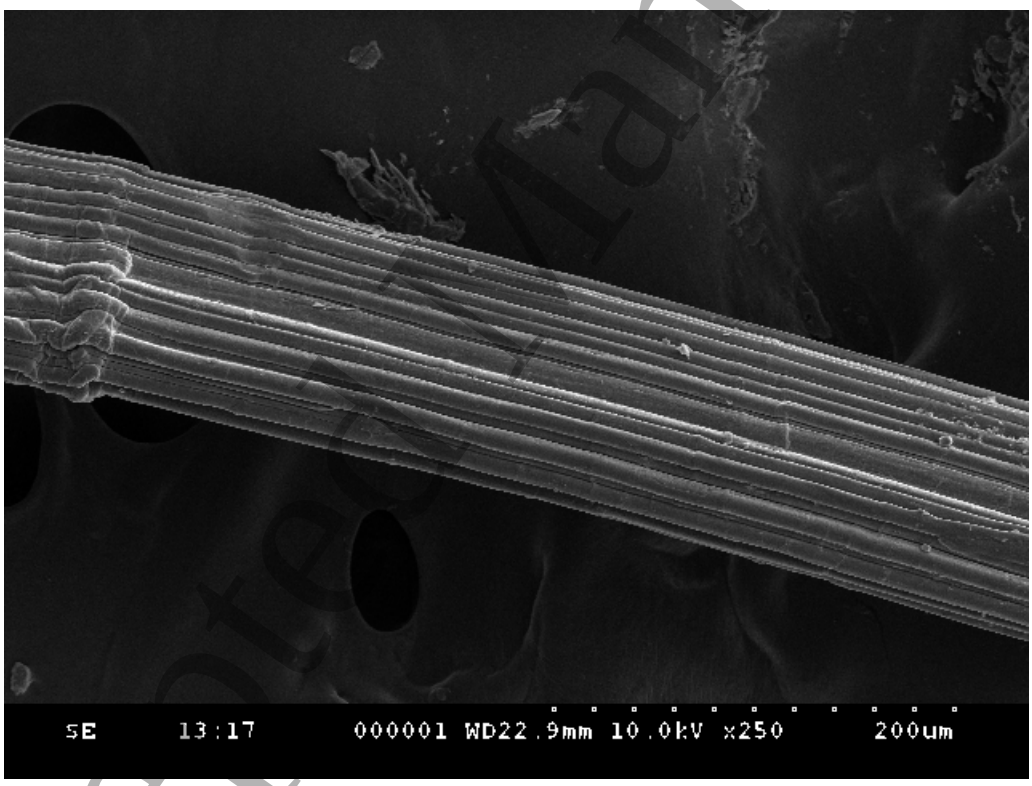
Pineapple Production-2016		Coconut Production-2016	
Country	Millions of tonnes	Country	Millions of tonnes
Costa Rica	2.9	Indonesia	17.7
Brazil	2.7	Philippines	13.8
Philippines	2.6	India	11.1
Thailand	1.9	Brazil	2.6
India	1.9	Sri Lanka	2.5
Indonesia	1.3		
World	25.8	World	59.0

Accepted Manuscript

2. Experimental

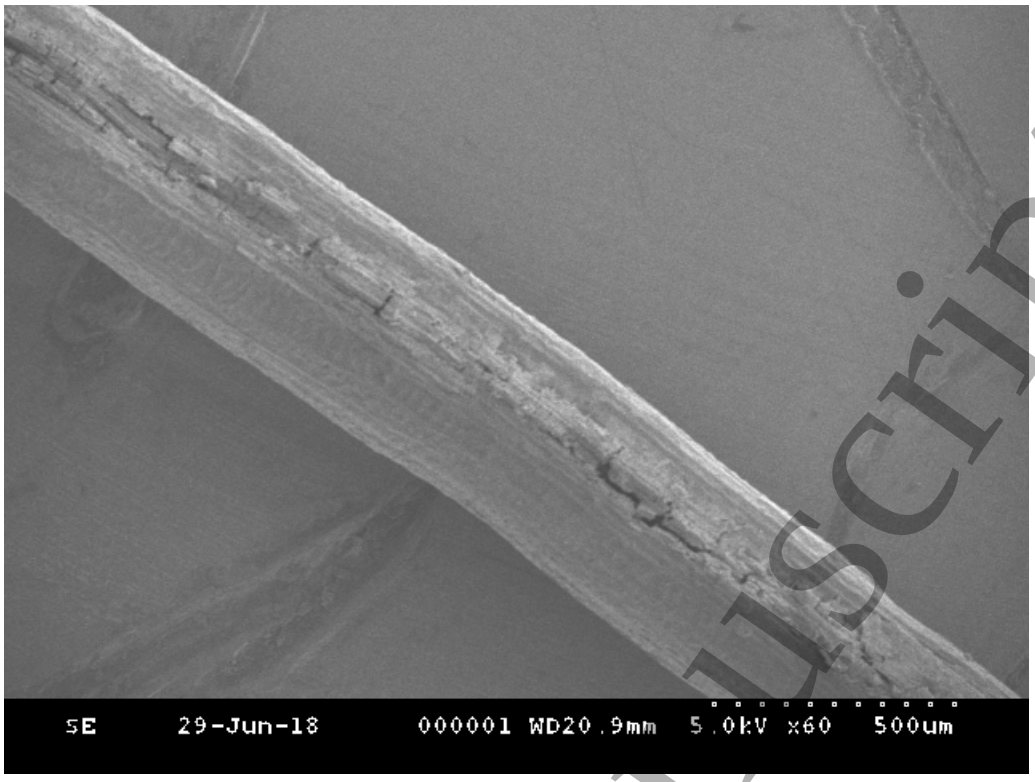
2.1 Materials

Pineapple leaf fiber (*Ananas-comosus*) and coconut husk fiber (*Cocos-nucifera*) were procured from M/s Go Green Products, Chennai, India. The fibers were washed with tap water to remove the dust and surface impurities, and dehydrated in sunlight for 3 days. The SEM micrographs of PALF and COIR are shown in Fig. 1. Table 1 shows the physical properties of both types of fibers. Epoxy adhesive (DGEBA, LY 556) with the cross-linking agent i.e. the hardener (TETA, HY 951) were obtained from M/s Sakshi Dies and Chemicals, Delhi, India. The epoxy thermoset was selected as a polymer matrix because of its low viscosity and shrinkage, excellent chemical, thermal, and moisture resistant properties, and ability to carry mechanical shocks. The curing reaction involved and the typical properties of the epoxy resin are depicted in Fig. 2 and Table 3 respectively.



(a)

Accepted Manuscript



(b)

Fig. 1 SEM micrograph: (a) pineapple leaf fiber (b) coconut husk fiber

Table 3. Properties of epoxy resin

The appearance of liquid resin	A clear pale yellow
Density (Kg/m^3)	1200-1250
Viscosity at 25^0C (cps)	550
Tensile strength (MPa)	11.14
Tensile modulus (GPa)	0.48
Flexural strength (MPa)	25.29
Flexural modulus (GPa)	1.53
Max. Elongation (%)	4.2-5.6
Impact strength (kJ/m^2)	1.2
Flashpoint	$> 200^0C$

Accepted Manuscript

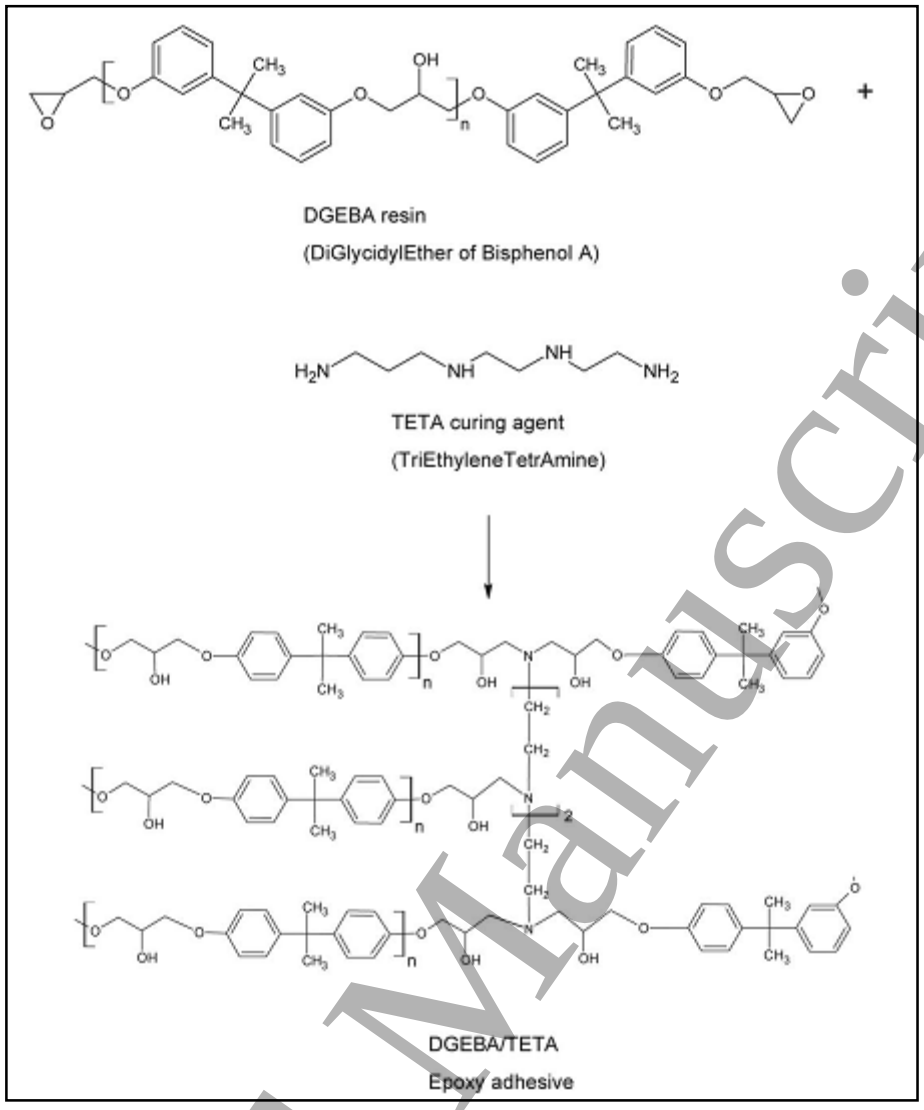


Fig.2 Curing reaction between epoxy resin (DGEBA) and hardener (TETA)

2.2 Fabrication of composites

The hybrid polymer composites were made by reinforcing the epoxy thermoset resin with pineapple leaf fiber (PALF) and coconut husk fiber (COIR) in an open mold cavity of dimensions 200 mm X 100 mm X 4 mm. Initially, cleaned the fibers with distilled water, dried in sunlight for three days, and chopped the fibers to a length of 20 mm by using scissors. The medium viscosity epoxy resin was used to impregnate the dried fibers of PALF and COIR. The epoxy resin (DGEBA) was thoroughly mixed with a curing agent (TETA) in a necked round glass flask through mechanical stirrer at a room temperature (approx. 30 ^{0}C) for 20 minutes. Before using the thermoset matrix for molding, ensure that the resin was mixed completely with hardener and no air bubbles were present. The volume ratio of resin and hardener was kept at 2:1 (v/v). In all composite samples, the volume ratio between fiber and matrix was taken as 40:60. The total of 11 specimens was fabricated by varying the

Accepted Manuscript

relative fiber volume percentage (shown in Table 4). The hand lay-up technique was employed to fabricate the biocomposite laminate sheet because of its huge number of advantageous features such as low tooling cost, easy and simple to use, least expensive and can be smoothly adopted by small-scale industries. The mold was covered with a Teflon sheet and the surface of the mold was sprayed with a releasing agent to prevent the sticking of adhesive polymer resin. The intimately mixed short and randomly oriented fibers are placed in a glass mold and impregnated with the premixed epoxy matrix. Ensure that the distribution of fibers and polymeric resin are uniforms throughout the mold space. During the fabrication of a composite sheet, the air bubbles are abstracted cautiously by using a paint steel roller. In the end, closed the mold and then pressed with 35 kg dead weight for 24 hrs curing at room temperature. The tensile, flexural, and impact test specimens were cut from the molded sheet as per ASTM standard.

Table 4. Designation and composition of PALF/COIR reinforced hybrid epoxy composites

Designation	Composition
EP	Epoxy matrix (100 vol.%)
P100	Epoxy (60 vol.%) + Pineapple leaf fiber (40 vol.%) + Coir fiber (0 vol.%)
P90-C10	Epoxy (60 vol.%) + Pineapple leaf fiber (36 vol.%) + Coir fiber (4 vol.%)
P80-C20	Epoxy (60 vol.%) + Pineapple leaf fiber (32 vol.%) + Coir fiber (8 vol.%)
P70-C30	Epoxy (60 vol.%) + Pineapple leaf fiber (28 vol.%) + Coir fiber (12 vol.%)
P60-C40	Epoxy (60 vol.%) + Pineapple leaf fiber (24 vol.%) + Coir fiber (16 vol.%)
P50-C50	Epoxy (60 vol.%) + Pineapple leaf fiber (20 vol.%) + Coir fiber (20 vol.%)
P40-C60	Epoxy (60 vol.%) + Pineapple leaf fiber (16 vol.%) + Coir fiber (24 vol.%)
P30-C70	Epoxy (60 vol.%) + Pineapple leaf fiber (12 vol.%) + Coir fiber (28 vol.%)
P20-C80	Epoxy (60 vol.%) + Pineapple leaf fiber (8 vol.%) + Coir fiber (32 vol.%)
P10-C90	Epoxy (60 vol.%) + Pineapple leaf fiber (4 vol.%) + Coir fiber (36 vol.%)
C100	Epoxy (60 vol.%) + Pineapple leaf fiber (0 vol.%) + Coir fiber (40 vol.%)

2.3 Tensile Testing

Tensile tests were performed on fabricated PALF/COIR-Epoxy composite specimens by using a Table Top Tinius Olsen Horizon H50KS, Universal Testing Machine in accordance with ASTM D 638 standard. The tensile properties (strength, modulus, breaking load, and elongation at break) of composite material were measured at room temperature with a

Accepted Manuscript

crosshead speed of 1 mm/min by using 50 N load cell. The standardized samples for tensile tests were in dumbbell shape of 50 mm gauge length, 13 mm width, and 4 mm thickness. The results are reported as an average of four identical test specimens. The tensile properties of reinforced fiber (PALF and COIR) and cured epoxy thermoset resin are reported in Table 1 and 3 respectively.

2.4 Flexural Testing

The composite samples of dimensions 80 mm X 15 mm X 4 mm with the span length to depth ratio 20:1 were subjected to a 3-point flexural bending test as per ASTM D 790 standard, to determine the flexural strength and modulus. The flexural test was executed on the above-mentioned UTM with the crosshead speed of 3 mm/min. Four similar sized test specimens were tested in each case and the mean value was reported.

2.5 Impact Testing

The impact fracture strength of unhybridized and hybridized notched composite specimens was determined by using an Izod impact testing machine (installed in the Strength of Material testing laboratory, DTU (Delhi) India) as per ASTM D 256 standard. The dimensions of impact test specimens are 65 mm X 13 mm X 4 mm. The total of four samples for each specimen was tested and the average values were reported.

2.6 Water Absorption Test

The water absorption test on the cured epoxy matrix and the developed PALF/COIR reinforced epoxy based hybrid composites was performed according to ASTM D 570 standard. Four samples of each composite specimen were immersed in distilled water at room temperature. Initially, dried all the samples in a vacuum oven at 60^0C for 24 hrs, and then cooled to room temperature. Repeat the drying process until the constant weight was obtained. The weight of dried and wet samples was measured by using a digital highly calibrated weighing machine (0.001 g accuracy, DTU (Delhi), India). After every 12 hrs of immersion, the samples were taken out from the water bath and wipe dry to remove redundant water molecules on the surfaces. The weight gain (%) was calculated by using the formula

$$Weight\ gain\ (\%) = \frac{(W_t - W_O)}{W_O} \times 100 \quad (1)$$

Accepted Manuscript

where W_t and W_O are the weight of the specimen after immersion time "t" and the oven dried weight respectively.

3. Results and Discussion

3.1 Mechanical Properties

In order to quantify the reinforcing ability of PALF, COIR, and PALF-COIR hybrid fiber in an epoxy thermoset polymer resin, the mechanical tests (tensile test, Flexural (3 point bending) test, and impact test) were performed on pure pineapple-epoxy, coir-epoxy, and pineapple/coir fiber hybrid epoxy composites. The mechanical properties of fiber reinforced polymer composites are dependent on the inherent properties of fiber and matrix, internal state and structure of composite, fiber-matrix interfacial bonding, and environmental conditions. Determination of mechanical properties can also provide indirect information about the interfacial interaction between fiber and polymer matrix.

3.1.1 Tensile Properties

The tensile properties (tensile strength and modulus) of unhybridized and hybridized pineapple leaf fiber and coir fiber reinforced epoxy based composites are shown in Fig. 3. It can be easily reported that the tensile strength and modulus of cured epoxy thermoset resin (shown in Table 3) was increased with the incorporation of pineapple leaf fiber and coir fiber. The positive reinforcing effect of PALF and COIR fibers on tensile properties of epoxy matrix can be observed in Table 5. The mechanical properties of PALF and COIR fiber are reported in Table 1 and revealed the higher strength and stiffness of pineapple leaf fiber than that of coconut husk fiber. Among the single fiber reinforced composites, PALF/Epoxy composite possess 61.2% and 49.5% higher tensile strength and tensile modulus as compared to the COIR/Epoxy composite respectively. This was because of the better inherent tensile properties of PALF than COIR fiber. Moreover, the diameter and microfibrillar angle of PALF is less than that of COIR fiber that leads to the higher surface area and reinforcement effectiveness of PALF than COIR. In the case of PALF/COIR hybrid composites of varying coir fiber volume content, it was observed that the tensile strength and modulus of COIR/Epoxy composite was elevated by the hybridization with pineapple leaf fiber. Fig. 3 makes clear that even the addition of 0.2 V_f (20% of the total fiber volume) pineapple leaf fiber in coir-epoxy composite, the tensile strength was increased from 12.34 MPa to 22.97 MPa. From the tensile test, we can conclude that the hybrid composite (P50-C50) containing relative volume fraction of two fibers PALF: COIR = 1:1 has maximum tensile strength and

Accepted Manuscript

modulus of elasticity. The tensile strength of pure PALF/Epoxy composite was increased by 16.74% with the incorporation of COIR fiber in equal volume content as PALF. This is due to the synergistic compatibility between pineapple leaf fiber and coir fiber and it is evident that the poor elongation at break (%) of PALF fiber can be overcome by the positive hybridization with high strain to failure COIR fiber. Moreover, the hybridization of a polymer matrix with optimum relative fiber volume improves the adhesion characteristics at the fiber surface. Previous research works [Sathishkumar et al. [14], C.A Kakou et al. [18], V.K Bhagat et al. [21], Siddika et al. [20]] were also reported the similar result. Fig. 3 clearly depicts the diminution in tensile strength with the further increase of PALF volume content beyond the 50% of total fiber volume ($V_f = 0.4V_c$) in the hybrid composite. This was attributed to the poor dispersion of fiber, an increase of void spaces, and poor stress transfer from the matrix to fiber. It is also found that at a very high concentration of PALF fiber in the hybrid composites, the processing is poor and fibers get agglomerated with an improper orientation which results in packing defects and rapid crack propagation.

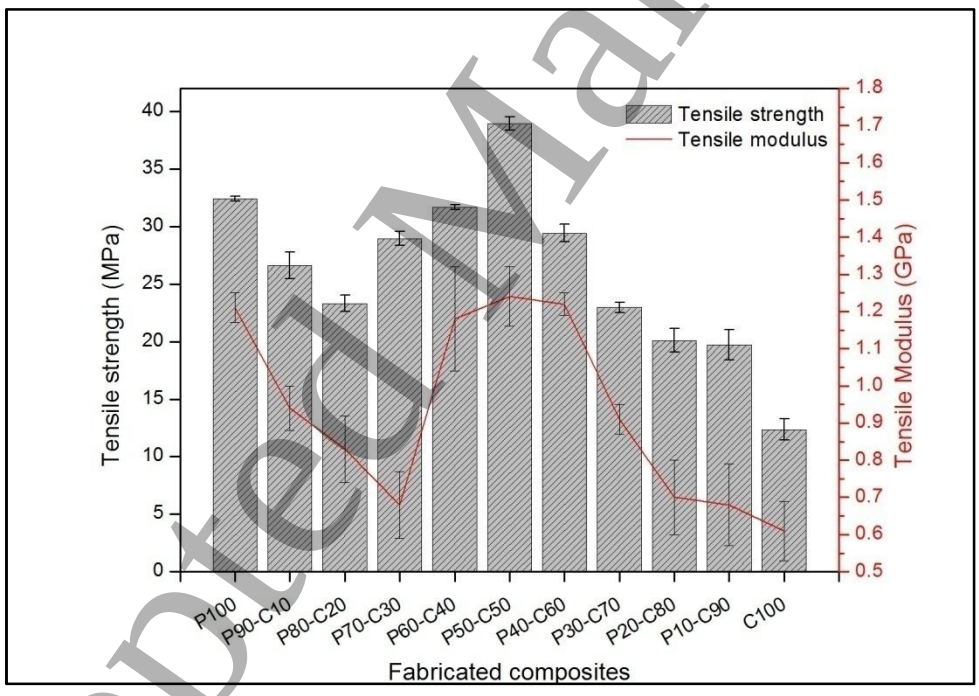


Fig.3 Tensile properties of fabricated composites as a function of coir fiber volume content

3.1.2 Flexural Properties

The flexural properties (flexural strength and modulus) of the formulated fiber reinforced composites were determined as per ASTM D 790 standard by using a Tinius Olsen H50KS, universal testing machine. Similar to the tensile properties, the improvement in flexural strength of epoxy thermoset matrix due to the addition of PALF and COIR fibers is shown in

Accepted Manuscript

Table 5. Fig. 4 shows the variation in flexural properties of PALF/COIR fiber reinforced hybrid composites with the increase of coir fiber content. It was observed that the flexural strength of pure COIR/Epoxy composite was significantly improved with the incorporation of pineapple leaf fiber, up to 50% of total fiber volume. The maximum flexural strength of 102.81 MPa is obtained for P50-C50 hybrid composite. The step-up of flexural strength by 23.44% and 77.03% was observed between the P100 to P50-C50 and C100 to P50-C50 composites respectively. This was because of the increased shearing resistance of fibers and effective stress transfer from the matrix to reinforcing fibers at this particular relative fiber volume content. Flexural strength is a combined effect of tensile, shear, and compressive strength. During the three-point bending test, the lower end of composite material experience tensile stress, upper end faced compressive stress, and the axisymmetric plane is subjected to a shear load. The flexural strength of a composite material almost follows the similar pattern as tensile strength. Similar to tensile strength, the flexural strength is also decreased with the further loading of pineapple leaf fiber from 0.5 V_f to 0.9 V_f. The flexural strength value was rapidly increased with the increase of pineapple leaf fiber volume content from 0.3 V_f to 0.5 V_f. This behavior is due to the synergistic compatibility between PALF and COIR in P40-C60 and P50-C50 composites. Moreover, an optimum balance between strength, stiffness, and ductility was obtained as a result of balanced combination of PALF and COIR fiber in a viscoelastic epoxy thermoset medium. However, the hybrid composite P60-C40 possess the higher value of flexural modulus than the P50-C50 composites. It was due to the little more accumulation of stiffer and stronger pineapple leaf fiber which results to restrict the plastic deformation.

Table 5. Percentage change in mechanical properties of epoxy matrix after the incorporation of pineapple leaf and coir fiber

Composites	Tensile strength (MPa)^{IM}	Tensile modulus (GPa)^{IM}	Flexural strength (MPa)	Flexural modulus (GPa)	Impact strength (kJ/m^2)^{IM}
P100	65.63	60.33	67.86	23.11	57.29
P90-C10	58.15	48.93	58.7	14.17^{DC}	66.19
P80-C20	52.2	42.16	69.48	3.77	74.63
P70-C30	61.51	29.41	67.31	28.5	77.01
P60-C40	64.83	59.32	73.79	62.68	84.59
P50-C50	71.39	61.29	75.4	58.87	86.63
P40-C60	62.12	60.65	71.62	50.32	86.28
P30-C70	51.5	47.25	46.63	12.57	83.02
P20-C80	44.6	31.42	29.57	16.79^{DC}	84.96
P10-C90	43.53	29.41	16.09	28.57^{DC}	85.54
C100	9.72	21.31	7.11^{DC}	47.11^{DC}	86.11

Accepted Manuscript

[IM] Improvement (%), [DC] Reduction (%)

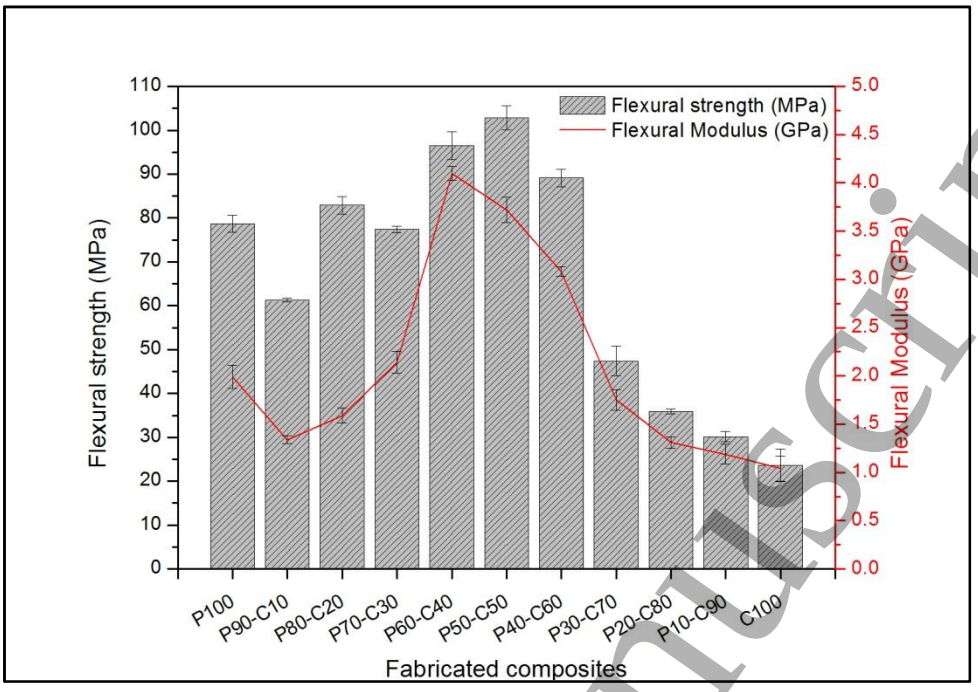


Fig.4 Flexural properties of fabricated composites as a function of coir fiber volume content

3.1.3 Impact Strength

The overall impact toughness of a fiber reinforced composite material is dependent on the nature of the constituent elements; internal structure and geometry of the composite; interfacial interaction between the reinforcing elements and the matrix; and experimental conditions. In order to study the effect of pineapple leaf fiber, coir fiber, and hybrid pineapple-coir fiber reinforcement on the impact performance of epoxy thermoset polymer, the fabricated composites were subjected to Izod Impact test as per ASTM D 256 standard. The impact strength of a virgin epoxy matrix (1.2 kJ/m^2) was significantly improved with the incorporation of PALF and COIR fiber (shown in Table 5). From Fig. 5, it can be reported that the pure COIR/Epoxy composite has 67.47% higher impact strength than that of pure PALF/Epoxy composite. This might be due to the inherent property of coir fiber (large value of spiral angle) and the dominating phenomena of "Fiber Pull-Out" than the "Fiber Fracture" in COIR/Epoxy composite. Moreover, the COIR fiber contains less cellulosic content with porous surface morphology (shown in Table 1 and Fig.1) which leads to a good interfacial adhesion and wetting of COIR fiber with thermoset epoxy matrix by capillary action. As observed in Fig. 5, the impact strength of PALF/Epoxy composites increases by the hybridization effect of Coir and PALF fiber in a polymer matrix. The hybrid composite (P50-C50) shows the maximum value of impact toughness (8.98 kJ/m^2) and it was due to the

Accepted Manuscript

synergistic combination of properties. The low elongation to failure property of PALF/Epoxy composite can be overcome with the incorporation of high strain to failure-COIR fiber. The strength of a material under impact loading is a function of stiffness and ductility. In the case of PALF/COIR hybrid composite, the inherent strength of pineapple leaf fiber makes a good balance with ductile nature of coir fiber which results the improvement in toughness of developed composites. Amongst all samples, the composite reinforced with an equal volume fraction of coir and pineapple leaf fiber (P50-C50) possess better mechanical properties than others. This was because of the good interfacial adhesion between fiber and matrix that leads hamper the crack propagation along the interface. It is worth to note that the impact strength of a material indicates the amount of energy dissipated in the form of heat during the process of crack initiation, growth, and propagation. In relative to pure COIR/Epoxy composite, the hybrid composite (P40-C60) shows better impact strength. It was attributed to the positive hybridization effect with intercede result (C100<P40-C60>P50-C50). As expected, the hybrid composites reinforced with optimum relative fiber volume content presents the better impact toughness than that of single fiber reinforced polymer composites.

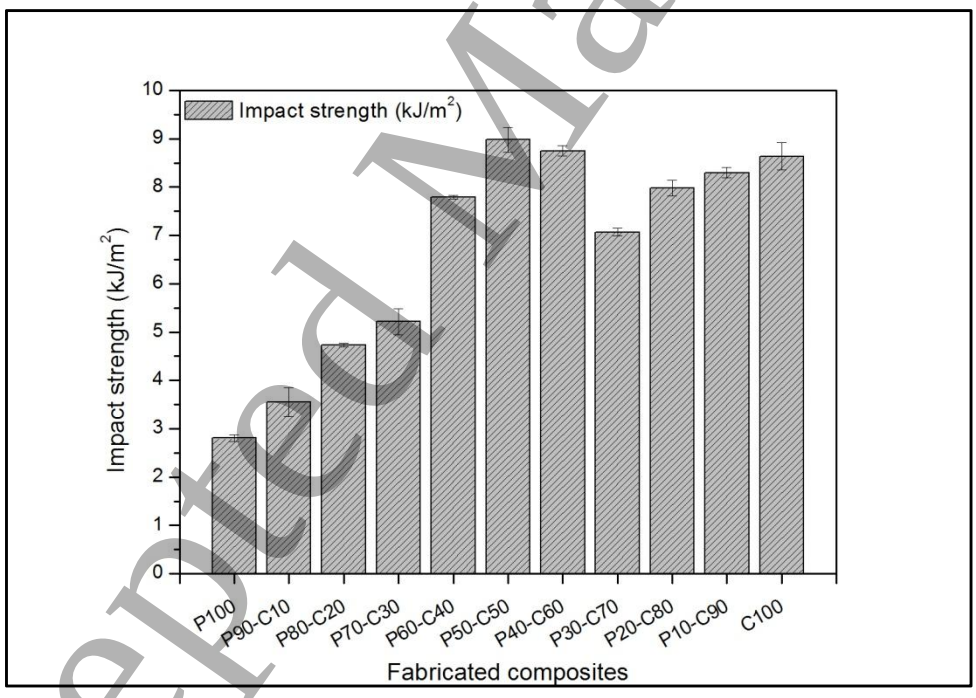


Fig.5 Impact strength of fabricated composites as a function of coir fiber volume content

3.2 Water Absorption

Water sorption characteristics of virgin epoxy, pineapple-epoxy, coir-epoxy, and pineapple-coir fiber hybrid epoxy composites were studied with the special reference of immersion time (in hrs) and the relative volume fraction of both types of cellulosic fiber. The water

Accepted Manuscript

absorption curves for all formulated composites as a function of immersion time and the coir fiber content are shown in Figs. 6 and 7 respectively. It was observed that the virgin epoxy resin absorbs the least amount of water among all the composite specimens. The moisture uptake ability and dimensional instability of thermoset polymer was increased with the incorporation of cellulosic fibers. The diffusion of water molecules in polymers is due to the presence of micro-gaps and cavities between polymeric chains. According to the Flory's proposed theory, the water diffusion in polymer and polymer-based composites are two step processes. In 1st step, the penetrant diffuses rapidly to induce elasticity in polymer chains and in 2nd step, the saturation level (equilibrium absorption) was achieved that leads to the polymer chain relaxation. Fig. 6 clearly shows that all the composites absorb water sharply up to 50 hr and reached to the saturation point at 120 hr. The maximum water absorption values (in %) of all the fabricated composites are reported in Table 6. It was observed that the pure pineapple-epoxy composite absorbs more water as compared to the coir fiber-epoxy composite. This was attributed to the presence of a large number of hydroxyl (-OH) and other polar functional groups on the surface of pineapple leaf fiber. Water uptake in NFRC's depends on the inherent sorption power of individual fibers and matrix, fiber length and content, fiber orientation, matrix cracking, and fiber-matrix interfacial adhesion. Our previous work already reported that the PALF absorb more water than the COIR fiber [31]. From Fig. 7, it can be concluded that the hybridization of a PALF-Epoxy composite with coir fiber leads to the improvement in resistance to water diffusion. The hybrid composite (P50-C50) has the least affinity to absorb water because of excellent structural integrity, good fiber-matrix interfacial adhesion, and less number of voids, cavities, and micro-pores. In comparison to the pure pineapple leaf fiber-epoxy and coir fiber-epoxy composite, the hybrid composite (P50-C50) absorbs 62% and 32% less amount of water respectively. It indicates that the hybrid fiber reinforcement in polymer matrix yields more packed structure which leads to the dimensionally stable high strength composite. Penetration of water molecules through the composite materials is mainly occurring by the capillary drive into the cracks, gaps, voids, and flaws at the interfacial plane between fibers and polymer matrix. The percent water absorption of single coir fiber reinforced epoxy composite is significantly reduced by hybridization with a $0.2V_f$ content of pineapple leaf fiber. This may be due to the improved fiber-matrix interaction and better-ordered dispersion of fiber in the matrix resin. In addition to the polymer and fiber swelling, the delamination of the interlayer between the fibers can also be the cause of free voids that leads to an important mechanism for water absorption in

Accepted Manuscript

the hybrid composites. To understand the kinetics and mechanism of water absorption, the experimental data are fitted into the following generalized equation

$$Log\frac{M_t}{M_m} = Log(k) + nLog(T) \quad (2)$$

where M_t and M_m are the water absorption at time t and at saturation level respectively, k is constant parameter and n is the diffusion coefficient. The water sorption constants n and k define the mechanism of water diffusion and the kind of interaction between polymer and water respectively. According to the value of n (n<0.5, n=0.5, 0.5<n<1, n>1), the water diffusion mechanism can be classified into less Fickian, Fickian, Anomalous, and Non-Fickian mechanism respectively. From Table 6, it can report that all the composites follow less Fickian water diffusion mechanism.

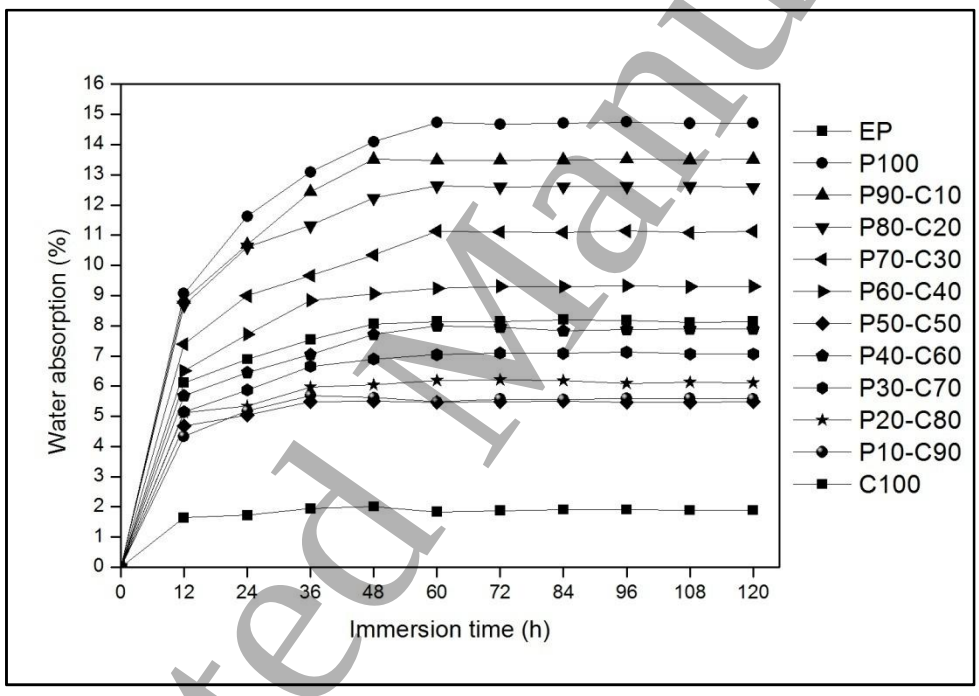


Fig.6 Water absorption behavior of fabricated composites as a function of immersion time

Accepted Manuscript

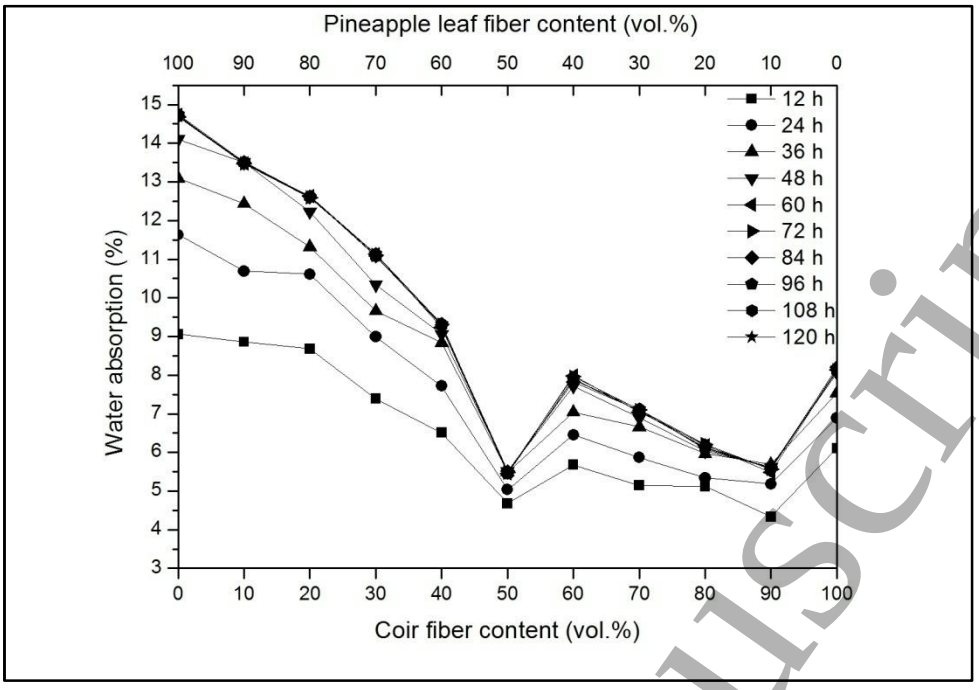


Fig.7 Water absorption behavior of fabricated composites as a function of coir fiber volume content

Table 6. Maximum water absorption, water sorption constants, and diffusion mechanism of the fabricated composites

Composites	Max. WA (%)	Diffusion exponent (n)	Intercept (k)	R^2	Type of diffusion mechanism
EP	2.01	0.14712	-0.65441	0.85839	Less Fickian diffusion
P100	14.75	0.1374	-0.61104	0.84521	
P90-C10	13.52	0.083	-0.37181	0.77271	
P80-C20	12.64	0.088	-0.38882	0.59984	
P70-C30	11.14	0.13	-0.55484	0.84624	
P60-C40	9.33	0.057	-0.25739	0.43924	
P50-C50	5.51	0.20	-0.8951	0.84861	
P40-C60	7.99	0.18	-0.78488	0.8108	
P30-C70	7.12	0.16	-0.68983	0.85101	
P20-C80	6.21	0.176	-0.78915	0.89529	
P10-C90	5.69	0.148	-0.65428	0.81936	
C100	8.2	0.062	-0.27329	0.69185	

Figs. 8 to 10 revealed the comparison of mechanical properties between the developed hybrid composite (P50-C50) and commonly used car bumper beam material, GMT. It was observed that the tensile, flexural, and impact strength of untreated PALF/COIR hybrid material is significantly less as compared to the pure glass fiber reinforced polypropylene composite. This was because of the improper interfacial adhesion between natural fibers and thermoset

Accepted Manuscript

matrix as well as the higher inherent properties of glass fiber compared to the PALF and COIR. In order to elevate the suitability of developed biocomposite material for structural applications, we need to form a strong covalent bond between reinforcing fibers and matrix. This can be achieved by modifying the surface morphology of pineapple leaf and coir fiber through chemical treatment process (alkaline, silane, acetylation, benzoylation, acrylonitrile grafting, peroxide, and permanganate treatments etc.). Furthermore, the compression at high pressure can also improve the adhesion between fibers and polymer matrix.

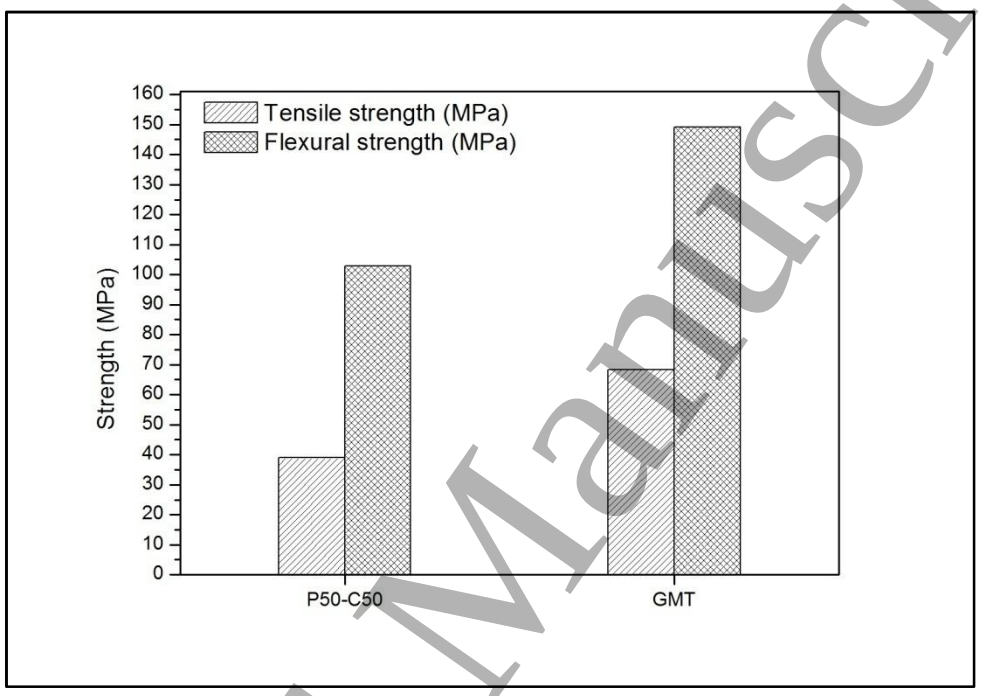


Fig.8 Comparison graph of tensile and flexural strength between P50-C50 hybrid composite and common car bumper beam material, GMT

Accepted Manuscript

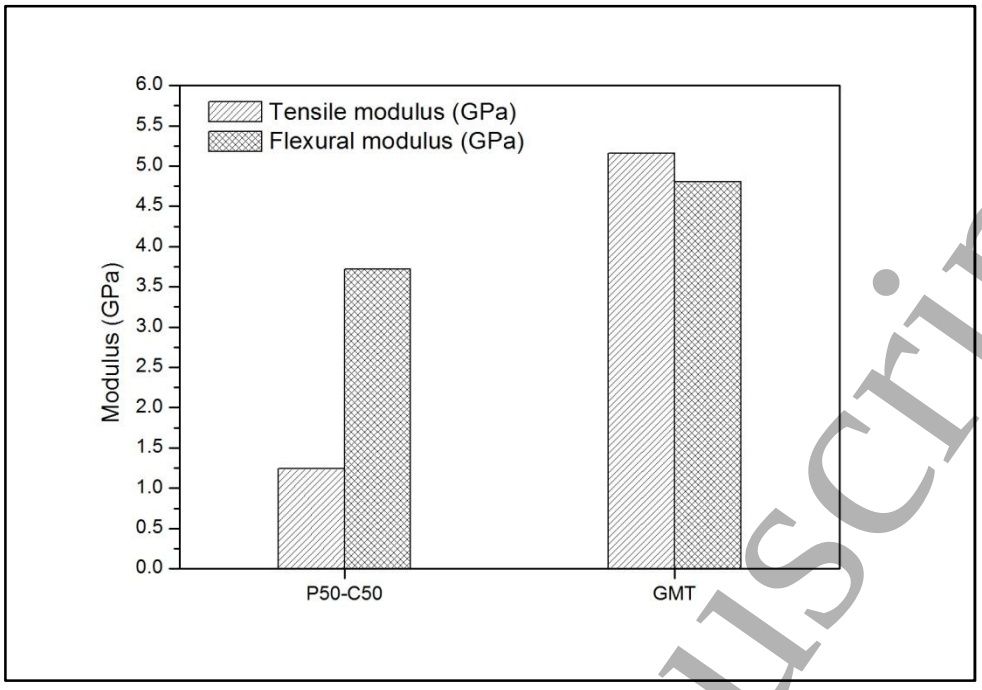


Fig.9 Comparison graph of tensile and flexural modulus between P50-C50 hybrid composite and common car bumper beam material, GMT

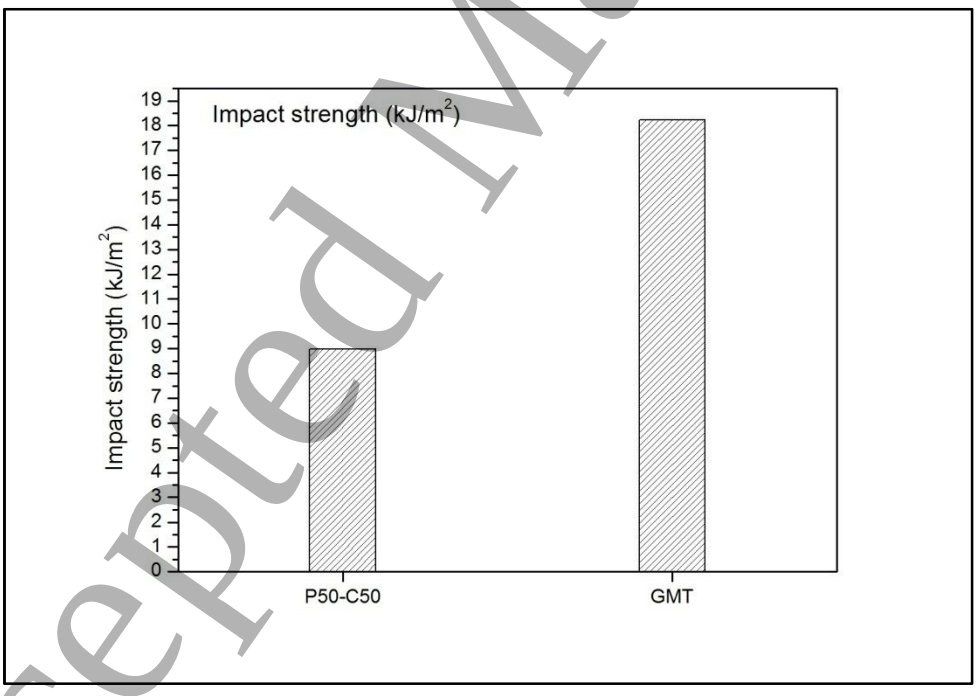


Fig.10 Comparison graph of impact strength between P50-C50 hybrid composite and common car bumper beam material, GMT

Accepted Manuscript

4. Conclusions

The current research work on the experimental investigation of the mechanical and water absorption behavior of PALF/COIR fiber reinforced hybrid epoxy composites leads to the following conclusions:

- The tensile and flexural properties (strength and modulus) of PALF/Epoxy composite are higher than that of COIR/Epoxy composite. However, the composite reinforced with coir fiber shows a greater value of impact strength and dimensional stability as compared to the pineapple leaf fiber reinforced composite.
- The mechanical properties and moisture sorption impedance of pineapple leaf fiber and coir fiber reinforced epoxy composites have been improved with the hybridization of PALF and COIR fiber in an optimum relative volume content. The overall results depict that the hybrid composites (P40-C60, P50-C50, and P60-C40) have better performance as compared to the other composite specimens.
- The maximum tensile strength, tensile modulus, flexural strength, and impact strength was observed for the composite reinforced with an equal volume fraction of pineapple leaf fiber and coconut husk fiber. Moreover, the hybrid composite (P50-C50) absorbs least amount of water than the other composite samples.
- The rate and maximum values of water absorption decrease with the increase of coir fiber content. According to the water diffusion exponent (n), it was concluded that a large number of composites exhibit Fickian diffusion mechanism.

On the basis of this experimental work, it can be concluded that the epoxy-based hybrid composite consists of equal volume content of pineapple leaf fiber and coir fiber demonstrate better mechanical properties as compared to the other hybrid formulations. The developed hybrid material can be used to fabricate helmets; roofing; floor furnishing materials; different parts of automobile such as door panels, head liner panel, package and floor trays, noise insulation panels; low cost cellular beam; raw material for particle board; printed circuit boards etc. However, the hybrid composite material P50-C50 has lower mechanical strength than that of common car bumper beam material, GMT. In order to increase the probability of PALF/COIR hybrid biocomposite for structural applications like automotive, aerospace, and constructions, we must require to improve the interfacial adhesion between reinforcing fibers and matrix by using appropriate chemical treatment process.

Accepted Manuscript

References

1. Asim, M., Abdan, K., Jawaid, M., Nasir, M., Dashtizadeh, Z., Ishak, M.R., and Hoque, M. (2015). A review on pineapple leaves fibre and its composites, *Int. J. Polym. Sci.* http://dx.doi.org/10.1155/2015/950567.

2. Statistical databases of "*Food and Agriculture Organization of the United Nations*", www.fao.org/statistics/databases/en/

3. Shahzad A (2011) Impact and fatigue properties of hemp-glass fiber hybrid biocomposites *J. Reinf. Plast. Compo.* **30**(16): 1389-1398.

4. Haq M, Burgueno R, Mohanty A K, and Misra M (2008) Hybrid bio-based composites from blends of unsaturated polyester and soybean oil reinforced with nanoclay and natural fibers *Compos. Sci. Technol.* **68**: 3344-3351.

5. Davoodi M M, Sapuan S M, Ahmad D, Ali A, Khalina A, and Jonoobi M (2010) Mechanical properties of hybrid kenaf/glass reinforced epoxy composite for passenger car bumper beam *Mater. Des.* **31**: 4927-4932.

6. Sreekala M S, George J, Kumaran M G, and Thomas S (2002) The mechanical performance of hybrid phenol-formaldehyde based composites reinforced with glass and oil palm fibers *Compos. Sci. Technol.* **62**: 339-353.

7. Burgueno R, Quagliata M J, Mohanty A K, Mehta G, Drzal L T, and Misra M (2005) Hybrid biofiber-based composites for structural cellular plates *Compos. Part A* **36**: 581-593.

8. Ranganathan N, Oksman K, Nayak S K, and Sain M (2015) Structure-property relation of hybrid biocomposites based on jute, viscose, and polypropylene: The effect of the fiber content and the length on the fracture toughness and the fatigue properties *Compos. Part A* http://dx.doi.org/10.1016/j.compositesa.2015.10.037.

9. Kaiser M R, Anuar H, and Abdul Razak S B (2014) Improvement in thermomechanical properties of injection molded nano-modified hybrid biocomposite *J. Thermoplast. Compos. Mater.* **27**(7): 992-1009.

Accepted Manuscript

10. Chevali V S, Nerenz B A, Ulven C A, and Kandare E (2015) Mechanical properties of hybrid lignocellulosic fiber-filled acrylonitrile butadiene styrene (ABS) biocomposites *Polym. Plast. Technol. Eng.* **54**: 375-382.

11. Panthapulakkal S and Sain M (2006) Injection-molded short hemp fiber/glass fiber-reinforced polypropylene hybrid composites- Mechanical, water absorption, and thermal properties *J. Appl. Polym. Sci.* **103**, 2432-2441.

12. Atiqah A, Maleque M A, Jawaid M, and Iqbal M (2014) Development of kenaf-glass reinforced unsaturated polyester hybrid composite for structural applications *Compos. Part B* **56**: 68-73.

13. Jawaid M, Abdul Khalil H P S, Hassan A, and Abdallah E (2012) Bi-layer hybrid biocomposites: Chemical resistant and physical properties *Bioresources* **7**(2): 2344-2355.

14. Sathishkumar TP, Navaneethakrishnan P, Shankar S, and Kumar J (2012) Mechanical properties of randomly oriented snake grass fiber with banana and coir fiber reinforced hybrid composites *J. Compos. Mater.* **0**(0): 1-11.

15. Khanam P N, Reddy G R, Raghu K, and Naidu S V (2009) Tensile, flexural, and compressive properties of Coir/Silk fiber reinforced hybrid composites *J. Reinf. Plast. Compo.* doi: 10.1177/0731684409345413.

16. Misra R K, Saw S K, and Datta C (2011) The influence of fiber treatment on the mechanical behavior of Jute-Coir reinforced epoxy resin hybrid composite plate *Mech. Adv. Mat. Struct.* **18**: 431-445.

17. Saw S K, Akhtar K, Yadav N, and Singh A K (2014) Hybrid composites made from Jute/Coir fibers: Water absorption, Thickness swelling, Density, Morphology, and Mechanical Properties *J. Nat. Fibers* **11**: 39-53.

18. Kakou C K, Essabir H, Bensalah M, Bouhfid R, Rodrigue D, and Qaiss A (2015) Hybrid composites based on polyethylene and coir/oil palm fibers *J. Reinf. Plast. Compo.* **0**(0): 1-14.

Accepted Manuscript

19. Siakeng R, Jawaid M, Ariffin H, and Sapuan S M (2018) Physical properties of coir and pineapple leaf fiber-reinforced polylactic acid hybrid composites *IOP Conf. Ser. Mater. Sci. Eng.* **290**, doi: 10.1088/1757-899X/290/1/012031.

20. Siddika S, Mansura F, Hasan M, and Hassan A (2014) Effect of reinforcement and chemical treatment of fiber on the properties of jute-coir fiber reinforced hybrid polypropylene composites *Fiber Polym.* **15**(5): 1023-1028.

21. Bhagat V K, Biswas S, and Dehury J (2013) Physical, Mechanical, and Water absorption behavior of Coir/Glass fiber reinforced epoxy based hybrid composites *Polym. Compo.* doi: 10.1002/pc.22736.

22. Daramola O O, Akintayo O S, Adewole T A, and Talabi H K (2017) Mechanical properties and water absorption behavior of polyester/soil retted banana fiber composites *ANNALS of Faculty Engineering Hunedoara Int. J. Eng.* 183-190.

23. Das G and Biswas S (2016) Physical, mechanical, and water absorption behavior of coir fiber reinforced epoxy composites filled with Al_2O_3 particulates *IOP Conf. Ser. Mater. Sci. Eng.* **115**, doi: 10.1088/1757-899X/115/1/012012.

24. Osman E, Vakhguelt A, Sbarski I, and Mutasher S Water absorption behavior and its effect on the mechanical properties of kenaf natural fiber unsaturated polyester composites, *18th International Conference on Composite Materials.*

25. Girisha K G, Akash Gupta N S V, and Rao K V (2017) A study on flammability and water absorption behavior of sisal/coir fiber reinforced hybrid composites *IOP Conf. Ser. Mater. Sci. Eng.* **191**, doi: 10.1088/1757-899X/191/1/012003.

26. Aji IS, Zainudin ES, Abdan K, Sapuan SM, and Khairul MD (2012) Mechanical properties and water absorption behavior of hybridized kenaf/pineapple leaf fiber reinforced high-density polyethylene composite *J. Compos. Mater.* **47**(8): 979-990.

27. Bakri B, Chandrabakty S, and Putra K A (2017) Evaluation of mechanical properties of Coir-Angustifolia Haw Agave fiber reinforced hybrid epoxy composite *Jurnal Mekanikal* **8**(1): 679-685.

Accepted Manuscript

28. Devi L U, Bhagawan S S, Nair K C M, and Thomas S (2011) Water absorption behavior of PALF/GF hybrid polyester composites *Polym. Compo.* doi: 10.1002/pc.21034.

29. Yahaya R, Sapuan S M, Jawaid M, Leman Z, and Zainudin E S (2016) Water absorption behavior and impact strength of kenaf-kevlar reinforced epoxy hybrid composites *Adv. Compo. Lett.* **25**(4): 98-102.

30. Rihayat T, Suryanil S, Fauzi T, Agusnar H, Wirjosentono B, Syafruddin Helmi Zulkifil, and Alam P N (2018) Mechanical properties evaluation of single and hybrid composites polyester reinforced bamboo, PALF, and Coir fiber *IOP Conf. Ser. Mater. Sci. Eng.* **334**, doi: 10.1088/1757-899X/334/1/012081.

31. Mittal M and Chaudhary R (2018) Experimental study on the water absorption and surface characteristics of alkali treated pineapple leaf fiber and coconut husk fiber *Int. J. Appl. Eng. Res.* **13**(15): 12237-12243.

Accepted Manuscript

Volume 6, Issue 3(2018) 212-213 ISSN 2347-3258
International Journal of Advance Research and Innovation

Geothermal Energy Resource of North-western Himalayas

Piyush Rawat, JP Kesari
Mechanical Engineering Department, Delhi Technological University, Delhi, India.

Article Info:
Article history
Received 18 August 2018
Received in revised form
28 August 2018
Accepted 02September 2018
Available online 15 September 2018

Keywords: Geothermal, energy, Himalaya, Environment

Abstract
Geothermal energy came into the picture after the oil crisis in the 1970s, with proper research and exploration took place in 1973. India's geothermal potential is entirely undeveloped with a power potential of 10,600 MWe. The capital cost of generating energy from geothermal sources in India is estimated to be US$1.6–2.0 million per MW, but the operating cost is minimal. This paper discusses geothermal heat source of different provinces of Jammu & Kashmir and Himachal Pradesh, with its direct use for production of electricity.

1. Introduction

Importance of geothermal energy came into the picture after the oil crisis in the year 1970s, with research and exploration started in 1973. In 1991 Geological Survey of India (GSI) prepared the 'Geothermal Atlas of India' describes 340 hot springs sites and identifies more than 300 sites with geothermal energy potential. The great Himalayas represents one of the best areas for geothermal energy and a detailed study of several thermal provinces [1] with geochemical characteristics of the thermal discharges and reservoir temperature has been estimated (Giggenabch, 1976; Giggenbach et. al., 1983; Nevada and Rao, 1991; Chandrasekharam et. al., 1989; 1992; 1996; Chandrasekharam and Antu, 1995; Chandrasekharam and Jayaprakash, 1996; Chandrasekharam et. al., 1997). These provinces are capable of generating 10,600 MW of power [1] with heat flow (>100 mW/m^2). Mixed magmatic fluids make the discharge saline (Alam et. al., 2004) with surface temperature varies 47 - 87^{o}C. These resources can be used for the direct application like space heating, drying crops, several industrial processes like pasteurizing milk and also for electricity production.

Importance of geothermal energy came into the picture after the oil crisis in the year 1970s, with research and exploration started in 1973. In 1991 Geological Survey of India (GSI) prepared the 'Geothermal Atlas of India' describes 340 hot springs sites and identifies more than 300 sites with geothermal energy potential. The great Himalayas represents one of the best areas for geothermal energy and a detailed study of several thermal provinces [1] with geochemical characteristics of the thermal discharges and reservoir temperature has been estimated (Giggenabch, 1976; Giggenbach et. al., 1983; Nevada and Rao, 1991; Chandrasekharam et. al., 1989; 1992; 1996; Chandrasekharam and Antu, 1995; Chandrasekharam and Jayaprakash, 1996; Chandrasekharam et. al., 1997). These provinces are capable of generating 10,600 MW of power [1] with heat flow (>100 mW/m^2). Mixed magmatic fluids make the discharge saline (Alam et. al., 2004) with surface temperature varies 47 - 87^{o}C. These resources can be used for the direct application like space heating, drying crops, several industrial processes like pasteurizing milk and also for electricity production.

2 Geothermal Heat Source

Himalayan region is one of the main sources of geothermal energy in India. It extends from the north-western part of India (Ladakh) to its north-eastern part (Assam) covering an area greater than 1500 km^2 and encloses over 150 thermal areas. The heat source available in Himalaya Geothermal Province is best suited for direct use as well as electricity generation [13].

2.1 Jammu and Kashmir

2.1.1 Puga and Chumathang

Located at an altitude of about 4400m above m.s.l., Puga Valley has more than a hundred of hot springs of temperatures range 28^{o}C to 84^{o}C [2] with the cumulative discharge of 30 l/s. It is the most promising and relatively high potential geothermal field in India located in the Indus Valley in eastern Ladakh. Geothermal resources are concentrated in an east-west elongated area of some 4 km^2, near the mouth of the 15 km long and 1 km wide Puga Valley. The area isat high altitude with snow desert during the winter season for about 4-5 months.

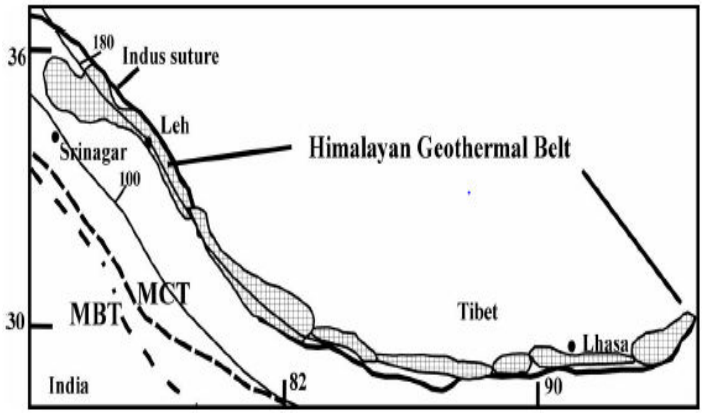


Fig.1: Regional structure and heat flow values (Varun Chandrasekhar and Chandrasekharam, 2007)

The surface temperature of the discharges at Puga ranges from 30^{o}C to 84^{o}C [2] with the boiling point of 84^{o}C at the altitude of (4400 m). More than 100 hot springs in an area of c. 5 km^2 with individual discharge rates of hot springs range from 30 to 40 l/s and an accumulative discharge of 300 l/s [3].

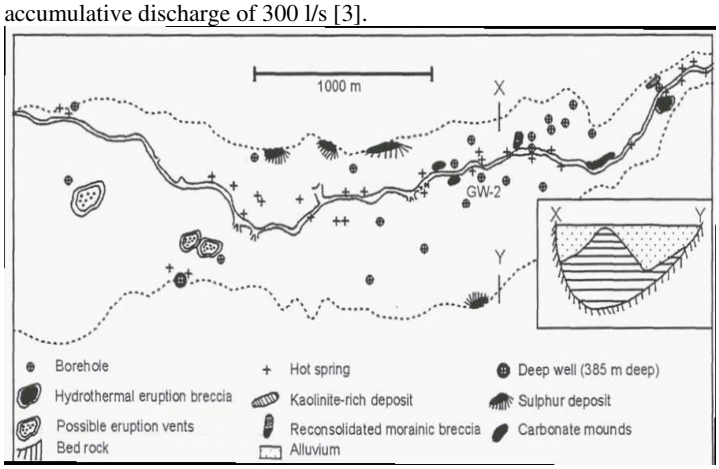


Fig. 2: Map showing surface alteration features of Puga valley geothermal system.

A pH range from 6.4 to 8.6 from the thermal water of the Puga geothermal field making thermal water neutral to weakly alkaline. Around 34 wells have been drilled in Puga geothermal field, with depths ranging from 28.5m to 384.7m [4] with 17 of 35 wells results in blow-outs. For 8 out of 17 flowing wells, wellhead measurements were taken with the discharge temperature of more than 120^{o}C.

***Corresponding Author,**
E-mail address: p.rawat8559@gmail.com
P. No.-+91-9654091741
All rights reserved: http://www.ijari.org

Discharge from a single a borehole was 1800 to 2400 l/min at Puga and 900 l/min at Chumanthang. Thermal studies and chemical thermometry suggest that at a depth of about 2.5 km below the Puga Valley temperature is more than 220°C [6]. While a reservoir modelling studies suggest that a temperature of up to 160°C at a depth of 450m [7].

Discharge surface temperature of Chumanthang range from 85 to 87°C and a boiling point of 87°C at an altitude of 3950m. The max recorded discharge is 90 l/min with the cumulative discharge of all the springs is 200 l/min [3]. 4 out of 6 wells blew out at Chumanthang with depths range 20 to 221m. The maximum temperature recorded is 110°C with reservoir temperature is expected to have at least 130°C. Around 73 hot springs have been identified with the max discharge of 90 l/min and 200 l/min of cumulative discharge [3]. The thermal water extracted from Chumathang has a slightly higher pH and a higher sulfate level. A geothermal power of 500MWt is considered to be available at Puga[5].

2.1.2 Shyok Valley

Shyok valley has hot springs at Panamik, Changlung, and Pulthang. Around 65 to 76°C of surface temperature with a discharge of 600 l/min were located at Panamik ranges. Two hot spring at Changlung has a discharge temperature of 61 and 66°C respectively with a cumulative discharge of 180 l/min, while the hot spring at Pulthang has a discharge of 180 l/min with a surface temperature of 28°C. Based on a subsurface temperature range of 130-150°C are considered at Panamik and Changlung with an installed capacity of 3-5 MWe at Panamik is considered feasible.

2.2 Himachal Pradesh

Himachal Pradesh geothermal provinces fall between Main Boundary Thrust (MBT) and Indo-Tsangpo Suture Zone (ITSZ) at an altitude of 1160 to 3660 m above the sea level [8] (Fig 3).

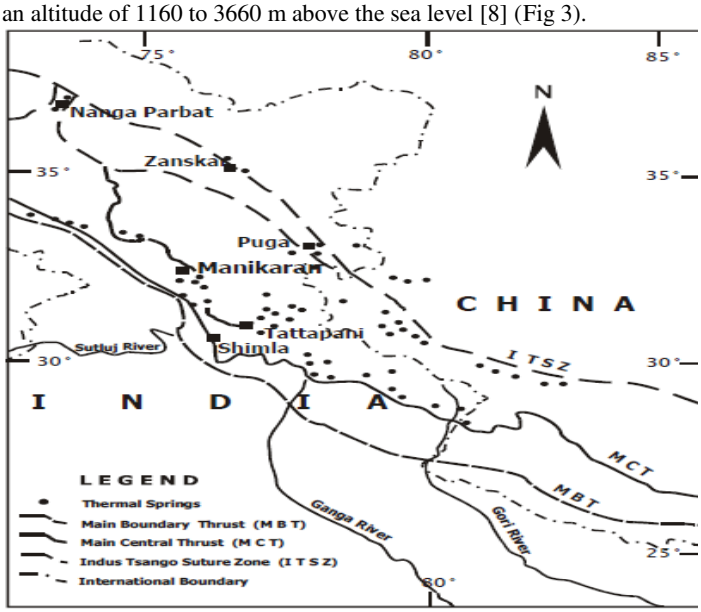


Fig. 3: Northwest part of Himalayas [9].

The Manikaran Quartzite inserted between the layers of phyllite constitutes the formation of the Rampur Group [10].Grey and green phyllite, with layers of carbonaceous schist and meta basics, are basal of Manikaran Quartzite [8]. Schist and gneiss of the Chail Group lie on the top of the Manikaran Quartzite [11].

Himachal Pradesh provinces reach a high geothermal gradient of about 260°C/km with a heat flow value ranges from 70 - 180 mW/m^2 [1]. The thermal springs of surface temperature vary from 57 to 98°C [12]. Thermal water from Manikaran was investigated and shows that it is considered as a mixture of two end members, one represented by calcium carbonate rich water and another by paleo-brine rich in NaCl [12]. It is a region with high altitudes and uneven surfaces, so it difficult to transmit power using conventional coal or hydropower grid. Though transmission cables are installed in rural areas, proper power demand is not fulfilled.

Himachal Pradesh is also famous for apple cultivation with more than 40% of total area and about 88% of total fruit production [9]. Processing of these fruits requires a high amount of energy which can be extracted from geothermal energy. Himachal Pradesh can become one of the major food producing and processing regions in the country with the proper use of geothermal energy [13-14]. For agribusiness industry, greenhouses, dehydration of fruits and vegetables and aquaculture are three primary uses of geothermal energy [15].

4. Conclusions

A great potential of geothermal energy lies within northern part of India, the Himalayas. The existing data on the geothermal resources indicates that direct application and power production is possible. It's a clean and renewable energy, hence development can occur without any hazardous effect on the environment.

References

[1] R Shankar. Heat flow map of India and its geological and economic significance. Indian Minerals 42, 1988, 89–110.

[2] J Craig, A Absar, G Bhat, G Cadel, M Hafiz, N Hakhoo, R Kashkari, J Moore, TE Ricchiuto, J Thurow, B Thusu. Hot springs and the geothermal energy potential of Jammu & Kashmir State, NW Himalaya, India, In Earth-Science Reviews, 126, 2013, 156-177.

[3] JL Thussu. Geothermal energy resources of India. Geological Survey of India Special Publication 69, 2002, 210.

[4] KJS Dua. Report on preliminary geothermal investigation, Jammu and Kashmir and Himachal Pradesh States, in: Report of the Geological Survey of India. Unpublished 1980.

[5] D Chandrasekharam. Geothermal energy resources of India: past and present. Proceedings of the World Geothermal Congress, Antalya, Turkey, 2005, 1–9.

[6] VK Saxena, F D'Amore. Aquifer chemistry of the Puga and Chumathang high temperature geothermal systems in India. Journal of Volcanological and Geothermal Research 21, 1984, 333–346.

[7] A Absar, V Kumar, IP Bajpal,AK Sinha, A Kapoor. Reservoir modeling of Puga geothermal system, Ladakh, Jammu and Kashmir. Geological Survey of India Special Publication 45, 1996, 69–74.

[8] BL Jangi, GPrakash, KJS Dua, JL Tussu, DB Dhimri, CS Pathak. Geothermal exploration of the Parbati valley geothermal field, Kulu district, Himachal Pradesh, India. In: Proceed. 2nd UN Symp. Development and Use of Geothermal Resources, San Francisco, 2, 1085-1094.

[9] D Chandrasekharam, MA Alam, A Minissale. Geothermal resource potential of Himachal Pradesh, India. International Geothermal Conference, Reykjavík, Sept. 2003.

[10] SV Srikantia, ON Bhargava. Geology of Himachal Pradesh. Geolo. Soc. India, 1998, 406.

[11] KA Sinha, DK Misra, SK Paul. Geology and tectonic features of Kulu and Spiti-Lahaul sector of NW Himalaya. Hmalayan Geol, Vol. 18, 1997, 1-16.

[12] AM Alam. Hydrogeochemistry of thermal springs in Manikaran, Kullu District, Himachal Pradesh (India). M.Sc. Dissertation (unpublished), IIT Bombay, 2002, 105.

[13] D Chandrasekharam. Use of geothermal energy for food processing-Indian status. Geo-heat Bull. 22, 2001, 8-12.

[14] D Chandrasekharam. Geothermal energy resources of India Country update. In: Proceedings, World Geothermal Congress, Japan 2000. E Iglesias, D Blackwell, T Hunt, J Lund, S Tamanyu, K Kimbara (Ed). 133-145.

[15] J Lund. Direct utilization of geothermal resources. In: Geothermal Energy Resources for Developing Countries, Dchandrasekharam, J Bundschuh (Ed)., AA Balkema, The Netherlands, 2002, 129-149.

IEEJ TRANSACTIONS ON ELECTRICAL AND ELECTRONIC ENGINEERING
IEEJ Trans 2018
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI:10.1002/tee.22821

Paper

Green Power Allocation for Cognitive Radio Networks with Spectrum Sensing

Ramnaresh Yadav*[a], Non-member
Ashwani Kumar*, Non-member
Keshav Singh**, Non-member

Green wireless communication has drawn ample interest in the 5G cellular networks. In this article, we investigate an energy-efficient power allocation with spectrum sensing for cognitive radio (CR) networks. The optimization problem is formulated as a ratio of the sum rate (SR) to the total power consumption in the signal transmission and spectrum sensing subject to the total power constraint. Because the original optimization problem is nonconvex, we adopt Dinkelback's method and propose an optimal energy-efficient power allocation scheme that iteratively improves the energy efficiency (EE) performance of the CR network and finally gives an optimal solution. In addition, an expression for the optimal solution is derived through the dual-decomposition method, and its performance is validated through simulation results. The impact of the probability of detection (P_D) and the probability of PU activeness (P_A) on the achievable average EE and SR is demonstrated when the CR network is designed from the energy-efficient perspective. © 2018 Institute of Electrical Engineers of Japan. Published by John Wiley & Sons, Inc.

Keywords: cognitive radio; spectrum sensing; energy efficiency; sensing time

Received 25 January 2017; Revised 1 June 2018

1. Introduction

With the rapid growth of wireless communication services, more and more spectral resources are needed. At present, most of spectral bands are licensed, but these spectral bands are underutilized. A cognitive radio (CR) has been developed to address the issues of paucity of spectral bands and underutilization of licensed frequency bands. In CR, the secondary user (SU) opportunistically utilizes the licensed spectrum keeping the interference to the primary user (PU) at an acceptable level.

The cognitive radio network (CRN) can be divided into three main paradigms: underlay, overlay, and interweave. The underlay paradigm allows cognitive users to operate if the interference caused to noncognitive users is below a given threshold. In overlay systems, the CRs use sophisticated signal processing and coding to maintain or improve the communication of noncognitive radios while also obtaining some additional bandwidth for their own communication. In interweave systems, the CRs opportunistically exploit spectral holes to communicate without disrupting other transmissions. Based on criteria such as interference, power, throughput, delay, and price, various combination access schemes have been developed for better resource utilization in CR [1–5].

Spectrum sensing is one of the most important tasks in CRN that helps to determine whether the PU is present or not, thus the SU can access the spectrum reasonably and the spectrum utilization can be enhanced at the same time. Different spectrum sensing techniques have been investigated in the literature to determine the presence or absence of the PU transmission. SU can either sense the spectrum continuously or on the fixed interval basis. With continuous spectrum sensing [6], SU grabs the spectrum opportunities completely but it requires more energy and hence it may not be energy efficient. On the other hand, the fixed spectrum sensing [7] requires less energy but SU may missed spectrum opportunities where the PU is inactive for a longer duration. Apart from energy required in spectrum sensing, the circuit energy consumption and energy required in data transmission are other factors that have a significant effect on the network performance. Recent research shows that the radio's power consumption reaches up to 50% of the mobile device's power consumption [8]. Hence, an energy-efficient reliable spectrum sensing in CRN for a higher sensing accuracy and energy-efficient data transmission are the most important key issues that need to be addressed.

Spectrum sensing has two associated performance metrics: the probability of detection (P_D) and the probability of false alarm (P_F). The probability that an SU declares the presence of a PU when the spectrum is occupied by the PU is called the probability of detection, whereas the probability that an SU declares the presence of the PU when the spectrum is idle is called the probability of false alarm. The probability of missed detection is $P_M = 1 - P_D$. The probability of detection increases with surging in the sensing period for a target probability of false alarm. The higher probability of detection indicates that the PU has better protection from SU. At the same time, for a target probability of detection, the probability of false alarm decreases when the sensing period increases. The lower probability of false alarm gives more opportunities to SU for data transmission. Therefore, sensing quality can be improved by using a longer sensing period. However, the increase in the sensing period results in a decline in the transmission period for the fixed frame length and overall it reduces the sum rate (SR) of the SU. The false alarms reduce spectral efficiency and missed detection causes interference with the PU. Thus, this indicates the existence of a trade-off between the sensing quality and the achievable SR [9,10]. This trade-off

[a] Correspondence to: Ramnaresh Yadav. E-mail: rnaresh@outlook.com

*Department of Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women, Delhi, India

**Institute for Digital Communications, School of Engineering, University of Edinburgh, Edinburgh, UK

© 2018 Institute of Electrical Engineers of Japan. Published by John Wiley & Sons, Inc.

can be utilized while allocating the optimal transmit power in CR together with spectrum sensing.

Recently, various spectrum sensing techniques were proposed for enhancing the sensing performance [10–14]. The optimal sensing time to maximize the achievable throughput under minimum detection probability constraint was determined in Ref. [10]. The adaptive spectrum sensing algorithm and sequential sensing policy for energy efficiency (EE) and guaranteed spectrum opportunity with reduced false alarm probability and limited missed detection probability were described in Ref. [11]. Effects of PU traffic on sensing throughput trade-off in CR networks was considered in Ref. [12]. The authors in Ref. [13] had extended [12] for MIMO CR networks under transmit power, probability of missed detection, and probability of false alarm constraints. SU's throughput was maximized for overlay CR networks by opportunistic spectrum sensing while limiting the disturbance to PU and characterizing the PU traffic by the Poisson birth-death process [14]. Transmission power is another important factor in CR that determines the system performance [15]. However, the authors in Refs [10–14] did not consider the power consumption model. In Ref. [16], the authors focused on the join optimization of sensing time and power for maximizing average throughput in a two-user amplify-and-forward overlay CR network. Statistical quality-of-service (QoS) driven power allocation for CR networks under PUs outage probability constraint was investigated in Ref. [17]. The authors in Ref. [18] proposed an energy-efficient transmission scheme in CR networks, but did not consider the sensing duration of CRs, which determines the sensing accuracy. In Ref. [19], the energy consumption in sensing for low-power sensor networks was optimized. The authors in Ref. [20] jointly optimized sensing and data transmission durations in an energy-efficient manner without using relay. The authors in Ref. [21] studied the design of sensing-access strategies and sensing order to improve EE in CR networks. In Ref. [22], the problem of energy-efficient sensing and transmission was formulated as a discrete time Markov decision process. The objective was to minimize both energy consumption and delay cost while meeting a target traffic payload through appropriate rate allocation. Authors in Ref. [23] focused on energy-efficient spectral sensing in cognitive relay networks. In Ref. [24], the authors calculated optimal sensing and data transmission duration to maximize EE under PU detection probability and interference constraints. It is worth mentioning here that the above works did not focus on joint energy-efficient power allocation and spectrum sensing in CR networks.

To the best of our knowledge, there has been no work reported yet on designing an energy-efficient power allocation scheme with spectrum sensing under probability of missed detection and false alarm constraint for the CR network. In this research work, we propose an energy-efficient power allocation scheme with spectrum sensing that takes probability of detection and probability of false alarm into consideration. The contributions of this article are summarized as follows:

- Based on probabilities associated with the sensing process, we propose a new method to utilize PUs subcarriers. These probabilities, P_{D}, P_{A}, and P_{F} are performance parameters of the sensing process. Higher the probability of detection (P_{D}), better the PUs are protected. However, from the perspectives of SUs, the lower the probability of false alarm (P_{F}), the more chances the channel can be reused when it is available, thus the higher the achievable throughput for the secondary network. Probability of PU activeness (P_{A}) increases interference to SU and also decreases its throughput. We utilize this trade-off between the sensing quality and the achievable SR while using PUs subcarriers and allocating the optimal transmit power.
- The optimization problem is formulated as a ratio of the achievable SR to the total power consumption in the signal transmission and spectrum sensing subject to the total power constraint. Interference is considered based on missed detection and reoccupancy of PU. We propose an optimal energy-efficient power allocation scheme that iteratively improves the EE performance and finally reaches an optimal solution. In addition, an expression for optimal solution is derived through dual decomposition.
- In contrast to Ref. [25], we extend the proposed system model to another two scenarios namely: (i) sum-rate maximization (SRM) and (ii) robust power allocation where CR users do not have a perfect channel state information (CSI) knowledge. Accordingly, we propose SRM and robust power allocation policies.
- The performance of the proposed scheme is validated through simulation results. The impact of P_{A}, P_{D} and α (sensing time) on the average EE and SRM performance of the considered CR network is demonstrated. Finally, the trade-off between the EE and the SRM is presented.

The remaining of this article is organized as follows. In Section 2, we introduce the system model and the problem formulation. The power allocation scheme for achieving the maximum EE is presented in Section 3. The complexity analysis is described in Section 4. In Section 6, numerical results are presented and finally conclusion is drawn in Section 7.

2. System Model

2.1. Transmission model As shown in Fig. 1, the CR network consists of single SU and PU. All users in the CR network are assumed to have a single antenna and the SU always has data to transmit. The sensing duration is far less than the frame duration. The CR network is assumed to operate on a frame-by-frame basis as shown in Fig. 2. All channels are assumed to be flat-fading and perfectly known to CR users. Before transmitting data, the SU transmitter (SU Tx) performs spectrum sensing at the beginning of each frame. Each frame is divided into two folds: (i) sensing time with a duration of α; and (ii) data transmission time with a duration of $(T-\alpha)$, where the total frame duration is T, as shown in Fig. 2.

Let us define G_i as the downlink channel from the SU Tx to the SU receiver (SU Rx) on ith subcarrier, for $i = 1, 2, \ldots, N$, where N denotes the available limited number of subcarrier for SU. Furthermore, the transmit power allocated at the ith subcarrier for the SU is given by P_i. From the capacity formula [26], the achievable SR or average reward during data transmission time $(T-\alpha)$ for SU user can be expressed as

$$R_{\mathrm{s}} = (T-\alpha) \sum_{i=1}^{N+\lfloor N_{\mathrm{add}} \rfloor} \log_2\left(1+\frac{P_i G_i}{\sigma^2}\right) \tag{1}$$

where σ^2 denotes the noise power. The additional subcarriers, N_{add}, sensed by SU Tx are given by

$$N_{\mathrm{add}} = [(1-P_{\mathrm{F}})(1-P_{\mathrm{A}}) + (1-P_{\mathrm{D}})P_{\mathrm{A}}] \times M \tag{2}$$

where P_{F}, P_{A}, P_{D}, and M indicate the probability of false alarm, the probability of PU activeness, the probability of detection, and subcarriers of PU, respectively. P_{A} and P_{F} are the two-performance metrics in the spectrum sensing. The probability of detection, P_{D}, increases with increasing the sensing period for a target probability of false alarm.

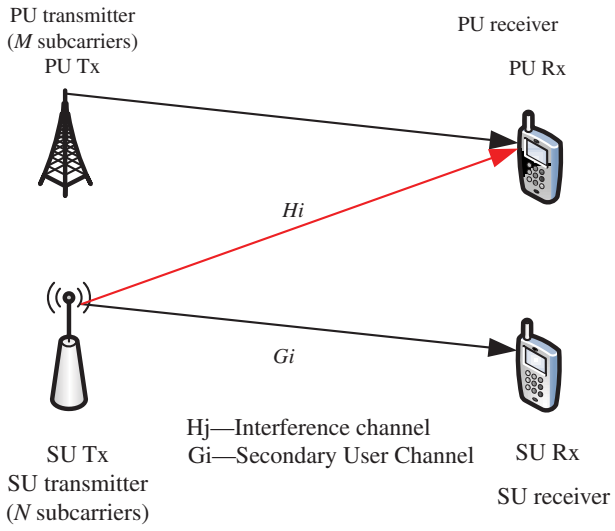


Fig. 1. Cognitive radio networks

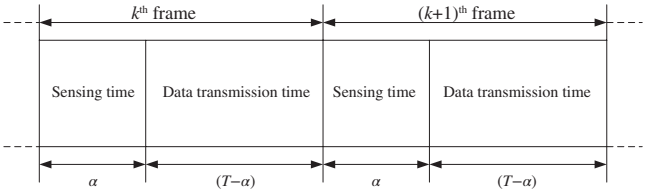


Fig. 2. Frame structure for cognitive radio networks with spectrum sensing

2.2. Energy consumption model The number of bits that can be transmitted per joule is a novel metric for designing an energy-efficient CR network. The energy consumption in the CR network depends on the frame structure. To design an energy-efficient CR network, it is important to take both transmit and spectrum sensing power into account. Spectrum sensing power, P_s, and transmission power, (P_i), of SU are two main parts in the power consumption. In general, the transmit power behaves dynamically with respect to the instantaneous channel gains and is exclusively used for data transmission to attain reliable communications. The power used in spectrum sensing, P_s, is fixed and is very less compared to data transmission power. The circuit power, P_c, is negligible in the considered CR network. The total energy consumed in one frame can be calculated as

$$E = (T-\alpha)\sum_{i=1}^{N+\lfloor N_{\text{add}}\rfloor} P_i + \alpha P_s. \tag{3}$$

here, P_s represents power used in spectrum sensing and P_i represents power used in transmission on ith subcarrier.

2.3. Problem formulation The EE of the considered CR network is defined as the ratio of the achievable SR over the total power consumption. Hence, by using (1) and (3), the EE is given as [27]

$$\eta_{\text{EE}} = \frac{(T-\alpha)\sum_{i=1}^{N+\lfloor N_{\text{add}}\rfloor}\log_2\left(1+\frac{P_iG_i}{\sigma^2}\right)}{(T-\alpha)\sum_{i=1}^{N+\lfloor N_{\text{add}}\rfloor}P_i+\alpha P_s} \tag{4}$$

The optimal power allocation policy $P_k^\star$, for $k = 1, 2, \ldots, N + \lfloor N_{\text{add}}\rfloor$, can be obtained by solving following optimization problem subject to the two conditions: (i) to limit the transmission power of the SU Tx; and (ii) by ensuring the interference power from SU to PU within a threshold limit:

$$(\mathbf{P}.1)\ \max_{\mathbf{P}} \frac{\overbrace{(T-\alpha)\sum_{i=1}^{N+\lfloor N_{\text{add}}\rfloor}\log_2\left(1+\frac{P_iG_i}{\sigma^2}\right)}^{F(\mathbf{P})}}{\underbrace{(T-\alpha)\sum_{i=1}^{N+\lfloor N_{\text{add}}\rfloor}P_i+\alpha P_s}_{F_{\text{TP}}(\mathbf{P})}} = \max_{\mathbf{P}}\frac{F(\mathbf{P})}{F_{\text{TP}}(\mathbf{P})} \tag{5}$$

$$\text{s.t.}\quad (C.1)(T-\alpha)\sum_{i=1}^{N+\lfloor N_{\text{add}}\rfloor}P_i+\alpha P_s \le P_{\max},$$

$$(C.2)\sigma^2 + P_jH_j \le I_o; \quad \text{for } j = 1, 2 \ldots \lfloor N_{\text{add}}\rfloor$$

$$(C.3)P_i \ge 0, \quad \text{for } i = 1, 2 \ldots (N + \lfloor N_{\text{add}}\rfloor)$$

where $\mathbf{P} = [P_1, P_2, \ldots, P_N, \ldots, P_{N+\lfloor N_{\text{add}}\rfloor}]$, I_0 is the predefined threshold with respect to the interference power, $P_{\max}$ is the maximum power budget of SU, and H_j is the interfering channel from SU Tx (Secondary User Transmitter) to PU Rx (Primary User Receiver), for $j = 1, 2, \ldots, \lfloor N_{\text{add}}\rfloor$. Interference arises due to the missed detection of the spectrum currently utilized by PU. If, after spectrum sensing, SU fails to detect the presence of PU then it will interfere with transmission of PU and effect the QoS of PU. Hence, the constraint $(C.2)$ controls the interference power level generated by SU toward the PU receiver and keeps it less than the predefined threshold value I_0.

3. Optimal Power Allocation Scheme

It can be observed from (5) that the optimization problem is nonconvex due to the fractional form of the objective function and there is no standard convex optimization approach to solve this nonconvex problem. To find an efficient power allocation scheme, we first transform the objective function of (5) into a parametrized convex form by applying Dinkelbachs' method [28]. The maximum EE, $q^\star$, of the considered CR network is defined as

$$q^\star = \frac{F(\mathbf{P}^\star)}{F_{\text{TP}}(\mathbf{P}^\star)} = \max_{\mathbf{P}}\frac{F(\mathbf{P})}{F_{\text{TP}}(\mathbf{P})} \tag{6}$$

Using (6), the optimization problem (5) can be rewritten as

$$(\mathbf{P}.2)\ \max_{\mathbf{P}}\ F(\mathbf{P}) - q^*F_{\text{TP}}(\mathbf{P})$$

$$\text{s.t. } (C.1),\ (C.2),\quad \text{and } (C.3) \tag{7}$$

Theorem. The maximum EE, $q\star$, of the considered network can be achieved, if and only if the following balance condition hold

$$\max_{\mathbf{P}}\ F(\mathbf{P}) - q^\star F_{\text{TP}}(\mathbf{P})$$

$$= F(\mathbf{P}^\star) - q^\star F_{\text{TP}}(\mathbf{P}^\star) = 0 \tag{8}$$

Proof: The proof of this theorem is provided through proof of forward and converge implication of Theorem.

Forward implication: Let S be a feasible set of the problem (**P. 1**). Consider $\mathbf{P}^\star$ as the optimal power allocation with respect to maximum EE, $q^\star$, we have

$$q_a = \frac{F(\mathbf{P}^\star)}{F_{\text{TP}}(\mathbf{P}^\star)} \ge \frac{F(\mathbf{P})}{F_{\text{TP}}(\mathbf{P})}, \quad \forall \mathbf{P} \in S \tag{9}$$

Hence, as a result

$$F(\mathbf{P}) - q_a F_{\mathrm{TP}}(\mathbf{P}) \leq 0 \ , \quad \forall \mathbf{P} \in S \tag{10}$$

$$F(\mathbf{P}^\star) - q_a F_{\mathrm{TP}}(\mathbf{P}^\star) = 0 \tag{11}$$

From (10) and (11), we conclude that we can achieve the maximum EE, q_{a}, through the optimal power ($\mathbf{P}^\star$).

Convergence implication: Let ($\mathbf{P}^\star$) be a solution of the optimization problem (7) such that it satisfies the balance equation, i.e. $F(\mathbf{P}^\star) - q^\star F_{\mathrm{TP}}(\mathbf{P}^\star) = 0$, it implies

$$\begin{aligned} &F(\mathbf{P}) - q^\star F_{\mathrm{TP}}(\mathbf{P}) \\ &\leq F(\mathbf{P}^\star) - q^\star F_{\mathrm{TP}}(\mathbf{P}^\star) = 0 \ , \quad \forall \mathbf{P} \in S \end{aligned} \tag{12}$$

$$\Rightarrow \qquad \frac{F(\mathbf{P})}{F_{\mathrm{TP}}(\mathbf{P})} \leq q^\star = \frac{F(\mathbf{P}^\star)}{F_{\mathrm{TP}}(\mathbf{P}^\star)} \ , \quad \forall \mathbf{P} \in S \tag{13}$$

Therefore, it is proven that $q^\star$ that satisfies the balance equation is the optimal EE of the considered CR network.

Lemma 1. $g(\mathbf{P},q)$ is a concave function of $\mathbf{P}$ for given q.

Proof: Function $g(\mathbf{P}, q)$ can be written as follows

$$g(\mathbf{p}, q) = \begin{cases} F(\mathbf{P}) - qF_{\mathrm{TP}}(\mathbf{P}) & \text{for} p_i > 0, \quad \forall i \\ 0 \ ; & \text{otherwise} \end{cases} \tag{14}$$

The Hessian of $g(\mathbf{p}, q)$ is given by

$$\nabla^2 g(\mathbf{P}, q) = -\left(\frac{T-\alpha}{\ln 2}\right) \frac{G_i^2}{(\alpha^2 + p_i G_i)^2} \leq 0$$

i.e. $\nabla^2 g(\mathbf{P}, q)$ is negative. Therefore, $g(\mathbf{P}, q)$ is a concave function of $\mathbf{P}$. ■

Lemma 2. $g(q) = \max\{F(\mathbf{P}) - qF_{\mathrm{TP}}(\mathbf{P}) — \mathbf{P} \in \mathbf{S}\}$ is strictly monotonic decreasing i.e. $g(\mathbf{P}, q_2) < g(\mathbf{P}, q_1)$; if $q_1 < q_2$.

Proof: Let $\mathbf{P}^* \in \mathbf{S}$ be the optimal solution that maximize $g(q_2)$. Assuming $F_{\mathrm{TP}}(\mathbf{P}) > 0$, we have

$$\begin{aligned} g(q_2) &= \max\{F(\mathbf{P}) - q_2 F_{\mathrm{TP}}(\mathbf{P}) | \mathbf{P} \in \mathbf{S}\} \\ &= F(\mathbf{P}^*) - q_2 F_{\mathrm{TP}}(\mathbf{P}^*) \\ &\leq F(\mathbf{P}^*) - q_1 F_{\mathrm{TP}}(\mathbf{P}^*) \\ &= \max\{F(\mathbf{P}) - q_1 F_{\mathrm{TP}}(\mathbf{P}) | \mathbf{P} \in \mathbf{S}\} \\ &= g(q_1) \end{aligned}$$

■

Lemma 3. Let $p^\star \in \mathbf{S}$ be a solution for $p^\star \in \mathbf{S}$ and $\widehat{q} = \frac{F(p^\star)}{F_{\mathrm{TP}}(p^\star)}$; Then, we have $g(\widehat{q}) \geq 0$.

Proof: $g(\widehat{q}) = \max\{F(\mathbf{P}) - \widehat{q} F_{\mathrm{TP}}(\mathbf{P}) | \mathbf{P} \in \mathbf{S}\}$
$\geq F(p^\star) - \widehat{q} F_{\mathrm{TP}}(p^\star) = 0$
Hence $\quad g(\widehat{q}) > 0$ ■

Theorem 2. For optimal p^*, if the parameter, q, can be updated in $(t+1)$th iteration as

$$q_{t+1} = \frac{F(p_t^*)}{F_{\mathrm{TP}}(p_t^*)}$$

then, q is a monotonic increasing function i.e. $q_{t+1} \geq q_t$.

Proof: First, we need to prove that $q_{t+1} > q_t$; $\forall t$ with $g(q_t) > \delta$. According to Lemma , $g(q_t) > 0$. By definition, we know that

$$\begin{aligned} &F(p_t) = q_{t+1} F_{\mathrm{TP}}(p_t) \\ \text{Hence} \qquad &g(q_t) = F(p_t) - q_t F_{\mathrm{TP}}(p_t) \\ &= q_{t+1} F(p_t) - q_t F_{\mathrm{TP}}(q_t) > 0 \end{aligned}$$

Table I. An iterative algorithm for energy-efficient power allocation in CR networks

Input:
$\mathbf{P} \leftarrow \mathbf{P}^{(0)}$; Convergence $= 0$;
Set the tolerance $\delta = 10^{-5}$;
Output:
Maximum energy efficiency: $q^\star$;
Optimal power allocation scheme: ($\mathbf{P}$).
Iterative Algorithm:
1. **repeat**
2. Update $\mathbf{P}$ by solving the optimization problem (**P. 2**)
3. **if** $F(\mathbf{P}) - qF_{\mathrm{TP}}(\mathbf{P}) \leq \delta$
4. $\mathbf{P}^\star = \mathbf{P}$, and $q^\star = q$;
convergence $= 1$;
break;
5. **else**
6. update q as follows:
$q = \frac{F(P^\star)}{F_{TP}(P^\star)}$
7. **end if**
8. **until** convergence $= 1$

As $F_{\mathrm{TP}}(p_t) > 0$, this implies that $q_{t+1} > q_t$.

Next, we need to show that $\lim_{t\to\infty} q_t = q(p_0) = q_0$. If this is not true, we must have $\lim_{t\to\infty} q_t = q^* < q_0$. By construction, we have a sequence p_t^* with q_t^*, such that $\lim_{t\to\infty} g(q_t^*) = g(q^*) = 0$. As $g(q)$ decreases monotonically, we get $0 = g(q^*) > g(q_0) = 0$. Hence, it follows that $\lim_{t\to\infty} g(q_t) = g(q_0)$ and $g(q)$ is continuous. Resulting in $\lim_{t\to\infty} q_t = q_0$.

The problem (**P. 2**) is a typical concave maximization problem and the optimal solution of this problem can be obtained using standard convex optimization techniques [29]. Here, the dual decomposition method is used to obtain optimal solution. The Lagrangian function of the primal problem (**P. 2**) is given by

$$\begin{aligned} \mathcal{L}(\lambda, \boldsymbol{\mu}, \mathbf{P}) &= F(\mathbf{P}) - qF_{\mathrm{TP}}(\mathbf{P}) \\ &-\lambda \left(\sum_{i=1}^{N+\lfloor N_{\mathrm{add}} \rfloor} P_i - \frac{(P_{\max} - \alpha P_{\mathrm{s}})}{(T-\alpha)} \right) \\ &- \sum_{j=1}^{\lfloor N_{\mathrm{add}} \rfloor} \mu_j \left(P_j - \frac{(I_0 - \sigma^2)}{H_j} \right) \end{aligned} \tag{15}$$

where λ is the Lagrangian multiplier associated with the constraint ($C.1$) and $\boldsymbol{\mu}$ is the Lagrangian multiplier vector accounting for constraint ($C.2$) with element μ_j for $j = 1, 2, \ldots, \lfloor N_{\mathrm{add}} \rfloor$. The iterative algorithm is summarized in Table I.

The dual problem may be written as

$$\min_{\lambda, \boldsymbol{\mu} \geq 0} \quad \max_{\mathbf{P}} \quad \mathcal{L}(\lambda, \boldsymbol{\mu}, \mathbf{P}) \tag{16}$$

The dual problem in Ref. [16] can be solved iteratively by solving subproblems for obtaining the power and by solving a master problem using the gradient method to update the Lagrangian multipliers λ and $\boldsymbol{\mu}$, until convergence is obtained.

3.1. Subproblem solution

The Karush–Kuhn–Tucker (KKT) conditions [29], which are first-order necessary and sufficient conditions for optimality, can be utilized for solving the subproblem. Taking the partial derivative of (15) with respect to P_i for fixed λ and $\boldsymbol{\mu}$ and equating the result to zero, i.e.

$$\frac{\partial \mathcal{L}(\lambda, \boldsymbol{\mu}, \mathbf{P})}{\partial P_i} = 0 \tag{17}$$

The optimal value of $P_i^{\star}$ can be obtained from [17] as

$$P_i^{\star} = \begin{cases} \left[\frac{\Phi}{\left(q+\frac{\lambda}{(T-\alpha)}\right)} - \frac{\sigma^2}{G_i}\right]^+; & \text{if } i \in \{1,2,\ldots,N\} \\ \left[\frac{\Phi}{\left(q+\frac{\lambda+\mu_j}{(T-\alpha)}\right)} - \frac{\sigma^2}{G_i}\right]^+; & \text{if } i \in \{N+1,\ldots,\lfloor N_{\text{add}}\rfloor\} \end{cases} \tag{18}$$

where $\Phi = \frac{1}{\ln 2}$ and $[x]^+ = \max(0,x)$. The optimal power allocation solution of the optimization problem (18) is in the form of dynamic water filling where the inverse of the parameter q plus the Lagrangian multiplier(s) can be regarded as a water-filling level that has to be chosen to meet the total power and/or interference temperature constraints. Using the bisection method or the gradient method, we can obtain the optimal values of Lagrangian multipliers λ and $\boldsymbol{\mu}$.

3.2. Solution of the master dual problem Because the dual function is differential, the gradient method can be used to solve the master problem in (16), i.e. to find λ and $\boldsymbol{\mu}$ for given $\mathbf{P}$. By taking the partial derivative of (15) with respect to λ and μ_j, we get

$$\frac{\partial \mathcal{L}(\lambda,\boldsymbol{\mu},\mathbf{P})}{\partial \lambda} = -\left(\sum_{i=1}^{N+\lfloor N_{\text{add}}\rfloor} P_i - \frac{(P_{\max}-\alpha P_s)}{(T-\alpha)}\right) \tag{19}$$

$$\frac{\partial \mathcal{L}(\lambda,\boldsymbol{\mu},\mathbf{P})}{\partial \mu_j} = -\left(P_j - \frac{\left(I_0-\sigma^2\right)}{H_j}\right) \tag{20}$$

From (19) and (20), the gradient update equations are given by

$$\lambda(1+t) = \left[\lambda(t) + \xi_1(t)\left(\sum_{i=1}^{N+\lfloor N_{\text{add}}\rfloor} P_i - \frac{(P_{\max}-\alpha P_s)}{(T-\alpha)}\right)\right]^+ \tag{21}$$

$$\mu_j(1+t) = \left[\mu_j(t) + \xi_2(t)\left(P_j - \frac{\left(I_0-\sigma^2\right)}{H_j}\right)\right]^+ \quad \forall j \tag{22}$$

where $j = 1, 2, \ldots, \lfloor N_{\text{add}}\rfloor$., $t \geq 0$ and $\xi_k(t)$, $k = 1, 2$ are the iteration number and positive step sizes, respectively. Through the iterative producer of (18), (21), and (22), we can obtain the optimal power allocation $\mathbf{P}^{\star}$ of the problem (**P. 2**). The process of obtaining the optimal power $\mathbf{P}^{\star}$ and updating the Lagrangian multipliers λ and μ_j is repeated until convergence is reached. Because the transformed problem (**P. 2**) is concave, the duality gap between dual optimal and primal optimal is zero. Therefore, the solution of the dual problem is equivalent to the solution of the primal problem.

4. Complexity Analysis

The complexity of the proposed scheme depends on the number of constraints, power allocations, and updating the Lagrangian multipliers [30]. The complexity of the total power constraint and optimal power allocation is given by $\mathcal{O}(K^3)$, where K indicates the maximum power level of SU [31]. Moreover, the complexity of updating a dual variable is $\mathcal{O}(K^{\varrho})$ (for example: $\varrho = 2$ for ellipsoidal method), thus total complexity of updating the dual variables is $\mathcal{O}(3K^{\varrho})$. If the dual problem converges in W iterations and the users operate over a time interval T, then total complexity of the optimal algorithm is given by $\mathcal{O}(6N_c TWK^{\varrho}(K^3))$.

5. Model Extension

5.1. SR maximization In the SRM case, we maximize total achievable SR of the proposed system. It can be observed from (P.2) that when 'q' approaches zero, the EEM problem resembles as a SRM problem that can be written as

$$\begin{aligned} \max_{\mathbf{P}} \quad & (T-\alpha)\sum_{i=1}^{N+\lfloor N_{\text{add}}\rfloor} \log_2\left(1+\frac{P_i G_i}{\sigma^2}\right) \\ \text{s.t.} \quad & (C.1)(T-\alpha)\sum_{i=1}^{N+\lfloor N_{\text{add}}\rfloor} P_i + \alpha P_s \leq P_{\max}, \\ & (C.2)\sigma^2 + P_j H_j \leq I_o; \quad \text{for } j = 1,2,\ldots,\lfloor N_{\text{add}}\rfloor \\ & (C.3)P_i \geq 0, \quad \text{for } i = 1,2,\ldots,(N+\lfloor N_{\text{add}}\rfloor). \end{aligned} \tag{23}$$

5.2. Robust power allocation In this subsection, we consider a more practical scenario where CR users do not have a perfect CSI knowledge [32,33]. In practice, perfect CSI is difficult to obtain due to the inherently time-varying nature of wireless channels, delay, noise, and limited bandwidth in feedback channels and other factors. Our proposed design framework can be easily extended to accommodate this scenario by replacing the channels with the following:

$$\widehat{G}_i = G_i + \varepsilon_i, \quad \forall i \tag{24}$$

$$\widehat{H}_j = H_j + \zeta_j, \quad \forall j \tag{25}$$

where $\widehat{G}_i$ and $\widehat{H}_j$ are the estimated channels and $\varepsilon_i \sim \mathcal{CN}(0,\sigma_{\text{e}}^2)$ and $\zeta_j \sim \mathcal{CN}(0,\sigma_{\text{e}}^2)$ are the channel estimation errors. Our proposed power allocation schemes obtained under imperfect CSI are useful in practical settings and give insights into the impact of channel uncertainty on the system performance. At the same time, the EE in the presence of perfect CSI of the transmission and interference links serves as a baseline to compare the performances attained under the assumption of imperfect CSI.

6. Simulation Results and Discussions

In this section, we evaluate the performance of the proposed power allocation scheme by computer simulations. The channel gains G_i and H_j are assumed to have a Rayleigh distribution and assumed to be perfectly known [34]. The sensing power, P_{s}, and the maximum available transmit power, $P_{\max}$, for the SUs are 0.1 and 10 mW, respectively. The value of noise power σ^2 is set to be 0.01. The maximum preset interference threshold, I_{o}, at PU is 2.5 mW. In the proposed scheme, the maximum number of iterations, $I_{\max}$, is 15 and the time duration for each frame is $T = 1$ second (s). The detection probability, P_{D}, and the false alarm probability, P_{F}, are set to 0.9 and 0.01, respectively [35–37].

Figure 3 shows the convergence behavior of the proposed power allocation scheme in the CR network for a single-channel realization. The total number of subcarriers, N, for SU is 5, while PU has $M = 256$ subcarriers. The sensing time, α, is 0.15 s. It is shown that the obtained EE is monotonically increased with the number of iterations, and the number of iterations required for attaining a convergence performance is less than five. As seen in Fig. 3, the EE is enhanced with decreasing P_{A} (i.e., increasing additional subcarriers, N_{add}).

The impact of P_{A} on the average EE and SR performance is presented in Fig. 4. The total number of subcarriers at SU and PU are set to $N = 5$, and $M = 256$. The values of P_{F} and P_{D} are 0.01 and 0.9, respectively, whereas P_{A} values are 0.1 and 0.5. In Fig. 4, the total number of additional subcarriers, N_{add}, are 231

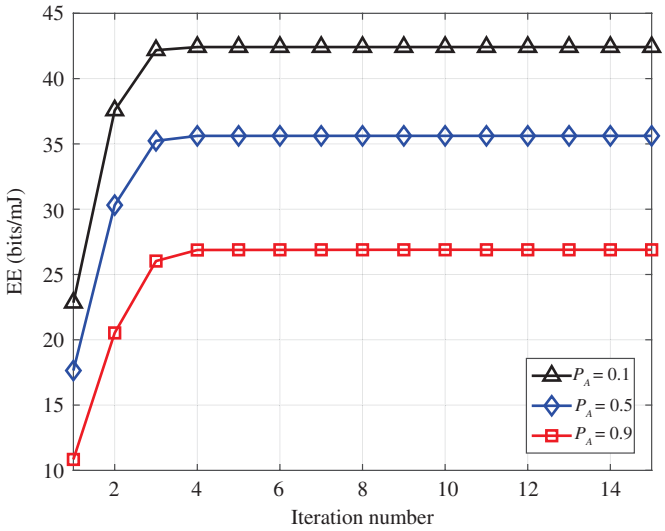


Fig. 3. Convergence behavior of the proposed iterative algorithm with $P_{\max} = 10$ mW, $N = 5$, $M = 256$, and $\alpha = 0.15$ s

for $P_{\mathrm{A}} = 0.1$ and 140 for $P_{\mathrm{A}} = 0.5$. As expected, upon increasing α, the average EE performance of the EE maximization (EEM) scheme is gradually decreased as shown in Fig. 4(a). It can be observed that for small sensing time (i.e., $\alpha < 0.6$), the average EE performance of the EEM scheme with $P_{\mathrm{A}} = 0.1$ is much better than the EEM scheme with $P_{\mathrm{A}} = 0.5$, whereas for a higher value of α (i.e., $\alpha > 0.65$ s), the SU spends more time on spectrum sensing and hence there is no or very less effect of P_{A} on the average EE performance. Figure 4(b) provides the effect of sensing time on the average SR. Average SR increases with decreasing of P_{A}. This also shows the similar trend like Fig. 4(a) for a higher value of α (i.e., $\alpha > 0.75$ s). This is because the SU user gets less time for data transmission.

Figure 5 shows the impact of P_{D} on the average EE and SR performance. P_{D} denotes the probability that SU declares the presence of PU when the spectrum is actually occupied by PU for a given probability of PU activeness (P_{A}). The total number of subcarriers at SU and PU are set to $N = 5$ and $M = 256$. The values of P_{F} and P_{A} are 0.01 and 0.1, respectively, whereas P_{D} values are 0.1 and 0.5. In Fig. 5, the total number of additional subcarriers, N_{add}, are 251 for $P_{\mathrm{D}} = 0.1$ and 241 for $P_{\mathrm{D}} = 0.5$. As expected, upon increasing α, the average EE performance of the EEM scheme is gradually decreased as shown in Fig. 5(a). It can also be observed that for a higher value of P_{D}, we have a better EE performance. This is in accordance with our expectation because with higher P_{D} there will be more accurate detection of PU and hence less interference. It can be observed that for a small sensing time (i.e., $\alpha < 0.5$), the average EE performance of the EEM scheme with $P_{\mathrm{D}} = 0.5$ is slightly better than the EEM scheme with $P_{\mathrm{D}} = 0.1$, whereas for a higher value of α (i.e., $\alpha > 0.55$ s), the SU spends more time on spectrum sensing and hence there is no or very less effect of P_{D} on average EE performance. Figure 5(b) provides the effect of sensing time on the average SR. Average SR increases with the value of P_{D}. This also shows a similar trend to Fig. 5(a) for a higher value of α (i.e., $\alpha > 0.75$ s). This is because the SU user gets less time for data transmission.

The EE ratio is defined as $\Gamma_r = \frac{\eta_{\mathrm{CR}}}{\eta_o}$, where η_{CR} indicates the EE of the CR network, whereas η_{o} presents the EE of non-CR networks. Similar to the sensing throughput trade-off [10], we can observe the sensing-energy-efficiency trade-off phenomenon in Fig. 6. When sensing time is small, the EE of CR networks is increased with the decreasing active probability P_{A}. The small value of P_{A} allows the SU to acquire a higher number of PU's subcarriers for data transmission. The small P_{A} and sensing time

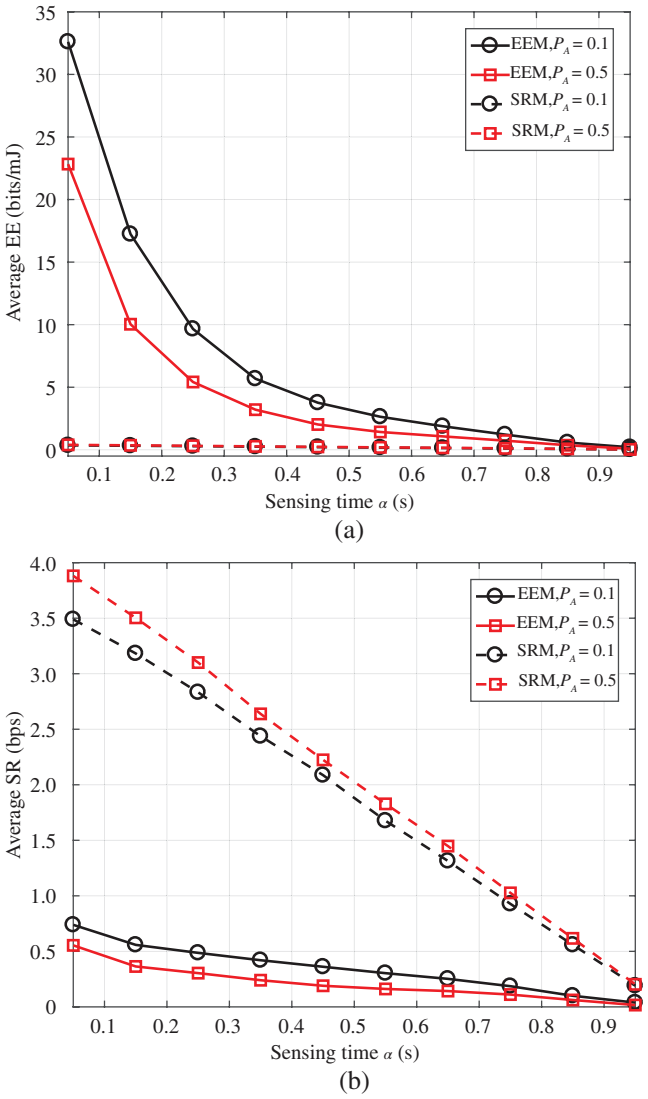


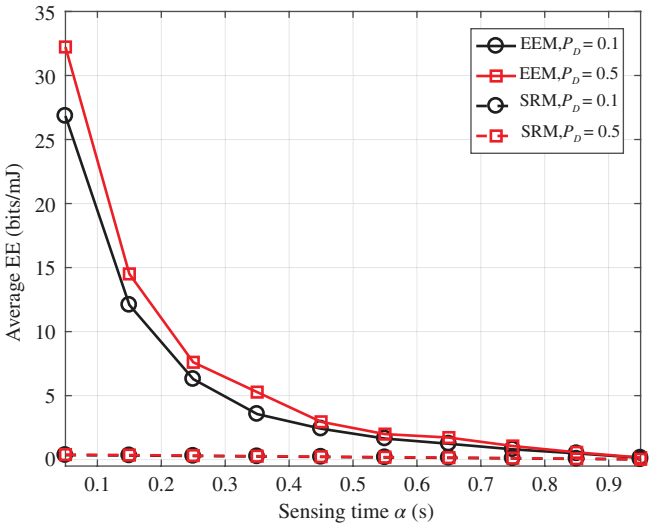
Fig. 4. Impact of P_{A} on the average EE and SR performance ($P_{\max} = 10$ mW, $N = 5$, and $M = 256$). (a) Average EE versus sensing time α. (b) Average SR versus sensing time α

allow the SU to achieve higher diversity with negligible sensing energy cost. On the other hand, when the sensing time is relatively large with lower P_{A}, more subcarriers do not help CR networks to enhance the EE and SE performance.

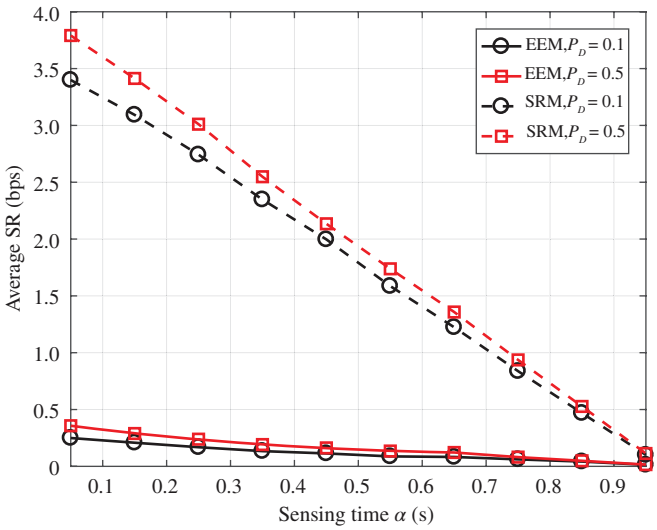
Figure 7 shows the effect of interference threshold I_{o} on average EE and average SR performance in the perfect and imperfect CSI scenarios. The parameter settings are as follows. We set $N = 5$, sensing time $\alpha = 0.5$, $P_{\mathrm{F}} = 0.5$, $P_{\mathrm{D}} = 0.5$, $P_{\mathrm{A}} = 0.1$, $P_{\mathrm{s}} = 0.1$, $M = 256$, and $\sigma_{\mathrm{e}}^2 = 0.5$. I_{o} varies between 0.05 and 0.4 [37]. We can observe that the average EE and SE performance increases with the increasing value of interference limit I_{o}. However, the average EE and SE performance of the algorithm in the case of imperfect CSI is lower than that of the perfect CSI scenario due to the channels estimation errors.

7. Conclusion

In this article, we proposed an energy-efficient power allocation scheme that utilized the spectrum sensing information. The power allocation problem was formulated for maximizing the EE in the CR network and transformed into a nonfractional problem through Dinkelbach's method. The optimal solution of the optimization problem is obtained by solving the problem in an iterative fashion. Moreover, an expression for optimal solution was provided. The



(a)



(b)

Fig. 5. Impact of P_{D} on the average EE and SR performance ($P_{\max} = 10$ mW, $N = 5$, and $M = 256$). (a) Average EE versus sensing time α. (b) Average SR versus sensing time α

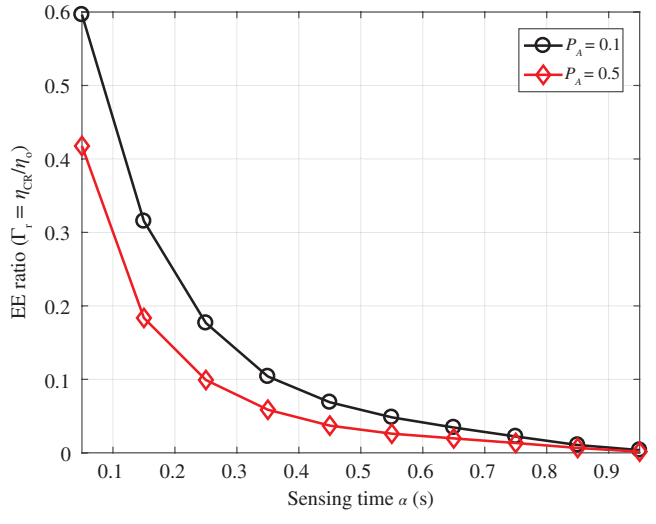


Fig. 6. EE ratio, $\Gamma_{\mathrm{r}} = \frac{\eta_{\mathrm{CR}}}{\eta_{\mathrm{o}}}$, versus sensing time α with $P_{\max} = 10$ mW, $N = 5$, and $M = 256$

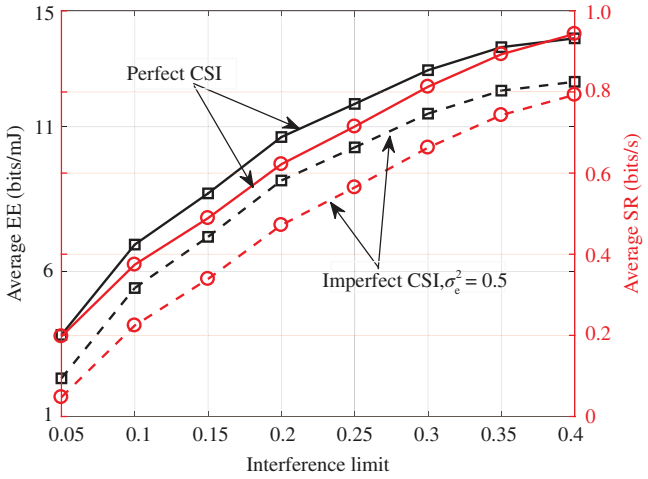


Fig. 7. Average EE and SR versus interference limit I_{o}

proposed system model is also extended to imperfect CSI scenario. Simulation results confirmed the convergence of the proposed scheme. The impact of P_{A} and P_{D} on the average EE and SE performance of the considered CR network is demonstrated. Additionally, we analyzed the effect of α on the network's achievable SR and EE. For lower value of α, the average EE and SR performances were significantly improved. The design framework can help in understanding for successful development of forthcoming energy-efficient CR networks.

References

(1) Mitola J, Maguire GQ. Cognitive radio: Making software radios more personal. *IEEE Personal Communications* 1999; **6**:13–18.

(2) Haykin S. Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications* 2005; **23**:201–220.

(3) Goldsmith A, Jafar SA, Maric I, Srinivasa S. Breaking spectrum gridlock with cognitive radios: An information theoretic perspective. *Proceedings of the IEEE* 2009; **97**:894–914.

(4) Lu L, Zhou X, Onunkwo U, Li GY. Ten years of research in spectrum sensing and sharing in cognitive radio. *EURASIP Journal on Wireless Communications and Networking* 2012; **12**:1–6.

(5) Tragos E, Zeadally S, Fragkiadakis AG, Siris VA. Spectrum assignment in cognitive radio: A comprehensive survey. *IEEE Communications Surveys and Tutorials* 2013; **15**:1108–1135.

(6) Su H, Zhang X. CREAM-MAC: An efficient cognitive radio enabled multi-channel MAC protocol for wireless networks. *Proceedings of the IEEE WoWMoM*, Newport Beach, CA, 2008; 1–8.

(7) Kim H, Shin KG. In-band spectrum sensing in cognitive radio networks: Energy detection or feature detection? *Proceedings of ACM MobiCom*, New York, 2008; 14–25.

(8) Anand A, Manikopoulos C, Jones Q, Borcea C. A quantitative analysis of power consumption for location-aware applications on phones. *Proceedings of IEEE ISIE*, Vigo, Spain, 2007; 1986–1991.

(9) Akyildiz IF, Lo BF, Balakrishan R. Cooperative spectrum sensing in cognitive radio networks: A survey. *Physical Communication* 2011; **14**:40–62.

(10) Liang YC, Zeng Y, Peh E, Hoang AT. Sensing-throughput tradeoff for cognitive radio networks. *IEEE Transactions on Wireless Communications* 2008; **7**:1326–1337.

(11) Su H., Zhang X. Energy-efficient spectrum sensing for cognitive radio networks. *Proceedings of IEEE ICC*, Cape Town, South Africa, 2010; 1–5.

(12) Tang L, Chen Y, Hines E, Alouini M. Effect of primary user traffic on sensing-throughput tradeoff for cognitive radios. *IEEE Transactions on Wireless Communications* 2011; **10**:1063–1068.

(13) Moghimi F, Mallik R, Schober R. Sensing time and power optimization in MIMO cognitive radio networks. *IEEE Transactions on Wireless Communications* 2012; **11**:3398–3408.

(14) Guerrini M, Rugini L, Banelli P. Sensing-throughput tradeoff for cognitive radios. *Proceedings of IEEE SPAWC*, Darmstadt, Germany, 2013; 115–119.
(15) Feng D, Jiang C, Lim G, Cimini LJ, Feng G, Li GY. A survey of energy-efficient wireless communications. *IEEE Communications Surveys & Tutorials* 2013; **15**:167–178.
(16) Zhao C, Kwak K. Joint sensing time and power allocation in cooperatively cognitive networks. *IEEE Communications Letters* 2010; **14**:163–165.
(17) Wang Y, Ren P, Du Q. Statistical QoS driven power allocation for cognitive networks under primary user's outage probability constraint. *Proceedings of IEEE PIMRC*, Sydney, Australia, 2012; 167–172.
(18) Li L, Zhou X, Xu H, Li GY, Wang D, Soong A. Energy-efficient transmission in cognitive radio networks. *Proceedings of IEEE CCNC*, Las Vegas, NV, 2010; 1–5.
(19) Stabellini L, Zander J. Energy-aware spectrum sensing in cognitive wireless sensor networks: a cross layer approach. *Proceedings of IEEE WCNC*, Sydney, Australia, 2010; 1–6.
(20) Wu Y, Tsang DHK. Energy-efficient spectrum sensing and transmission for cognitive radio system. *IEEE Communications Letters* 2011; **15**:545–547.
(21) Pei Y, Liang YC, Teh KC, Li KH. Energy-efficient design of sequential channel sensing in cognitive radio networks: Optimal sensing strategy, power allocation, and sensing order. *IEEE Journal on Selected Areas in Communications* 2011; **29**:1648–1659.
(22) Wu Y, Lau VKN, Tsang DHK, Qian LP. Energy-efficient delay-constrained transmission and sensing for cognitive radio systems. *IEEE Transactions on Vehicular Technology* 2012; **61**: 3100–3113.
(23) Huang S, Chen H, Zhang Y, Zhao F. Energy-efficient cooperative spectrum sensing with amplify-and-forward relaying. *IEEE Communications Letters* 2012; **16**:450–453.
(24) Shi Z, Teh KC, Li KH. Joint design of sensing and transmission in energy-efficient cognitive radio systems over fading channels. *IET Communications* 2013; **7**:577–584.
(25) Yadav R, Singh K, Kumar A, Kumar T. Energy-efficient power allocation with spectrum sensing in cognitive radio networks. *Proceedings of IEEE ICCC*, Shenzhen, China, 2015; 1–6.
(26) Tse D, Viswanath P. *Fundamentals of Wireless Communication*. Cambridge University Press: Cambridge, UK; 2005.
(27) Singh K, Gupta A, Ratnarajah T. A utility-based joint subcarrier and power allocation for green communications in multi-user two-way regenerative relay networks. *IEEE Transactions on Communications* 2017; **65**(9):3705–3722.
(28) Dinkelbach W. On nonlinear fractional programming. *Management Science* 1967; **13**:492–498.
(29) Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge University Press: Cambridge, England; 2004.
(30) Singh K, Gupta A, Ratnarajah T, Ku M. A general approach toward green resource allocation in relay-assisted multiuser communication networks. *IEEE Transactions on Wireless Communications* 2018; **17**(2):848–862.
(31) Jang YU, Jeong ER, Lee YH. A two-step approach to power allocation for OFDM signals over two-way amplify-and-forward relay. *IEEE Transactions on Signal Processing* 2010; **58**: 2426–2430.
(32) Masmoudi R, Belmega EV, Fijalkow I. Impact of imperfect CSI on resource allocation in cognitive radio channels. *Proceedings of IEEE WiMo*, Rome, Italy, 2017; 293–299.
(33) Kwon T, Wong VWS, Schober R. Secure MISO cognitive radio system with perfect and imperfect CSI. *Proceedings of IEEE GLOBECOM*, Anaheim, CA, 2012; 1236–1241.
(34) Forouzan N, Ghorashi SA. New algorithm for joint subchannel and power allocation in multi-cell OFDMA-based cognitive radio networks. *IET Communications* 2014; **8**:508–515.
(35) Wang X, Ekin S, Serpedin E. Joint spectrum sensing and resource allocation in multi-band-multi-user cognitive radio networks. *IEEE Transactions on Communications* 2018; **99**:1.
(36) Kako S, Takyu O, Fuji T. Joint resource allocation with interference constraint for cognitive multi-hop wireless networks. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 2015; **E98.A**:569–577.
(37) Yang Y, Scutari G, Song P, Palomar DP. Robust MIMO cognitive radio systems under interference temperature constraints. *IEEE Journal on Selected Areas in Communications* 2013; **31**:2465–2482.

Ramnaresh Yadav (Non-member) received his B.Tech. degree from the Kurukshetra University, Kurukshetra, India, his M.Tech. degree from the National Institute of Technology, Kurukshetra, India and currently he is pursuing his Ph.D. degree at Indira Gandhi Delhi Technical University for Women, Delhi, India. His main research interests are green communication, cognitive radio networks, and full duplex wireless communication.

Ashwani Kumar (Non-member) received his B.E. and M.E. degrees in Electronics and communication Engineering from the Delhi College of Engineering, Delhi University, India in 1991 and 1993, respectively. He received his Ph.D. degree in Digital Signal Processing from the Delhi college of Engineering, Delhi University, India in 2002. Since 2008, he has been with the Indira Gandhi Delhi Technical University for Women, Delhi, India, where he is currently working as a Professor. His current research interest includes signal processing in wireless communication, energy-efficient wireless communication, and full duplex wireless communication.

Keshav Singh (Non-member) received his degree of Master of Technology (with first-class honors) in Computer Science from the Devi Ahilya Vishwavidyalaya, Indore, India, in 2006, his M.Sc. in Information & Telecommunications Technologies from Athens Information Technology, Greece, in 2009, and his Ph.D. degree in Communication ngineering from the National Central University, Taiwan, in 2015. Since 2016, he has been with Institute for Digital Communications, School of Engineering, the University of Edinburgh, where he is currently working as a Research Associate. He is a member of IEEE. He has also served as a Technical Program Committee Member for numerous IEEE conferences. His current research interests are in the areas of green communications, resource allocation, full-duplex radio, cooperative and energy harvesting networks, multiple-input multiple-output (MIMO) systems, optimization of radio access, caching in wireless communications, and NOMA.

PLOS ONE

RESEARCH ARTICLE

Health and economic growth: Evidence from dynamic panel data of 143 years

Rajesh Sharma*

University School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

* rajesh.sharma@dtu.ac.in

Check for updates

OPEN ACCESS

Citation: Sharma R (2018) Health and economic growth: Evidence from dynamic panel data of 143 years. PLoS ONE 13(10): e0204940. https://doi.org/10.1371/journal.pone.0204940

Editor: Wu Zeng, Brandeis University, UNITED STATES

Received: May 22, 2018

Accepted: September 16, 2018

Published: October 17, 2018

Copyright: © 2018 Rajesh Sharma. 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.

Data Availability Statement: Data are available at: Macroeconomic Data: http://www.macrohistory.net/data/; Data corresponding to Fig 1: http://databank.worldbank.org/data/home.aspx; Life Expectancy at Birth: http://www.mortality.org/; Data Corresponding to Fig 2: http://apps.who.int/gho/data/.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

This paper re-examines health-growth relationship using an unbalanced panel of 17 advanced economies for the period 1870–2013 and employs panel generalised method of moments estimator that takes care of endogeneity issues, which arise due to reverse causality. We utilise macroeconomic data corresponding to inflation, government expenditure, trade and schooling in sample countries that takes care of omitted variable bias in growth regression. With alternate model specifications, we show that population health proxied by life expectancy exert a positive and significant effect on both real income per capita as well as growth. Our results are in conformity with the existing empirical evidence on the relationship between health and economic growth, they, however, are more robust due to the presence of long-term data, appropriate econometric procedure and alternate model specifications. We also show a strong role of endogeneity in driving standard results in growth empirics. In addition to life expectancy, other constituent of human capital, education proxied by schooling is also positively associated with real per capita income. Policy implication that follows from this paper is that per capita income can be boosted through focussed policy attention on population health. The results, however, posit differing policy implications for advanced and developing economies.

Introduction

Researchers and policymakers strive hard to identify factors that influence economic growth to aid policymaking and implementation. Human capital made inroads into growth framework following endogenous growth revolution and is identified as one of the most important contributors to economic growth [1–3]. Recently, there are more evidences to the positive effect of health and healthcare investments in economic growth [4–5] and the importance of human capital for economic growth is re-emphasised by the World Bank [6–7]. Health and education are the most important constituents of human capital. Focussing on health, Fig 1A and 1B illustrates that rich countries are healthier when compared to poor countries, which gives rise to a question: whether rich countries are healthier because they are rich, or they are rich because they are healthier?

This question is analysed using a variety of population health indicators such as life expectancy [8], adult survival rate [9], child mortality [10] and adult height [11]. Theoretically, a

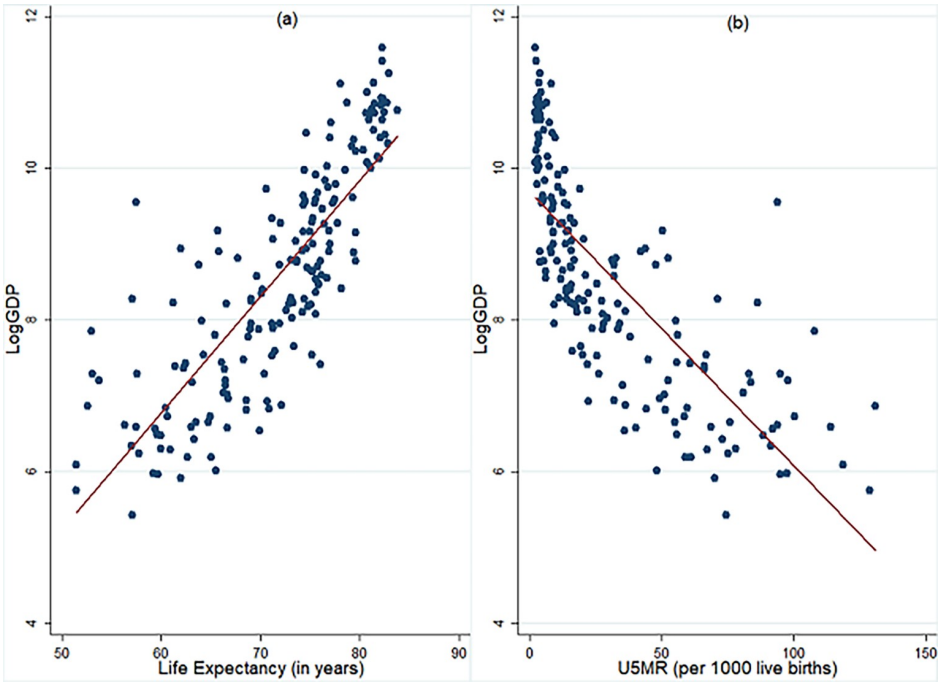


Fig 1. Relation between income per capita and indicators of population health. a) LogGDP vs Life Expectancy b) LogGDP vs U5MR. The data pertains to 193 countries for 2015 and is procured from World Development Indicators (WDI) database of World Bank. LE: Life Expectancy at birth (in years). U5MR: Under-five mortality rate is the probability per 1,000 that a new-born child will die before reaching age five, if subject to age-specific mortality rates of the specified year. LogGDP: Natural Logarithm of real GDP per capita at constant 2010 USD.

https://doi.org/10.1371/journal.pone.0204940.g001

healthy workforce contributes to greater economic output due to increased longevity as well as due to lesser number of working days lost due to ill-health. Moreover, a healthy workforce can produce more ideas and thereby contributes towards improved productivity of existing capital [1]. However, increased longevity may depress per capita income due to increased population and thin spreading of existing resources over large population [12]. Moreover, improved health may lead to substitution in favour of leisure as the same amount of income can be earned through lesser number of days of work. Which channel works, or which force dominates? Is increased longevity sufficient to compensate for its depressing effect or is there a plateau beyond which the effect of improved health starts diminishing, or it continues to provide productivity gains over infinite horizon as predicted in AK type endogenous growth models? In this paper, we aim to provide answers to these questions by re-examining health-growth relationship using a data-set of 17 advanced economies for the period 1870–2013. While previous studies were limited by data availability as they primarily relied on World Health Organisation (WHO) data-set for population health and Penn World Table (PWT) dataset for economic indicators, this paper draws strength from availability of 143 years of macroeconomic data that allows us to examine long-term relationship between economic growth and population health proxied by life expectancy. Working on a larger canvas of data is expected to unveil better insights into the long-term relationship between population health and economic growth.

The data corresponding to per capita income is made available for long time series by the Maddison project and the data pertaining to life expectancy is made available by Human Mortality Database [13] that provides data from 18th century for a set of countries. However, only a bivariate analysis between life expectancy and economic growth would have resulted in biased

and inconsistent estimates owing to omitted variable bias. We overcome the omitted variable bias in this paper by using historical macroeconomic data from Jorda, Schularik and Taylor ([14], JST from hereon) database which provides data for 17 advanced economies from 1870. The JST database contains data for consumer price index (CPI), merchandise trade, government expenditure and investment to GDP ratio, all of these factors have been observed to be robust determinants of economic growth.

In addition to resolving omitted variable concerns, we also take care of endogeneity issue in our estimation framework. This is because, growth determinants such as life expectancy or inflation may be affected by per capita income, thereby resulting in biased estimates due to reverse causality concerns. Concerning endogeneity in growth regression Caselli et al. [15] writes "*At a more abstract level, we wonder whether the very notion of exogenous variables is at all useful in a growth framework*." To control for endogeneity issues arising out of reverse causality from income to life expectancy, we apply panel generalised method of moments (GMM) technique [16–18]. The advantage with panel GMM methodologies is that we do not require separate instrument variable for each endogenous variable, rather, we can use internal instruments, which are made available by the panel nature of the dataset. In this paper, we also make a distinction between levels *versus* growth effect. Generally, researchers consider either real per capita income or its growth rate as the dependent variable to identify whether a given macroeconomic, demographic or any other factor influences levels of per capita income or its growth rate. In this paper, we consider both variables as dependent variables separately, and identify whether population health proxied by life expectancy has significant association with the levels of real per capita income and growth rate.

The main contribution of the paper is to assemble data on life expectancy and per capita income for the period beginning 1870. Moreover, in order to abate omitted variable concerns in growth regression, we could also gather data of macroeconomic variables, which are found to be significant in growth regression. The presence of data-set for such a long period allows us to examine long-run relationship between health and growth which was earlier not possible due to data-set dating back to 1960, only. Our data-set covers a period of time in which countries experienced changes in both economic and demographic profile which provides us with sufficient dynamics to study health-growth relationship as well as to provide few policy implications. Second, we also employ one of the most appropriate econometric procedures that takes care of possible endogeneity concerns while examining health-growth relationship.

Rest of this paper is structured as follows. The section titled "Materials and Methods" is devoted to data and econometric methodology employed in the paper. The section titled "Results and Discussion" provides empirical results and ensuing discussion of results along with the limitations of the paper. The section titled "Concluding Remarks" sums up the paper and provides few avenues of future research.

Materials and methods

Data and variables

We employ an unbalanced panel of 17 advanced economies for the period 1870–2013. The sample countries are: Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and United States (S1 Table provides key data for sample countries). The choice of sample countries is purely governed by data availability as long time series data exceeding a century of modern macroeconomic history is available only in these set of countries.

The level of economic development is represented by GDP per capita at purchasing power parity, expressed in natural logarithmic terms, as is commonly used in growth literature. As

control variables, we include investment to GDP ratio (INVEST), average years of total schooling (SCHOOLING), total merchandise trade (exports plus imports) to GDP ratio (OPEN), INFLATION and government expenditure to GDP ratio (GOVT_EXP). The choice of control variables is dictated by growth theory in which these variables along with measures of human capital are found to be the most robust determinants of economic growth (see Table 1 for variable definitions and sources of data). All the data is averaged over 10 years to examine the long-term effects of population health on income and to take care of short-term fluctuations [15]. In growth literature, it is a general practice to work with data averaged over five years to arrive at long-term effects of growth determinants [15, 19, 20]. However, the presence of long time series data in this paper allows us to work with ten-year averaged data, which is expected to provide more robust evidence of long-term relationship between population health and economic growth.

Data is averaged over 1871–1880, 1881–1890 and so on with last observation averaged over the period 2001–2013 providing us with a maximum of 14 observations per country. As schooling data is available at five-year frequency, we used data of median year as representative of the decade. For instance, data of schooling in 1875 serves as one observation in the decadal average of 1871–1880. Natural logarithm of real income per capita in 1870 serves as *initial income* for the decade 1871–1880 and so on with natural logarithm of real income per capita in 2000 serving as initial income for the period 2001–2013 (see Table 2 for descriptive statistics).

Life expectancy as indicator of population health. We use life expectancy at birth as a proxy of population health. Although health is a multi-dimensional concept and life expectancy is one of the most widely used indicator of population health [8, 12, 25, 26], it is not without limitations. For instance, a country may have high life expectancy but majority of its population might be suffering from illness and might not be productive. For instance, in poor economies, due to poor nutritional availability during childhood, a child will be lesser productive in performing tasks in future life expectancy may not be high because of available medical care [9]. A better measure may be health adjusted life expectancy (HALE) that calculates healthy years of life that a child at birth is expected to live [27]. However, data pertaining to HALE or other indicators of population health such as adult mortality rate (AMR) or child mortality (U5MR) is available only for shorter duration.

Table 1. Variable description and sources of data.

Variable	Variable Description	Data Source
LogGDP	Natural Logarithm of real per capita income (at PPP prices)	JST Database· (Schularik and Taylor [21] and Jorda, Schularik and Taylor [14, 22])
GROWTH	Ln (GDP_t)- Ln (GDP_{t-1})	JST Database
INFLATION	Ln (CPI_t)- Ln (CPI_{t-1})	JST Database
INVEST	Investment to GDP Ratio	JST Database
GOVT_EXP	Government Expenditure to GDP ratio	JST Database
OPEN	Ratio of [Exports +Imports] to GDP	JST Database
LIFE EXPECTANCY	Life Expectancy at birth (in years)	HMD Database and WDI Database*
SCHOOLING	Mean years of total schooling in adult population (15–64 years old)	Lee and Lee [23, 24]

* Data for Germany is procured from World Development Indicators which is available from 1950 whereas data from HMD was available only from 1990.

https://doi.org/10.1371/journal.pone.0204940.t001

Table 2. Descriptive statistics.

Variable	No. of Obs.	Mean	Std. Dev	Min.	Max.
GROWTH	238	0.019	0.017	-0.045	0.087
LogGDP	238	8.613	0.896	6.669	10.333
INFLATION	238	0.049	0.175	-0.062	2.573
INVEST	225	0.184	0.060	0.052	0.356
GOVT_EXP	233	0.318	0.190	0.019	0.849
OPEN	237	0.427	0.349	0.060	2.855
LIFE EXPECTANCY	194	64.650	12.372	32.666	82.515
SCHOOLING	238	5.983	3.273	0.150	13.010

Based on ten-year averaged data. GROWTH: growth rate of real GDP per capita (at PPP prices); LogGDP: natural logarithm of real GDP per capita at PPP prices; INFLATION: calculated as difference in natural logarithm of CPI; INVEST: investment to GDP ratio; GOVT_EXP: government expenditure to GDP ratio; OPEN: total merchandise trade to GDP ratio; LIFE EXPECTANCY: life expectancy at birth; SCHOOLING: average number of total years of schooling.

https://doi.org/10.1371/journal.pone.0204940.t002

We employ life expectancy at birth as the indicator of population health for two reasons. **First**, the data pertaining to life expectancy extends back to 19th century which enables a better empirical examination between population health and economic growth. **Second**, life expectancy has a strong, albeit not perfect, correlation with most other indicators of population health. For instance, the correlation of life expectancy at birth with healthy life expectancy at birth (HALE), adult mortality rate (AMR) and child mortality (U5MR) is 0.99, -0.96 and 0.93, respectively (See Fig 2A–2C). Therefore, life expectancy may be regarded as a good, although

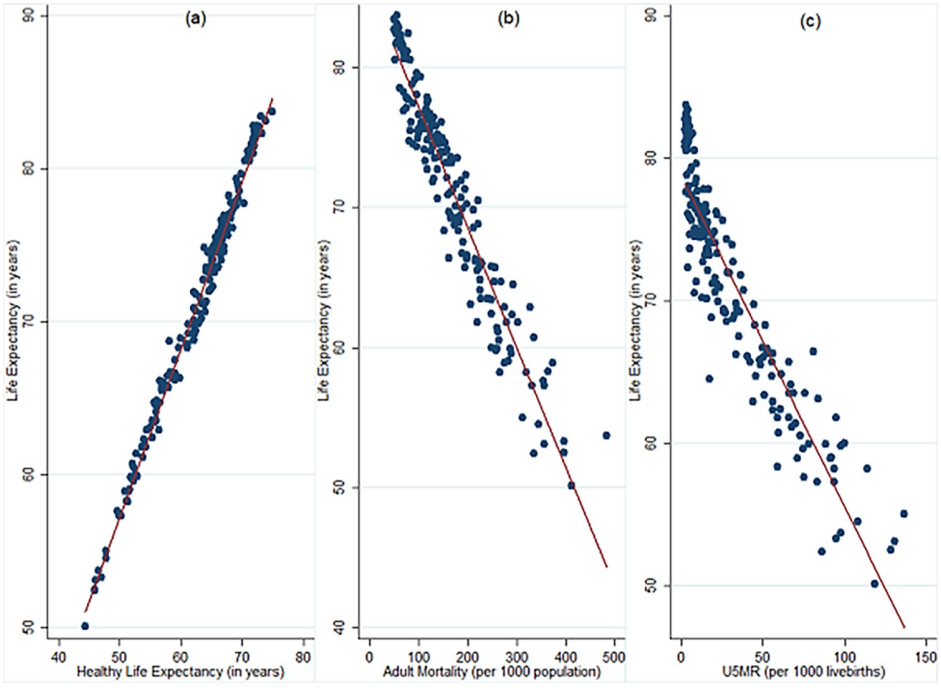


Fig 2. Relationship between life expectancy and other indicators of population health. a) LE vs HALE b) LE vs AMR c) LE vs U5MR. LE: Life Expectancy at birth (in years), HALE: Healthy Life Expectancy (at birth), AMR: Adult Mortality Rate (per 1000 population) and U5MR: Under-5 mortality Rate (per 1000 livebirths). The data pertains to 183 countries in 2015 and is procured from WHO mortality and global health estimates database.

https://doi.org/10.1371/journal.pone.0204940.g002

not perfect, representative indicator of population health (See S2 Table for data used in the figure).

Econometric methodology

Our basic regression framework builds upon Barro-type [28] growth regression model as specified below:

$$y_i = \alpha + \gamma X + \varepsilon_i \quad (1)$$

Here, i = 1,2,3. . .. . .n indexes cross-sections, y_i is the per capita income of i^{th} cross-section, X is the vector of growth determinants and ε_i is the stochastic error term. Above regression specified in cross-sectional framework suffers from two main limitations. *First*, it doesn't take care of individual heterogeneity and the assumption of identical production function may result in biased coefficient estimates of γ [15]. This happens because of ignoring country-specific fixed effects in above regression, which may be correlated with regressors, therefore, estimation of cross-sectional regression Eq (1) through ordinary least squares (OLS) leads to biased coefficient estimates [15]. *Second*, it doesn't exploit the time dimension of the data-set. Panel estimation which relaxes the restrictive assumption of identical production function can take care of both limitations. Growth regression in panel framework is specified as below:

$$Y_{i,t} = \mu_i + \gamma X_{i,t} + \varepsilon_{i,t} \quad (2)$$

Here i = 1,2,3. . . . . indexes country; t = 1,2,3. . .. indexes time; μ_i is the country-specific fixed effect that accounts for cross-sectional heterogeneity. In growth regression, income in previous period is also a significant determinant of income in the following period [15], therefore, it is more appropriate to specify growth regression in a dynamic panel framework as below.

$$Y_{i,t} = \mu_i + \beta Y_{i,t-1} + \gamma X_{i,t} + \varepsilon_{i,t} \quad (3)$$

However, estimation of Eq (3) using OLS will result in biased coefficient estimates due to correlation between fixed effects and lagged dependent variable [29]. Part of this issue can be resolved by differencing the data, which eliminates fixed effects.

$$Y_{i,t} - Y_{i,t-1} = \beta(Y_{i,t-1} - Y_{i,t-2}) + \gamma(X_{i,t} - X_{i,t-1}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1}) \quad (4)$$

Estimation of Eq (4) through OLS still leads to biased estimates due to correlation between $(Y_{i,t-1}-Y_{i,t-2})$ and $(\varepsilon_{i,t}-\varepsilon_{i,t-1})$. There are also endogeneity concerns due to potential reverse causality from income to growth determinants such as inflation, health and schooling. The estimation of Eq (4) using OLS fails to address the concerns imposed by endogeneity of explanatory variables. The endogeneity issue is resolved by using $Y_{i,t-2}$ as instrument for $(Y_{i,t-1}-Y_{i,t-2})$ which is correlated with $(Y_{i,t-1}-Y_{i,t-2})$ by construction but is uncorrelated with $(\varepsilon_{i,t}-\varepsilon_{i,t-1})$ provided that new error $(\varepsilon_{i,t}-\varepsilon_{i,t-1})$ term is serially uncorrelated (Arellano and Bond [16], Eq (4) also marks the first step in first difference GMM estimation). Therefore, following moment conditions may be exploited to arrive at the estimates of β and γ.

$$E[y_{it-s}(\varepsilon_{it} - \varepsilon_{it-1})] = 0 \text{ for } s \geq 2, t = 3, 4 \ldots\ldots\ldots T. \quad (5)$$

A similar instrumental variable strategy can be pursued if the explanatory variables (X's) are treated as endogenous providing us with following moment conditions.

$$E[x_{it-s}(\varepsilon_{it} - \varepsilon_{it-1})] = 0 \text{ for } s \geq 2, t = 3, 4 \ldots\ldots T. \quad (6)$$

However, moment conditions specified in Eqs (5) and (6) result in more moment conditions than the number of parameters to be estimated, thereby results in an over-identified system of equations. Such an over-identified system of equations is solved through generalised method of moments (GMM) estimation to arrive at coefficient estimates of β and γ [16].

Model estimation using panel GMM estimators gives valid estimates, provided they pass a battery of specifications tests (Arellano and Bond [16]). *First*, validity of lagged values instrumenting for the differenced variable depends crucially on their being uncorrelated with differenced error term ($\varepsilon_{it}-\varepsilon_{it-1}$) which can be tested using Sargan test of overidentifying restrictions [30]. In this paper, we rely on the results from Hansen test of over-identifying restrictions rather than Sargan test, as the later not robust to heteroscedasticity or autocorrelation [31]. Therefore, following Roodman [31], for one-step, robust estimation, we report Hansen J statistic, which is the minimized value of two-step GMM criterion function and is robust. *Second*, instrument validity depends upon the assumption of error term being serially uncorrelated [16]. We, however, require errors to be serially uncorrelated of order two only, as by construction, errors in the differenced Eq (4) are correlated of order one i.e. ($\varepsilon_{it}-\varepsilon_{it-1}$) is correlated with ($\varepsilon_{it-1}-\varepsilon_{it-2}$). Therefore, we just require errors to be uncorrelated of order 2 which is tested using Arellano and Bond test [16].

Results and discussion

Preliminary observations

Fig 3A and 3B presents bivariate relation between life expectancy and per capita income as well as its growth for sample countries. Life expectancy exhibits positive relation with both natural logarithm of per capita income (LogGDP) and its growth. The positive relation between life expectancy and per capita income seems to be more robust when compared to that between life expectancy and growth.

To gain a further visual traction, Fig 4 depicts movement of LogGDP and life expectancy for each individual country in the sample which again shows a positive relation between life expectancy and per capita income.

Pairwise correlation. Before investigating the empirical relation between health and economic growth, it is instructive to examine the pairwise correlation between explanatory variables to be used in the regression set-up which helps in checking whether the regression results are distorted by perfect multicollinearity in regression or not. Table 3 illustrates that we can rule out the evidence of perfect multicollinearity between explanatory variables except for a high correlation (0.87), albeit not perfect, between the two components of human capital i.e. life expectancy and schooling. We will take up this issue in panel GMM estimation.

OLS results

Before moving towards panel GMM estimation, we provide results corresponding to pooled OLS (Col I, Table 4) which helps in illuminating how the coefficient estimates change by inclusion of initial income and by accounting for endogeneity of the explanatory variables. Pooled OLS results (Col I, Table 4) demonstrate that both life expectancy and schooling enter the regression positively and are significant ($p<0.01$). This model explains 90% of the variation in the logarithm of real per capita income (LogGDP). The other control variables such as openness (OPEN) and government expenditure (GOVT_EXP) enter the regression positively, while being statistically significant at 1% level. As discussed previously, this regression produces biased estimates as cross-sectional heterogeneity is not being accounted for in this. However, when we add fixed effects in regression framework (Col II of Table 4), new model explains 95% of variation in real per capita income (LogGDP) and the standard errors get

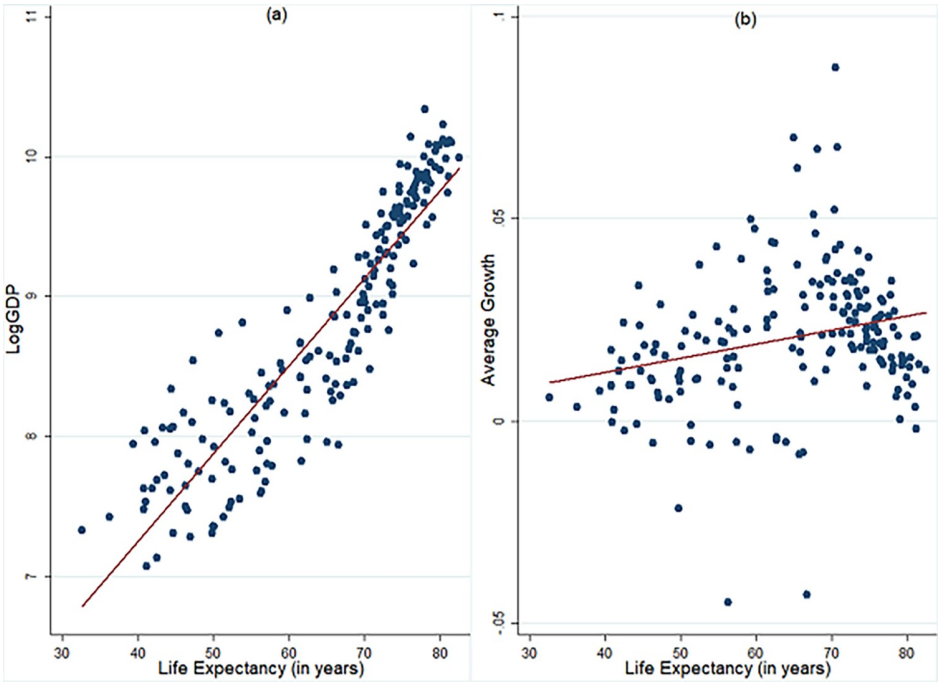


Fig 3. Relationship between average life expectancy and per capita income, growth. a) Scatter plot between LogGDP and Life Expectancy b) Scatter plot between Growth and Life Expectancy. LogGDP is the natural logarithm of real per capita income (at PPP prices). Growth: Growth rate of real per capita income calculated as difference of natural logarithm of real per capita income. Life Expectancy: Life Expectancy at Birth (in years). The data presented here is ten-year averaged.

https://doi.org/10.1371/journal.pone.0204940.g003

smaller in the case of all explanatory variables. The measures of human capital, life expectancy and schooling still exert positive and statistically significant influence on per capita income ($p<0.01$), the coefficient size on life expectancy, however, is slightly diminished. We also observe that inflation exerts significant and negative influence on per capita income and the investment to GDP ratio (INVEST) is positively associated with per capita income (LogGDP). In column III-IV of Table 4, we replaced the dependent variable to economic growth and regressed it over the same set of explanatory variables as in column I-II. Observe that coefficient sign and statistical significance remains unchanged in most of the variables except schooling which enters the pooled OLS and fixed effect regression negatively and government expenditure no longer remains statistically significant($p>0.1$). Even after adding fixed effects in growth regression, the model explains only 33% variation in growth of real per capita income as against 95% variation explained by these factors with per capita income as the dependent variable. One striking observation is that schooling exerts negative and statistically significant influence on growth of per capita income (Col III-IV, Table 4). The result seems to be at odds with the conventional wisdom of positive association between human capital and economic growth. However, such kind of anomalous observation had also been experienced previously in which human capital proxied by schooling was found to be either statistically insignificant or negative in growth regression [3, 19, 32]. As mentioned in Islam [19], the reason lies in the theoretical variable H (human capital) employed in the production function and the actual variable used in growth regression. As years of schooling may not fully account for the quality of education, an imperfect proxy for education may, therefore, manifests in the inappropriate coefficient sign or significance in growth regression.

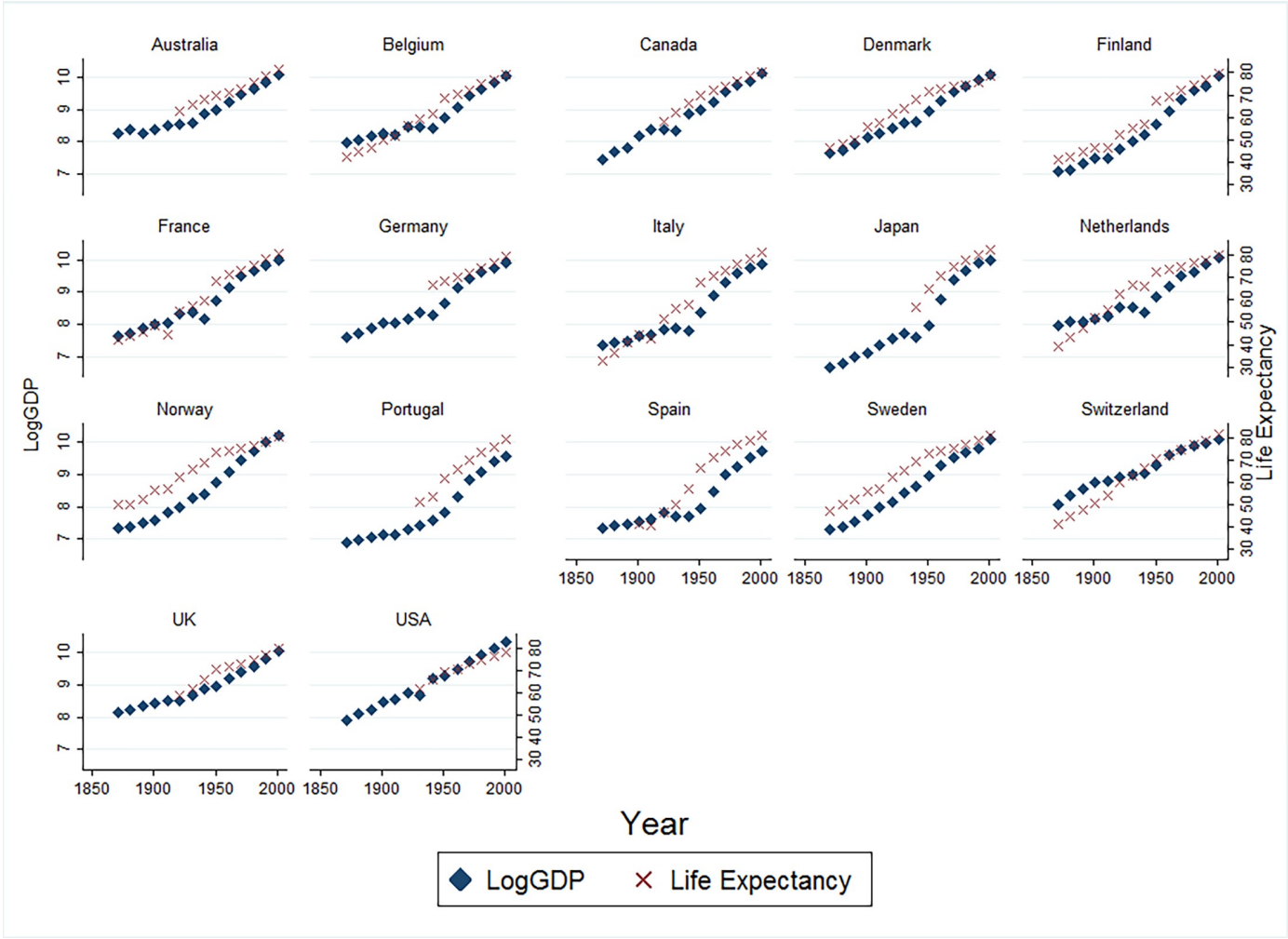


Fig 4. Movement of life expectancy and per capita income: By sample country. The Data pertains to 17 advanced economies in the sample. LogGDP: Natural logarithm of per capita GDP (at PPP prices).

https://doi.org/10.1371/journal.pone.0204940.g004

Table 3. Correlation matrix.

	GROWTH	LogGDP	INFLATION	INVEST	GOVT_EXP	OPEN	LIFE EXPECTANCY	SCHOOLING
GROWTH	1							
LogGDP	0.1778	1						
INFLATION	-0.0132	-0.016	1					
INVEST	0.3791	0.5985	0.0335	1				
GOVT_EXP	0.1370	0.6301	0.0460	0.4076	1			
OPEN	-0.0468	0.1821	-0.0402	0.0111	0.1486	1		
LIFE EXPECTANCY	0.2575	0.9062	0.0809	0.6875	0.6122	-0.0341	1	
SCHOOLING	0.1567	0.9133	0.0077	0.5753	0.5635	0.0899	0.8711	1

Based on 10-year averaged data. GROWTH: growth rate of real GDP per capita (at PPP prices); LogGDP: natural logarithm of real GDP per capita at PPP prices; INFLATION: calculated as difference in natural logarithm of CPI; INVEST: investment to GDP ratio; GOVT_EXP: government expenditure to GDP ratio; OPEN: total merchandise trade to GDP ratio; LIFE EXPECTANCY: life expectancy at birth; SCHOOLING: average number of total years of schooling.

https://doi.org/10.1371/journal.pone.0204940.t003

Table 4. OLS results.

	I	II	III	IV
	LogGDP as dependent variable		**Growth as dependent variable**	
INFLATION	-0.6315	-0.9329***	-0.0686**	-0.0787**
	(0.4623)	(0.2655)	(0.0279)	(0.0330)
INVEST	0.0297	1.2533***	0.1114***	0.1278***
	(0.5929)	(0.4293)	(0.0306)	(0.0372)
GOVT_EXP	0.4040***	0.5903***	0.0005	0.0126
	(0.1470)	(0.1967)	(0.0066)	(0.0102)
OPEN	0.1883***	0.2480***	-0.0040**	0.0010
	(0.0437)	(0.0662)	(0.0019)	(0.0038)
LIFE EXPECTANCY	0.0317***	0.0223***	0.0006***	0.0008***
	(0.0053)	(0.0037)	(0.0002)	(0.0003)
SCHOOLING	0.1261***	0.1505***	-0.0028***	-0.0048***
	(0.0140)	(0.0151)	(0.0007)	(0.0009)
Intercept	5.6947***	5.7208***	-0.0176**	-0.0262**
	(0.2100)	(0.1381)	(0.0076)	(0.0085)
Fixed Effects	No	Yes	No	Yes
No. of Obs.	185	185	Yes	Yes
R^2	0.90	0.95	0.27	0.33

Dependent variable is natural logarithm of real GDP per capita at PPP prices for column I-II and growth rate of real GDP per capita (at PPP prices) for column III-IV. The explanatory variables are: INFLATION: difference in natural logarithm of CPI; INVEST: investment to GDP ratio; GOVT_EXP: government expenditure to GDP ratio; LIFE EXPECTANCY: life expectancy at birth; OPEN: total merchandise trade to GDP ratio; SCHOOLING: average number of total years of schooling. Standard errors are heteroscedasticity corrected robust errors and are presented in parentheses.

*/**/*** denote statistical significance at 10/5/1 percent, respectively.

https://doi.org/10.1371/journal.pone.0204940.t004

Although not perfect, high correlation between schooling and life expectancy may be distorting the coefficient estimates, the results corresponding to OLS excluding schooling are presented in S3 Table, which again demonstrates that life expectancy still exerts positive and statistically significant influence on real per capita income as well as its growth.

Panel GMM results: Effect of health on per capita income

Results presented in Table 4 provide evidence of significant effect of population health on economic growth. The estimates, however, cast few doubts on appropriateness of model specification as rich countries are healthier than poorer ones i.e. there may be evidence of reverse causality from income to health. Moreover, some of the effects picked up by the explanatory variables may be the underlying ones of lagged income as economic growth may sustain due to growth momentum. Therefore, appropriate model of economic growth must account for the effects of lagged income as well as must take care of endogeneity issues that arise due to reverse causality concerns. Therefore, we estimate the Eq (3) using first difference GMM in which all the variables are considered endogenous. The model may suffer from instrument proliferation issue, therefore, in order to reduce the instrument count, we use only two lags as instruments and collapse the instruments as suggested by Roodman [31, 33].

Table 5 shows that lagged income is positively associated with current income which provides evidence of persistence in the dependent variable i.e. real income per capita (Col I-II). The effect of life expectancy on per capita income is statistically insignificant when it is treated as exogenous ($p>0.1$). However, the regression in which it is treated as endogenous, life expectancy exerts positive and statistically significant ($p<0.05$) effect on per capita income,

schooling, however, is negative although statistically insignificant ($p>0.1$). We observe that inflation exerts negative and statistically significant ($p<0.01$) effect on per capita income, while the effect of trade openness is positive although significant only at 10% level. The model diagnostic tests show that errors are serially uncorrelated of order 2 (p-value = 0.473). Hansen test has p-value of 0.440 which implies that we fail to reject null hypothesis of no overidentifying restrictions.

The model estimated using first difference GMM model and presented in column 1-II of Table 5 may suffer from weak instrumental bias problems [18, 34]. Therefore, we also estimate the growth regression using system GMM which provides lesser biased coefficient estimates [18, 34, 35]. The system GMM approach [18] combines in a system, a regression in differences with regression in levels with the aim that additional moment conditions would be generated increasing the efficiency of resulting estimators. The idea of the system GMM is to estimate level equation and first difference equation simultaneously, and suggests using lags as instrument for differenced equation and differenced variables as instrument for the level equation. Thus, additional moment conditions for second part of the system (level equation) are:

$$E[\Delta y_{it-i}(\mu_i + \varepsilon_{it})] = 0 \quad (7)$$

$$E[\Delta X_{it-i}(\mu_i + \varepsilon_{it})] = 0 \quad (8)$$

Table 5. Panel GMM results with LogGDP as dependent variable.

	First Difference GMM (LE Exogenous)	First Difference GMM (LE Endogenous)	System GMM (LE Exogenous)	System GMM (LE Endogenous)
INITIAL	0.8704***	0.7229***	0.8031***	0.7422***
	(0.0704)	(0.0829)	(0.0490)	(0.0588)
INFLATION	-0.8349***	-0.6772***	-0.7961***	-0.6773***
	(0.2229)	(0.1873)	(0.1919)	(0.1950)
INVEST	1.6645**	0.5937	1.4314***	1.0232**
	(0.7005)	(0.6453)	(0.4738)	(0.4878)
GOVT_EXP	0.2525	0.1392	0.1539	0.0325
	(0.2966)	(0.2099)	(0.1678)	(0.1615)
OPEN	0.1578**	0.2827*	0.0698*	0.0815**
	(0.0788)	(0.1678)	(0.0416)	(0.0368)
LIFE EXPECTANCY	0.0039	0.0195**	0.0014	0.0096**
	(0.0064)	(0.00997)	(0.0041)	(0.0040)
SCHOOLING	-0.0034	-0.0129	0.0311	0.0244*
	(0.0244)	(0.0338)	(0.0194)	(0.0146)
No. of Obs.	166	166	185	185
No. of Instruments	13	14	13	22
Hansen Test p-value	0.420	0.440	0.395	0.544
AR (2) test p-value	0.339	0.473	0.660	0.892

Dependent variable LogGDP: natural logarithm of real GDP per capita at PPP prices. The explanatory variables are: INITIAL: natural logarithm of initial real per capita income (at PPP prices) at period t-1; INFLATION: difference in natural logarithm of CPI; INVEST: investment to GDP ratio; GOVT_EXP: government expenditure to GDP ratio; LIFE EXPECTANCY: life expectancy at birth; OPEN: total merchandise trade to GDP ratio; SCHOOLING: average number of total years of schooling. Standard errors are heteroscedasticity corrected robust errors and presented in parentheses.

*/**/*** denote statistical significance at 10/5/1 percent respectively. Null hypothesis of AR (2) test: errors are serially uncorrelated at order 2. Null hypothesis of Hansen test: there are no overidentifying restrictions.

https://doi.org/10.1371/journal.pone.0204940.t005

In system GMM, additional moment conditions specified by Eqs (7) and (8) along with those presented in Eqs (5) and (6) generate consistent and relatively more efficient estimates compared to those obtained in the first difference GMM model.

We observe that in system GMM results (Col III-IV of Table 5), the standard errors are substantially reduced in most of the variables. With system GMM results too, we observe that life expectancy exerts positive effect on per capita income when it is treated as endogenous which is justifiable given the reverse causality concerns mentioned previously. In system GMM results in which the life expectancy is considered endogenous, schooling also exerts positive and significant influence on per capita income ($p<0.1$). The other explanatory variables such as inflation, investment to GDP ratio (INVEST), openness and government expenditure to GDP ratio (GOVT_EXP) have similar sign and statistical significance as observed before (Table 4 and Col I-II of Table 5). Model specification tests also demonstrate that moment conditions employed in the system GMM are appropriate–errors are uncorrelated of order 2 (AB test p-value >0.1) and there are no overidentifying restrictions (Hansen test p-value > 0.1).

Panel GMM results: Effect of health on economic growth

In addition to investigating the effect of health on per capita income, we also examine the effect of life expectancy on economic growth using panel GMM regression (Table 6). In regression Eq (3), we replace LogGDP with growth in per capita income as the dependent variable. We

Table 6. Panel GMM results with growth as dependent variable.

	First Difference GMM	System GMM Estimation
INITIAL	-0.0407***	-0.0296***
	(0.0156)	(0.0105)
INFLATION	-0.0841*	-0.0937***
	(0.0491)	(0.0334)
INVEST	-0.0327	0.1280*
	(0.0896)	(0.0739)
GOVT_EXP	0.0248	0.0057
	(0.0390)	(0.0168)
OPEN	0.0568*	0.0090
	(0.0337)	(0.0058)
LIFE EXPECTANCY	0.0049**	0.0018**
	(0.0020)	(0.0007)
SCHOOLING	-0.0095	-0.0001
	(0.0063)	(0.0026)
No. of Obs.	166	185
No. of Instruments	14	22
Hansen Test p-value	0.321	0.443
AR (2) test p-value	0.301	0.086

Dependent variable: GROWTH: growth rate of real GDP per capita (at PPP prices) calculated as difference in natural logarithm of real per capita income (at PPP prices). The explanatory variables are: INITIAL: natural logarithm of initial real per capita income (at PPP prices) at time t-1; INFLATION: is calculated as difference in natural logarithm of CPI; invest: investment to GDP ratio; GOVT_EXP: government expenditure to GDP ratio; LIFE EXPECTANCY: life expectancy at birth; OPEN: total merchandise trade to GDP ratio; SCHOOLING: average number of total years of schooling. Standard errors are heteroscedasticity corrected robust errors and are presented in parentheses.
*/**/*** denote statistical significance at 10/5/1 percent, respectively. Null hypothesis of AR (2) test: errors are serially uncorrelated at order 2. Null hypothesis of Hansen test: there are no overidentifying restrictions.

https://doi.org/10.1371/journal.pone.0204940.t006

Table 7. Panel GMM results excluding schooling.

	First Difference GMM	System GMM	First Difference GMM	System GMM
INITIAL	0.6194***	0.8274***	-0.0596***	-0.0292***
	(0.1299)	(0.0395)	(0.0197)	(0.0067)
INFLATION	-0.6256***	-0.7611***	-0.0914**	-0.0978**
	(0.2343)	(0.1716)	(0.0463)	(0.0390)
INVEST	0.2291	1.0962**	-0.0231	0.1397*
	(0.8185)	(0.4627)	(0.1097)	(0.0819)
GOVT_EXP	0.2133	0.0740	0.0276	0.0062
	(0.2401)	(0.1463)	(0.0365)	(0.0177)
OPEN	0.3376*	0.0856***	0.0485	0.0096**
	(0.1839)	(0.0254)	(0.0301)	(0.0043)
LIFE EXPECTANCY	0.0239**	0.0086**	0.0037**	0.0016**
	(0.0113)	(0.0039)	(0.0017)	(0.0008)
No. of Obs.	166	166	185	185
No. of Instruments	12	19	12	19
Hansen Test p-value	0.595	0.644	0.113	0.256
AR (2) test p-value	0.340	0.681	0.267	0.113

Dependent variable is natural logarithm of real GDP per capita at PPP prices for column I-II and growth rate of real GDP per capita (at PPP prices) for column III-IV. The explanatory variables are: INITIAL: natural logarithm of initial real per capita income (at PPP prices); INFLATION: is calculated as difference in natural logarithm of CPI, INVEST: investment to GDP ratio; GOVT_EXP: government expenditure to GDP ratio; LIFE EXPECTANCY: life expectancy at birth; OPEN: total merchandise trade to GDP ratio. Standard errors are heteroscedasticity corrected robust errors and presented in parentheses.

*/**/*** denote statistical significance at 10/5/1 percent, respectively. Null hypothesis of AR (2) test: errors are serially uncorrelated at order 2. Null hypothesis of Hansen test: there are no overidentifying restrictions.

https://doi.org/10.1371/journal.pone.0204940.t007

obtain few results similar to those obtained with real per capita income (LogGDP) as the dependent variable such as positive and significant effect of life expectancy ($p<0.05$) on economic growth in both first difference and system GMM regression. Initial income per capita (INITIAL) has negative effect on subsequent growth which is in accordance with convergence hypothesis (Caselli et al [15]). Inflation has statistically significant negative influence on economic growth with both GMM estimators and investment to GDP ratio is significant ($p<0.1$) only in system GMM results (Col II, Table 6). In both the models, the model diagnostic tests show that errors are serially uncorrelated of order 2 ($p>0.1$). We also fail to reject the null hypothesis of no overidentifying restriction as the Hansen-J statistic has p-value greater than 0.10.

Panel GMM results: Effect of health excluding schooling

From Table 3, we observed high correlation between life expectancy and schooling. Therefore, to extract more robust estimates of life expectancy on per capita income and economic growth, we re-ran regression (3) using first difference and system GMM regression excluding schooling as one of the explanatory variables (Table 7). Comparing results of regression including schooling (Tables 5 and 6) and with those after excluding it (Table 7), the coefficient on life expectancy is still significant and there is not much change in its coefficient estimate. For instance, in system GMM regression, the effect of life expectancy on per capita income changes only from 0.0096 (Col IV, Table 5) to 0.0086 (Col II, Table 7) and on growth it changes marginally from 0.0018 (Col II, Table 6) to 0.0016 (Col IV, Table 7).

Table 8. Health expenditure and health status in country groupings.

Country Groups	Domestic general government health expenditure (% of GDP)	Out-of-pocket expenditure (% of current health expenditure)	Immunization, DPT (% of children ages 12–23 months)	Life expectancy at birth, total (years)
Low Income	1.26	40.42	77.54	62.11
Lower middle income	1.30	57.28	80.48	67.70
Upper middle income	3.34	31.56	94.42	75.14
High income	7.82	13.50	95.76	80.50

Note: Data is from World Development Indicators: Database of World Bank

https://doi.org/10.1371/journal.pone.0204940.t008

Discussion

Through alternate model specifications and econometric procedures, we found that life expectancy has positive and statistically significant effect on per capita income as well its growth. Our regression results also indicate a strong role of endogeneity in driving standard results in growth empirics [15]. What are the policy implications going forward and what developing countries can learn from the experience of developed countries? Our results implicate a scope of boosting economic growth through focussed policy attention on population health by increased investments in healthcare systems. Secondly, as shown in Fig 1, there is very strong correlation between life expectancy and other indicators of population health which implies that till the time we get a better summary indicator of population health for which data is available for longer periods, a focus on boosting longevity seems to be an optimal strategy. From S2 Table, we notice that life expectancy ranges from the lows of 50.2 and 52.4 in Sierra Leone and Angola, respectively, to 83.4 and 83.7 in Switzerland and Japan, respectively in 2015. Such low levels of life expectancy imply high child mortality, lower adult survival rate and high prevalence of disease burden in working age-group which implies a significant loss to human capital which significantly impairs its contribution to economic production. Moreover, low levels of life expectancy lead to higher fertility and consequent population growth leading to further thinning of existing resources per capita. Therefore, developing countries have a scope to boost their income levels by focussing on population health. This raises a question: how to boost population health? What are the challenges to healthcare systems in advanced economies and low-resource economies? In terms of boosting longevity, reduction in child mortality has played a crucial role in the past [36] and is expected to provide further gains in future for low- and low-middle income economies. Moreover, many of the diseases that account for premature deaths as well as a cause of morbidity are preventable as well as treatable with available low-cost interventions [37]. For instance, 19.5 million children under the age 1 did not receive the three recommended doses of Diphtheria, Tetanus and Pertussis (DTP3) in 2016, and 20.8 million children below the age 1 didn't receive a single dose of measles-containing vaccine [38]. Lastly, if we compare general government expenditure on health (Table 8) it is found to be significantly lower in poor economies. Moreover, in the light of stagnating development assistance for health (DAH) to poor economies, there is going to be significant pressure on already constrained public resources in poor economies [39].

Table 8 also demonstrates that in the absence of insurance cover, a large chunk of health expenditure is incurred as out-of-pocket expenditures (OOP) which further pushes people below the poverty line (see Fig 5). Health expenditure is thus a significant drain on poor household's resources and significantly impairs their human capital, which has further negative bearing on economic growth. In the light of high OOP expenditure in low and lower-middle

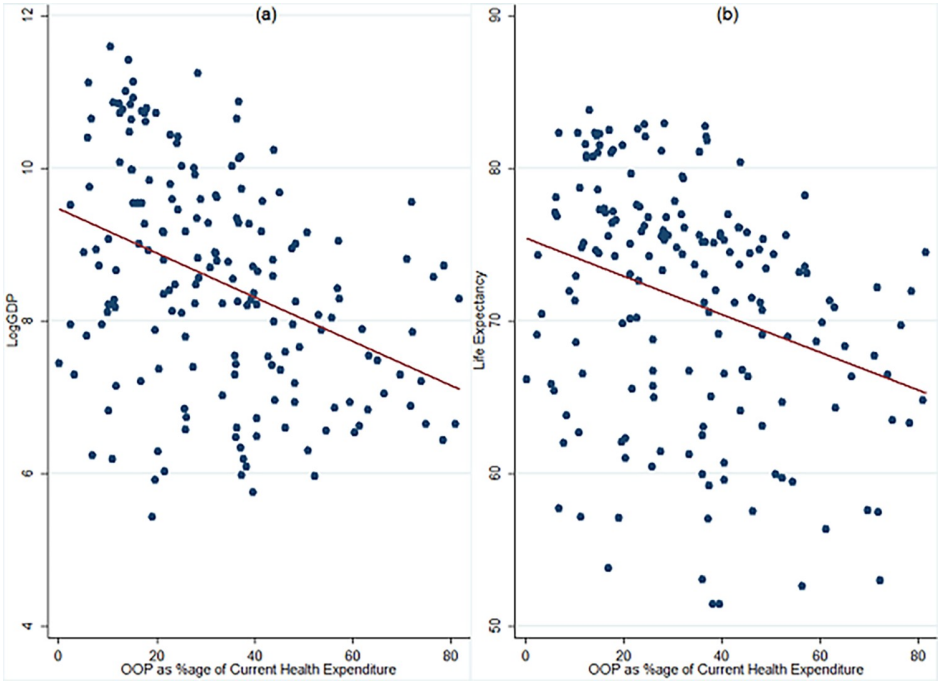


Fig 5. Relationship between income, life expectancy and OOP. a) Scatter plot between LogGDP and OOP b) Scatter plot between LE and OOP. The data pertains to 193 countries and is procured from World Development Indicators (WDI) database of World Bank. LogGDP: Natural Logarithm of real GDP per capita at constant 2010 USD. LE: Life Expectancy at Birth. OOP: Out-of-pocket expenditure as percentage of current health expenditure (CHE). Data of all variables pertains to 2015 and is procured from World Development Indicators (WDI) database of World Bank.

https://doi.org/10.1371/journal.pone.0204940.g005

income economies, the access to universal health coverage (UHC) is felt most urgently which is also emphasised in Sustainable Development Goal 3.8 [40].

Next question that demands urgent attention is: will life expectancy gains continue to persist and continue to provide economic gains forever? The answer to this question seems to be not clear as endogenous growth theory involving AK type of production function assume that investments in human capital continues to provide positive results forever by boosting productivity (A). However, economic rationale of diminishing marginal returns suggests that improvements in human capital may eventually lead to non-significant economic gains in future. Now, advanced economies which have witnessed substantial gains in life expectancy over last 140 years are standing on the verge of stagnating gains as most of health gains due to reduction in mortality from communicable causes have almost vanished and their place has increasingly been taken up by non-communicable diseases such as cancer, cardiovascular diseases and neurological disorders etc. [41]. The signs of non-significant gains in longevity may already be visible in developed countries (Fig 4). Secondly, there may be evidence of diminishing marginal returns to improved longevity evidenced by Global Burden of Disease study 2016 [42, 43], which shows that longer lives may not be translating into healthy longer lives. This is because we may have found solutions to many of leading communicable causes of death but the burden of non-communicable diseases such as cardiovascular diseases, diabetes, and neoplasms is increasing alongwith the burden of disorders such as neurological disorders, musculoskeletal disorders and substance abuse disorders, the later may not be claiming lives but are responsible for a surge in burden due to disability. This is particularly challenging as much of this burden falls on the working age group (15–64) causing a significant loss to human capital.

To summarise, the priority of public health policy in developing countries must be on boosting longevity, the advanced economies, on the other hand, must focus on ensuring healthier life to their long-living populations.

Limitations

The main limitation of this paper is that comprehensive macroeconomic data spanning over a century is available only for a subset of advanced economies. Although, the experience of advanced economies can shed some light on health-growth nexus and has policy implications for developing countries too, still, availability of data for developing countries for long time series is expected to illuminate health-growth relationship in a more robust manner. Second, although life expectancy is strongly correlated with other indicators of population health, still it is not a perfect measure of population health [9].

Concluding remarks

Sustainable Development Goal 3 seeks to *ensure healthy lives and promote well-being for all at all ages*. In this paper, we investigate whether population health has causal effect on per capita income and its growth or not, using an unbalanced panel of 17 advanced economies for the period 1870–2013. We use life expectancy at birth as a proxy for population health and control for endogeneity issues using panel generalised method of moments (GMM) technique [16–18]. With alternate model specifications, we demonstrated that population health has positive and significant effect on both real income per capita as well as its growth. In addition to life expectancy, other constituent of human capital, schooling is also positively associated with per capita income. Investment has positive effect on growth which is consistent with the findings of existing literature. Government expenditure, however, has either statistically insignificant ($p>0.1$) or positive effect on per capita income and growth which seems to be at odds with the hypothesis of crowding out effect of government expenditure on real investments and hence, must exerts negative effect on growth. Our results may be driven by the positive effect of government expenditure during great depression and recessionary times as the data-set employed in this study contains many of the recessionary episodes. This observation cannot be generalised, however, and the dynamics of government's size and its implications on real investments and growth demand rigorous empirical scrutiny by focussing on specific components of government expenditure. Inflation exerts negative and statistically significant influence on both per capita income and its growth and seems to be consistent with the existing literature [44]. Regarding population health, our results are consistent with previous results. Our results, however, seem to be more robust as these are based on 143 years of data, appropriate econometric specification and controlling for majority of macroeconomic determinants of growth. The probable channels through which life expectancy is envisaged to foster growth is through capital accumulation and boosting productivity of existing capital. In general, improved life expectancy leads to elongation of number of working years, also evident from worldwide debates on increasing the age of retirement. Second, majority of gains in life expectancy can be attributed to reductions in child as well as adult mortality; providing for greater number of working years and hence higher level of savings (and investments), which further propels the economic growth. This longevity-saving-growth channel has been identified as a source of East Asian growth miracle [45–46]. Health itself is not only a part of human capital but also contributes to human capita such as education [47]. Improved health (measured using life expectancy) allows the working population to be more productive in terms of new ideas and innovations. This further leads to enhanced productivity of the physical capital when employed by more skilful managers in the economy. Subsequently, longer living population invest more in human

capital such as education as it expects to live greater number of years [48] leading to a virtuous cycle of improved productivity and economic growth. The empirical investigation of channels through which population health affects per capita income and growth may be an important agenda for future research.

Lastly, life expectancy may have provided gains in per capita income in the past, however, due to elevating burden of NCDs, the disability adjusted life years (DALYs) burden–sum of years of life lost (YLL) due to premature death and years lived with disability (YLD) due to injury or illness–which was earlier used to be dominated by YLLs is now increasingly getting dominated by lived with disability (YLD) even in developing nations [49]. Moreover, the increasing burden of old age diseases ranging from cancers to neurological diseases as well as life style related diseases is exacting a toll on not only elderly population but also on working age-group [37]. How health system can tackle these challenges going forward? This is particularly a challenging problem for poor countries whose health system are stressed due to twin burden of communicable diseases as well as rising burden of non-communicable diseases such as cardiovascular disorders and cancers. Going forward, an analysis of economic burden or loss in GDP due to these diseases may be an agenda for future research.

Supporting information

S1 Table. Key data for sample countries. Data Source: Based on authors' calculation on data reported in the paper.
(DOCX)

S2 Table. Key indicators of population health for 183 countries. LE: Life Expectancy at Birth (in years); LE60: Life Expectancy at the age 60 (in years); HALE: Healthy Life Expectancy at Birth (in years); HALE60: Healthy Life Expectancy at the age 60 (in years); AMR: Adult Mortality Rate (per 100,000 population); U5MR: Under 5 Mortality Rate (per 1000 livebirths).
(DOCX)

S3 Table. Pooled OLS results excluding schooling as explanatory variable. Dependent variable is natural logarithm of real GDP per capita at PPP prices for column I-II and growth rate of real GDP per capita (at PPP prices) for column III-IV. The explanatory variables are: INFLATION: difference in natural logarithm of CPI; INVEST: investment to GDP ratio; GOVT_EXP: government expenditure to GDP ratio; LIFE EXPECTANCY: life expectancy at birth; OPEN: total merchandise trade to GDP ratio; SCHOOLING: average number of total years of schooling. Standard errors are heteroscedasticity corrected robust errors and are presented in parentheses. */**/*** denote statistical significance at 10/5/1 percent, respectively.
(DOCX)

S4 Table. Pooled OLS results excluding life expectancy as explanatory variable. Dependent variable is natural logarithm of real GDP per capita at PPP prices for column I-II and growth rate of real GDP per capita (at PPP prices) for column III-IV. The explanatory variables are: INFLATION: difference in natural logarithm of CPI; INVEST: investment to GDP ratio; GOVT_EXP: government expenditure to GDP ratio; OPEN: total merchandise trade to GDP ratio; SCHOOLING: average number of total years of schooling. Standard errors are heteroscedasticity corrected robust errors and are presented in parentheses. */**/*** denote statistical significance at 10/5/1 percent, respectively.
(DOCX)

S5 Table. Panel GMM results excluding schooling as explanatory variable. Dependent variable is natural logarithm of real GDP per capita at PPP prices for column I-II and growth rate

of real GDP per capita (at PPP prices) for column III-IV. The explanatory variables are: INITIAL: natural logarithm of initial real per capita income (at PPP prices); INFLATION: calculated as difference in natural logarithm of CPI; INVEST: investment to GDP ratio; GOVT_EXP: government expenditure to GDP ratio; LIFE EXPECTANCY: life expectancy at birth; OPEN: total merchandise trade to GDP ratio. Standard errors are heteroscedasticity corrected robust errors and are presented in parentheses. */**/*** denote statistical significance at 10/5/1 percent, respectively. Null hypothesis of AR (2) test: errors are serially uncorrelated of order 2. Null hypothesis of Hansen test: there are no overidentifying restrictions.
(DOCX)

Acknowledgments

I thank the editor and two anonymous referees for comments that greatly improved the paper. I also want to thank Òscar Jordà, Moritz Schularik and Alan M. Taylor for providing macroeconomic history database in public domain which was quite critical in this study.

Author Contributions

Conceptualization: Rajesh Sharma.

Data curation: Rajesh Sharma.

Formal analysis: Rajesh Sharma.

Funding acquisition: Rajesh Sharma.

Investigation: Rajesh Sharma.

Methodology: Rajesh Sharma.

Project administration: Rajesh Sharma.

Resources: Rajesh Sharma.

Software: Rajesh Sharma.

Supervision: Rajesh Sharma.

Validation: Rajesh Sharma.

Visualization: Rajesh Sharma.

Writing – original draft: Rajesh Sharma.

Writing – review & editing: Rajesh Sharma.

References

1. Romer PM. Human capital and growth: theory and evidence. In Carnegie-Rochester conference series on public policy 1990 Mar 1 (Vol. 32, pp. 251–286). North-Holland.
2. Mankiw NG, Romer D, Weil DN. A contribution to the empirics of economic growth. The quarterly journal of economics. 1992 May 1; 107(2):407–37.
3. Benhabib J, Spiegel MM. The role of human capital in economic development evidence from aggregate cross-country data. Journal of Monetary economics. 1994 Oct 1; 34(2):143–73.
4. Boachie MK. Health and Economic Growth in Ghana: An Empirical Investigation. Fudan Journal of the Humanities and Social Sciences. 2017 Jun 1; 10(2):253–65.
5. Rosendo Silva F, Simões M, Sousa Andrade J. Health investments and economic growth: a quantile regression approach. International Journal of Development Issues. 2018 Jun 1; 17(2): 220–245. https://doi.org/10.1108/IJDI-12-2017-0200

6. The World Bank. Countries commit to strong action on human capital to drive economic growth. World Bank. http://www.worldbank.org/en/news/feature/2017/10/20/countries-commit-to-strong-action-on-human-capital-to-drive-economic-growth (accessed May 7, 2018).
7. The World Bank. Human Capital: The Greatest Asset of Economies on the Rise. World Bank. http://www.worldbank.org/en/news/opinion/2017/04/03/human-capital-the-greatest-asset-of-economies-on-the-rise (accessed May 7, 2018).
8. Bloom DE, Canning D, Sevilla J. The effect of health on economic growth: a production function approach. World development. 2004 Jan 1; 32(1):1–3.
9. Bhargava A, Jamison DT, Lau LJ, Murray CJ. Modelling the effects of health on economic growth. Journal of health economics. 2001 May 1; 20(3):423–40. PMID: 11373839
10. Pritchett L, Summers LH. Wealthier is healthier. Journal of Human resources. 1996 Oct 1:841–68.
11. Subramanian SV, Özaltin E, Finlay JE. Height of nations: a socioeconomic analysis of cohort differences and patterns among women in 54 low-to middle-income countries. PLoS One. 2011 Apr 20; 6(4): e18962. https://doi.org/10.1371/journal.pone.0018962 PMID: 21533104
12. Acemoglu D, Johnson S. Disease and development: the effect of life expectancy on economic growth. Journal of political Economy. 2007 Dec; 115(6):925–85.
13. Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de. (Accessed: August 2017)
14. Jordà Ò, Schularick M, Taylor AM. Macrofinancial history and the new business cycle facts. NBER Macroeconomics Annual. 2017 Jan 1; 31(1):213–63.
15. Caselli F, Esquivel G, Lefort F. Reopening the convergence debate: a new look at cross-country growth empirics. Journal of economic growth. 1996 Sep 1; 1(3):363–89.
16. Arellano M, Bond S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. The review of economic studies. 1991 Apr 1; 58(2):277–97.
17. Arellano M, Bover O. Another look at the instrumental variable estimation of error-components models. Journal of econometrics. 1995 Jul 1; 68(1):29–51.
18. Blundell R, Bond S. Initial conditions and moment restrictions in dynamic panel data models. Journal of econometrics. 1998 Nov 1; 87(1):115–43.
19. Islam N. Growth empirics: a panel data approach. The Quarterly Journal of Economics. 1995 Nov 1; 110(4):1127–70.
20. Levine R, Loayza N & Beck T. Financial Intermediation and growth: causality and causes. Journal of Monetary Economics. (2000) Aug 1; 46(1):31–77.
21. Schularick M, Taylor AM. Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870–2008. American Economic Review. 2012 Apr; 102(2):1029–61.
22. Jordà Ò, Schularick MH, Taylor AM. When credit bites back: leverage, business cycles, and crises. National Bureau of Economic Research; 2011 Nov 23.
23. Lee JW, Lee H. Human capital in the long run. Journal of Development Economics. 2016 Sep 1; 122:147–69.
24. Barro Robert J. and Lee Jong-Wha, Education Matters: Global Schooling Gains from the 19th to the 21st Century (Oxford University Press, 2015)
25. Castelló-Climent A, Doménech R. Human capital inequality, life expectancy and economic growth. The Economic Journal. 2008 Apr 1; 118(528):653–77.
26. Cervellati M, Sunde U. Life expectancy and economic growth: the role of the demographic transition. Journal of economic growth. 2011 Jun 1; 16(2):99–133.
27. Sullivan DF. A single index of mortality and morbidity. HSMHA health reports. 1971 Apr; 86(4):347–354. PMID: 5554262
28. Barro RJ. Economic growth in a cross-section of countries. Quarterly Journal of Economics. (1991) May 1; 106(2):407–443.
29. Nickell S. Biases in dynamic models with fixed effects. Econometrica: Journal of the Econometric Society. 1981 Nov 1; 49(6):1417–26.
30. Sargan JD. The estimation of economic relationships using instrumental variables. Econometrica: Journal of the Econometric Society. 1958 Jul 1; 26(3):393–415.
31. Roodman D. How to do xtabond2: An introduction to difference and system GMM in Stata. Stata Journal. 2009a; 9(1):86–136.
32. De Gregorio J. Economic growth in Latin America. Journal of development economics. 1992 Jul 1; 39 (1):59–84.

33. Roodman D. A note on the theme of too many instruments. Oxford Bulletin of Economics and statistics. 2009b Feb 1; 71(1):135–58.
34. Blundell R, Bond S. GMM estimation with persistent panel data: an application to production functions. Econometric reviews. 2000 Jan 1; 19(3):321–40.
35. Soto M. System GMM estimation with a small sample. BEWPS Working Paper No. 395, 2009
36. Weil DN. Health and economic growth. In Handbook of economic growth 2014 Jan 1 (Vol. 2, pp. 623–682). Elsevier.
37. GBD 2015 Healthcare Access and Quality Collaborators. Healthcare Access and Quality Index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990–2015: a novel analysis from the Global Burden of Disease Study 2015. The Lancet. 2017 Jul 15; 390 (10091):231–66.
38. https://data.unicef.org/topic/child-health/immunization (accessed 17th May, 2018)
39. Micah AE, Zlavog B, Friedman S, Reynolds A, Chapin AL, Schneider MT, Dieleman JL. The US Provided $13 Billion In Development Assistance For Health In 2016, Less Per Person Than Many Peer Nations. Health Affairs. 2017 Dec 1; 36(12):2133–41. https://doi.org/10.1377/hlthaff.2017.1055 PMID: 29200357
40. Sustainable Development Goals: https://www.un.org/sustainabledevelopment/health/ (Accessed 17th May, 2018)
41. GBD 2016 cause of death collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017 Sep 16; 390(10100):1151–1210.
42. GBD 2016 Mortality Collaborators. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970–2016: A systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017 Sep 16; 390(10100):1084–1150.
43. GBD 2016 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017 Sep 16; 390(10100):1260–1344.
44. Fischer S. The role of macroeconomic factors in growth. Journal of monetary economics. 1993 Dec 1; 32(3):485–512.
45. Bloom DE, Canning D, Graham B. Longevity and life-cycle savings. Scandinavian Journal of Economics. 2003 Sep; 105(3):319–38.
46. Bloom DE, Canning D, Malaney PN. Population dynamics and economic growth in Asia. Population and development review. 2000 Jan 1; 26:257–90.
47. Bleakley H. Health, human capital, and development. Annual Review of Economics. Mar 8; 2(1): 283:310.
48. Kalemli-Ozcan S, Ryder HE, Weil DN. Mortality decline, human capital investment, and economic growth. Journal of Development Economics. 2000 Jun 1; 62(1):1–23.
49. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017 Sep 16; 390(10100):1211–59.

69th International Astronautical Congress (IAC), Bremen, Germany, 1-5 October 2018.
Copyright ©2018 by the International Astronautical Federation (IAF). All rights reserved.

IAC-18-B3.9.5

Manned Mars Mission Risks Evaluation

Laura Bettiol [a], Andrea De La Torre[b], Divyesh Patel[c], Funmilola Oluwafemi[d], Guzel Kamaletdinova[e]*, Robin Kumar Singh[f], Uthpala Heshani[g], Yasith Lakmal[h], Aureliano Rivolta[i], Andrey Sorokin[j]

[a] *SGAC (Space Generation Advisory Council), c/o European Space Policy Institute, 6 Schwarzenbergplatz, Vienna, Austria, 1030, laura.bettiol@spacegeneration.org*
[b] *SGAC (Space Generation Advisory Council), c/o European Space Policy Institute, 6 Schwarzenbergplatz, Vienna, Austria, 1030, andrea.delatorre97@hotmail.com*
[c] *SGAC (Space Generation Advisory Council), c/o European Space Policy Institute, 6 Schwarzenbergplatz, Vienna, Austria, 1030, pateldivyesh6859@yahoo.in*
[d] *Space Life Science Unit, Engineering and Space Systems Department, National Space Research and Development Agency (NASRDA), Km 17 Airport Road, Abuja, Nigeria, P.M.B. 437, oluwafemifunmilola@gmail.com*
[e] *MAI (Moscow Aviation Institute), 4 Volokolamskoe shosse, Moscow, Russia, 125993, kamaletdinova.guzel@gmail.com*
[f] *DTU (Delhi Technological University) Shahbad Daulatpur, Main Bawana Road, Delhi, India, 110042, robinvishen@gmail.com*
[g] *SGAC (Space Generation Advisory Council), c/o European Space Policy Institute, Schwarzenbergplatz 6, 1030 Vienna, Austria, uthpalahesh@gmail.com*
[h] *SGAC (Space Generation Advisory Council), c/o European Space Policy Institute, Schwarzenbergplatz 6, 1030 Vienna, Austria, yasith.lakmal@spacegeneration.org*
[i] *SGAC (Space Generation Advisory Council), c/o European Space Policy Institute, Schwarzenbergplatz 6, 1030 Vienna, Austria, aureliano.rivolta@spacegeneration.org*
[j] *MAI (Moscow Aviation Institute), 4 Volokolamskoe shosse, Moscow, Russia, 125993, sorokin@mai,ru*
** Corresponding Author*

Abstract

Mars missions can be seen as a natural step in space exploration, as Earth-like environmental conditions and the length of the interplanetary flight are the most crucial for planetary exploration compared to other celestial bodies of the Solar System. On the other hand, these missions require detailed planning and worldwide collaboration, because of the need of extremely high financial resources and technical capabilities. Many researches focused on different aspects of Mars missions are being conducted in these recent years. They include flight preparation, lift-off, interplanetary journey, habitat and Life Support Systems (LSS) design, Extravehicular Activities (EVA), precise landing, planetary exploration, etc. This paper summarizes identified safety issues that can arise during future Mars missions. Based on the analysis of previous studies, where conditions of the interplanetary flight were studied, including environmental issues, Mars habitat and the spacecraft design were discussed and the spacesuit concept was analyzed, the most critical hazards have been defined and the whole mission has been taken into consideration. In particular, possible failures and hazards for the habitat on Mars, space station and spacesuit, including off-nominal situations and their influence on the safety of the astronauts have been investigated. This work is based on the results of previous projects carried out within the Space Generation Advisory Council's (SGAC) Space Safety and Sustainability project group, which aims to bring an international and interdisciplinary vision to this topic and discuss it from different perspectives, creating a foundation for further studies on safety risk assessment. Technological gaps are identified for further discussion and possible solutions for risks reduction are proposed with regard to onboard systems (including LSS) and layout design.

Keywords: safety, risks, Manned Mars Mission

Acronyms/Abbreviations
Extravehicular Activity (EVA)
International Space Station (ISS)
Low Earth Orbit (LEO)
Life Support System (LSS)
Nuclear thermal propulsion (NTP)

Copyright ©2018 by the International Astronautical Federation (IAF). All rights reserved.

1. Introduction

This paper summarizes reviews made by the authors with regards to their experience and the previous project dedicated to the One-way mission to Mars. This study is mostly focused on the mission itself with the priority of human safety during the journey, as the most important element in a manned mission is the astronaut. The spacecraft, Life Support Systems (LSSs) and the environment inside the space station are designed to provide humans with essential elements to live and to guarantee their working efficiency during the mission. Therefore, the man-made ecosystem for the interplanetary journey is built for human existence and survival.

There are many technological gaps related to the mission. This work stresses challenges of advanced closed-loop LSS design with the view to the breakthrough made by ESA [1] taking into consideration approximate calculation of needed supplies. Results of our previous project regarding the spacesuit design [2] act as a basis for this study on EVA with an accent to human safety and medical aspects.

The human health topic combines challenges rising particularly in space (gravity and radiation effects) with psychology. Offered astronautical hygiene solutions based on Health Stabilization program [3] for astronaut are combined with overview of supplement and nutrition effects. The biggest challenges related to radiation protection are still open, but offered solutions by "Radiation shielding of composite space enclosures" (SIDER) project [4] were taking into consideration.

Power system and engine play crucial roles in the mission and broach an ethical topic because of nuclear power usage, radiation debris, and consequence of accidents [5]. Results of the Kilopower project [6] which was offered for Mars Mission are discussed. As there is no off-the-shelf technology, the paper gives an overview of nuclear thermal propulsion, chemical and electric propulsion systems including its pros and cons.

2. Methods

The man-made ecosystem includes an LSS combined with the environment. The key element is the human who lives in it. Together LSS, environment and human can be defined as a complex system [7].

A man-made ecosystem is significantly different compared to a natural ecosystem. The human is a decision-making element and his/her safety is a key issue. At the same time it is important to control human's actions in case of an emergency, as stress can affect human control and behavior. Moreover, redundancy should always be adopted for higher safety [7].

As the ecosystem is geared towards the outer envelope, the importance of the materials used for spacecraft is taken into consideration as well as resources analysis.

Radiation effects are analyzed from different points of view: human health, electronics behavior and spacecraft protection.

Engine and power systems are discussed, including existing solutions and solutions that are being developed.

This paper summarizes the most important risks stressing their correlations and demonstrates that they have to be analyzed all together.

3. Theory and risks description

Considering previous experience on hazard analysis, the following points were taken as the basis of the whole mission analysis: human health, LSS operation, EVA, radiation influence, power systems and engine solutions, and resources analysis.

3.1 Risks associated with human health

Risks related to stress can be divided into five categories: gravity fields, isolation/confinement, hostile/closed environments, space radiation, and distance from Earth [8].

3.1.1 Gravity challenge and physiology

Most of the exploration missions so far have been carried out in Low Earth Orbit (LEO). This environment does not allow to study and explain many unknown situations and behaviors that would occur during a long-duration interplanetary mission [8].

During a mission to Mars, astronauts transition from a gravity pull of 9.807 m/s^2, to microgravity and then to 3.711 m/s^2 on the Red Planet. This will have important physiological effects on their body. The most important and known consequences of long permanence in a microgravity environment are: shift in body fluids, space motion sickness, muscle atrophy, bone demineralization, immune system deregulations [9] and disruption of senses such as vision and taste [10]. Upon arrival on Mars, the astronaut's body will have to adjust to its level of gravitational pull. This will also affect the astronauts' bone density, muscle strength, and circulation.

3.1.2 Psychological issues

Several psychological and social issues have been demonstrated to affect the crew in an isolated and confined environment during long periods of separation from society. In particular these elements have been identified: cultural aspects, interpersonal stressors, effects of long-term microgravity and radiation, extreme isolation and loneliness, limited social contacts and novelty, lack of support from Earth due to communication delays, problems including negative news coming from Earth (such as a family member or

Copyright ©2018 by the International Astronautical Federation (IAF). All rights reserved.

close friend's death), increased home sickness, depression, food, sleep, sexual attraction/tension, ethical issues, etc [10].

Much is known about the psychosocial issues that affect astronauts in space from anecdotal reports, scientific studies performed in space analog environments on Earth [7] (e.g., Antarctica, submarines, analog missions (e.g. isolation studies)), and experiments conducted during space missions. However, there is still a big knowledge gap to fill due to new stressors that appear only during the real long-duration missions and new flight conditions. Researches show that astronauts could experience a serious form of adjustment reaction called "asthenization", which is related to a psychiatric disorder called neurasthenia [11]. This disorder is defined as a weakness of the nervous system that produces fatigue, irritability and emotional liability, attention and concentration difficulties, restlessness, heightened perceptual sensitivities, palpitations and blood pressure instability, physical weakness, and sleep and appetite problems.

3.2 Life support systems

Life support systems can be divided into 3 categories: environmental monitoring, atmosphere management and water management [12]. A malfunction, damage or exhaustion of technical equipment, for example: filters, regenerating cartridges, water supply (tanks) or different gas balloons could compromise the spacecraft environment, making it unsafe for humans [7]. Combining with a challenge related to food supply a huge supply issue rises as it is not possible to receive additional sources support from Earth during an interplanetary flight.

The approximate time for a mission to Mars with the current level of technologies is 240 days of travel and 240 days of return. Summing the time that the astronauts will spend on the Red Planet, the mission would take from 2 to 3 years [11, 13].

An average calculation of the most important resources for a mission to Mars is based on human and equipment needs. The astronaut in space needs 30.60 kg of food, water, and air daily to stay healthy. A crew of 3 astronauts on a 2-year mission will need 66,000 kilograms of food, water, and air [14, 15]. It would be impossible to carry that amount of supplies from Earth.

Water is the most important resource for human beings. The amount of drinking water would be 2.3 kg-4.6 kg and washing water would be 1.1 kg-5.4 kg per person daily in the mission approximately. The approximate O_2 needed per day is 0.8 kg-0.9 kg per person [16]. An astronaut needs approximately 1.56 kg of food (1.83 kg including packaging) per day. [15] According to these data, a thousand-day mission to Mars will require 1,830 kg of food and packaging for each astronaut. [14, 15]

Depending upon the degree of recycling that can be achieved by the life support system, these requirements can be reduced by carefully using waste products, which are given per person per day:

- CO_2 : 0.8 kg-1.1 kg
- Water (H_20): 1.8 kg-2.5 kg
- Solid Waste: 0.13 kg-0.2kg [16]

The transportation of so much waste and so many resources would be almost impossible.

Current LSSs implemented on the International Space Station (ISS) are not fully closed-loop systems [12]. In addition, systems' elements of the heating or ventilation units have a limited service life [7]. These parameters should be taken into consideration.

3.2.1 Environmental monitoring

The first fundamental difference between a long interplanetary travel and the ISS in Low Earth Orbit (LEO) is the distance from a mission control center.

On the ISS, if an emergency that could compromise the health of the crew (such as leakage, malfunction of air filtering systems, unexpected chemical reactions etc.) occurs, current environmental controls alert crewmembers as well as the mission control center instantaneously. Current monitoring procedures include the shipping of air and water samples back to Earth. This procedure is not possible during long-distance spaceflight. Therefore, it is important to have the capability of continuously analyzing the air enclosed in the space station and the water available for the crew autonomously and in near-real time. [12]

3.2.2 Atmosphere management

The most critical function of an LSS is maintaining the atmosphere inside the vehicle safely breathable by humans. It must remove carbon dioxide and add oxygen, as well as prevent some gases that humans produce in small quantities (i.e. ammonia and acetone) from accumulating. [12] The LSS must also detect and remove potentially dangerous gases accidentally produced by chemical reactions in on-board scientific experiments or equipment. [12] The trace contaminant system on board the ISS currently uses absorbent media and catalysts that are commercially obsolete. Some alternatives are currently being identified and evaluated. Scroll filter technologies for particulate filtration are also under study. [17] Although all the atmosphere revitalization systems on board the ISS have reached TRL 9 [18], improvements must be made for long interplanetary missions [7, 17, 18].

Currently, the LSS system of the ISS generates/recycles about 40% of the total amount of oxygen needed by the astronauts using a Sabatier process reactor that extracts about 42% of O_2 from CO_2. It is essential that this percentage dramatically increases for interplanetary missions [12, 14].

Copyright ©2018 by the International Astronautical Federation (IAF). All rights reserved.

In addition to maintaining an optimal oxygen level, the LSS must reduce the amount of carbon dioxide that is mostly exhaled by the astronauts with their breath. Both improvements on the current ISS system and assessment of alternate technologies are now being examined by NASA [12]. Current ISS systems can support about 2.3 mmHg $ppCO_2$ for 4 people crew. The objective for interplanetary missions is to reduce it to <2 mmHg $ppCO_2$. Reliability must also be increased from the current <6 months mean time between failures [14].

Another key technology that has to be improved is the trace contaminant and particulate control system. Its aim is to keep the concentration of chemicals and particulates/aerosols below the spacecraft's maximum allowable amount. To improve this technology it is essential to increase its performances and reduce the consumables that are typically used in technologies such as filters and scrubbers [12].

3.2.3 Water management

The Water Recovery Subsystem of the ISS has reached TRL 9 [18] and is currently capable of recycling about 93% of the crewmembers' urine, wastewater and humidity [17]. The Urine Processing Assembly currently recovers about 80%; brine is stored for disposal [14].

The goal is to reach 98% of water loop closure. This value cannot be reached unless major developments on the water extraction from brine are achieved. This is challenging because the process of extracting water from wastewater brine causes a cascade of other problems and faults on traditional systems. [17]

Another aspect to consider is the water potability conservation during storage periods. Some developments with respect to the state-of-the-art ISS system are currently undergoing technology testing. It has been demonstrated that low concentration of silver can kill bacteria and maintain the water potability over long periods. Silver also brings a great advantage in terms of reducing consumables, as it is not necessary to remove it from the water before consumption. [17]

3.3 Extravehicular Activities

EVAs will play a crucial role in a Mars mission. EVAs will help to run maintenance activities supporting the whole operation of the habitat or station and carry out all needed scientific activities. Based on [19], the most relevant hazards for the spacesuit design include structural failures, power, thermal control and life support systems off-nominal situations, communication problems, loss in data management and problems related with humans in the crew.

Spacesuits are required to ensure the safety of the astronauts (from radiation, to supply oxygen if needed) and at the same time the efficiency of the planned operations.

Thermal extremes must be considered. Contacting an external hardware, which exceeds the touch temperature limit, can cause injury or even a loss of crewmember.

When touching a hot surface the pain threshold of the crew is reached before the physical injury. However, when the crew gets in contact with a cold surface, his hands and feet become numb before the injury or frostbite [19, 20].

All EVA spacesuits currently in use are designed to operate in microgravity and vacuum, since they are only being used in LEO. Even with a perfectly functioning suit, the astronaut's condition in the suit is important as uncomfortability in suit may lead to mission failure or to dangerous situations [20].

During the EVA, the reasons for potential electrical shock are identified as molten metal concerns. If the voltage is greater than 32V that can be a potential electrical hazard. [19, 20]

Due to the failure of the EVA safety tether, there is a potential to detach the EVA crew member. The safety tether has been designed to withstand the load of a crewmember. Before the crewmember reaches the end of the tether, the crewmember will be slowed down by the safety tether to reduce the load coming to the underlying structure. The crew members are trained to identify the structure that the safety tether can be attached. A test is performed to confirm that the EVA crew could get out of the articulating portable foot restraint easily and use the safety tether to return to the structure. If safety tether becomes snarled with the Orbiter Boom Sensor System, the EVA crewmember should release the safety tether and return.

3.3.1 Medical Risks of Extravehicular Activity

Decompression sickness known as "the bends" or "caisson disease" occurs when the human body is exposed to a low ambient pressure environment. When the crewmembers exit the pressurized spacecraft into the vacuum in a spacesuit for EVAs, decompression sickness can occur. Pressure reduction happens when spacewalkers exit the spacecraft. This can produce inert gas in the human bloodstream and tissues to come out of solution and form bubbles. These gas bubbles can embolize different body tissues including skin, lungs, joints and even brain damage. Also a nitrogen poisoning appears and affects body [19, 20].

3.4 Radiation effects

Due to the absence of magnetosphere, one of the biggest hazards associated to journeys between Earth and Mars is the exposure to high doses of radiation. Especially it becomes relevant for passing the Van Allen radiation belts [13, 21].

Galactic Cosmic Rays come from Milky Way or other distant galaxies. These particles are mostly

Copyright ©2018 by the International Astronautical Federation (IAF). All rights reserved.

protons and some of them are heavier ranging from Helium to heaviest elements. They come with the speed of light and knock apart the material they contact such as hard spacecraft wall, spacesuit of astronaut, habitat on Mars, and vehicle. This radiation is known as secondary [22].

Along with solar flares, the Sun's rapidly traveling rays expose astronauts to high levels of ionizing radiation. This form of radiation can damage the DNA in human cells, leading to mutations and higher risk of cancer as well as for cataracts, cardiovascular diseases, musculoskeletal issues and damage to the central nervous system and even impact of space radiation on neurodegeneration, in particular, the biological processes in the brain that contribute to the development of Alzheimer's disease. Exposure to radiation and high-energy particles were found to be associated with cognitive problems [13].

Another issue related to radiation is its influence on materials, its damage and human protection. The material damage could happen at atomic level. Moreover, charge accumulation and surface erosion could damage the whole satellite. The localized damage could be in the form of solid-state microelectronics failure or malfunctioning of a component. The macro-level failure could happen due to accumulation of localized damage that could lead to the failure of the overall mission [21].

Another risk that may cause failures during a mission is the accumulation of electric charges on the satellite surface. The main sources of spacecraft charging are ambient plasmas, sun rays, and the Earth's electromagnetic field. Charging happens when the spacecraft is traveling in space and interacts with external ionic environment. Accumulation of charges on the surface of satellite can cause electrostatic discharge and electronics components could stop working. Dielectric thermal coating and composite spacecraft structure are the main risk factors behind charging [23].

The following effects can appear:

- *Operational anomalies*: they include component failure, degraded sensor performance, phantom commands, missing signals, logic upsets, and telemetry glitches.
- *Physical damage*: spacecraft's surface can be damaged due to arc discharge, that causes surface heating and ejection of surface material from the discharging site. The ejected material can contaminate other surfaces of the same spacecraft.
- *Degradation damage*: the overall electrical and thermal properties of the satellite can be degraded. The highly charged (negative) surfaces will accelerate other positive ions towards it and this event will cause physical removal of top layer atoms.

Spacecraft has to be in the darkness to discharge itself. The charging of spacecraft is mainly for bigger satellites as the surface area increases, the possibility of charging increases [23].

3.5 Risks and Challenges associated with using Nuclear Power

One of the most important risks is dispersion of radioactive material in the atmosphere in the event of a catastrophic failure during launch. Due to multiple accidents in the past, the government and general public might oppose the use of nuclear power in spacecraft and satellites [24, 25].

As was said in Section 3.4, radiation shielding is very important to prevent problems with onboard electronics. However, onboard electronics systems are already complicated and to cover them with another layer is very difficult. Any small mistake leading to interference with the electronics might lead to failure of the system that can lead to accidents [26].

Another challenge is related to the crew. For safety of the crew members their compartment needs to be provided with radiation shielding which adds up to the total cost and complexity of the entire ship. Important to note that Uranium and Plutonium as well as all the systems required is expensive due to which funding becomes a big hindrance.

3.6 Engine

The engine is a critical as well as the biggest element of an interplanetary mission. Chemical and non-chemical propulsion systems, advanced and supporting technologies have been analyzed in [27].

Supporting technologies support an in-space propulsion system or subsystem but are not directly propulsive. The supporting technology areas are given significant consideration by the ISPSTA team including pervasive technologies (Integrated System Health Management, Materials and Structures, Heat Rejection and Power) and Cryogenic fluid management for propellants [27].

4. Possible solutions

Section 3 described the main challenges and hazards of a manned Mars mission. Based on this and following the sources review, the following solutions were marked.

4.1 Solutions to the risks associated with human health

Protecting astronauts' health from microgravity effects includes eating antioxidant-rich and high nutritious foods [13] and having required daily exercise to keep muscles and bones from deteriorating. Also the use of system of straps and buckles to help maintain an upright position should not be underestimated.

Copyright ©2018 by the International Astronautical Federation (IAF). All rights reserved.

Plant growth will be helpful for the crew as a relaxation procedure and as an Earth-like activity.

Space medicine has unique challenges, as space environment is characterized by microgravity and radiation as well as by isolation from the mission control with a time-delay.

4.1.1 Astronautical Hygiene

Astronautical hygiene and space medicine are the duet that ensure the health of astronauts in space travel.

Pre-travel infection control education should be given to astronauts. As part of the current Health Stabilization program [3], astronauts are isolated from others in the community for a period of time greater than the incubation period for most common viral upper respiratory tract, gastrointestinal, viral, and bacterial infections. Interventions for mitigating the risk of infection prior to flight include: health screening, including a complete dental exam, tuberculosis [3], Staphylococcus aureus [28], human immunodeficiency virus (HIV), fungal infections [29] and parasitic infections that may be endemic in the astronauts' home countries.

Consideration should be made for positive or neutral pressure within the containment vessel to reduce the risk of airborne microbes entering the containment vessel after docking if technically feasible.

Astronauts with signs or symptoms of respiratory tract infection should wear surgical masks to mitigate risk of transmission to other astronauts [30, 31]. Astronauts routinely take a vitamin D supplement, which may enhance immune function [32] and reduce risk of reactivation of herpes viruses [32, 33], and this practice should continue.

An unmet need that is currently being investigated is for astronauts to have access to equipment needed to detect microbial pathogens that cause infections, in order to direct appropriate therapeutic interventions and mitigate transmission risk [30, 31]. Challenges include testing for common viruses and bacteria, ease of use, durability of reagents, and output that does not require incubation.

Other hygiene practices include frequently sampling the space-cabin air and surfaces to detect early signs of a rise in microbial contamination, keeping surfaces clean using disinfected clothes, ensuring that all equipment is well maintained (in particular the LSSs) and regularly vacuuming the spacecraft to remove dust etc.

4.1.2 Psychology

Homesickness prevention should include sufficient, effective and suitable workload, coupled with frequent opportunities for private communication with families and self-controlled individual schedules [11].

The "Earth-out-of-view phenomenon" theorizes that when humans are traveling in outer space they might start to feel unconnected to the Earth and to those left behind [34]. This can be dealt with by having a telescope on board with which the crewmembers could clearly see Earth and its features.

Psychiatric screening needs to be done on people who apply to be astronauts or who are selected for important space missions to prevent suicidal and homicidal intent [11].

4.2 Existing and possible solutions for LSS

Resource management will play a huge role in the mission preparation. Supply calculation should be made based on the possible failures analysis and lifetime of the equipment. LSS should be tending to the close-loop concept [12, 14, 17].

Reaching 100% recycling is one of the top technical challenges that were identified by NASA in [17], as well as improving reliability (close to 100%) and making repairs easy to implement. In addition, it is extremely important to reduce faults with respect to the high fault rate of the ISS equipment, both on the US and Russian segments. Current fault rate is unacceptable for long duration interplanetary missions.

Regarding the analysis of the air in the station, NASA is currently developing modular integration of multiple sensing modalities, employing a combination of simple and rugged technologies, and using highly capable complex approaches where needed. [12] The objective is to achieve on-orbit identification and quantification utilizing non-culture based systems, with a fast response and a minimal crew time requirement.

In addition to air and water monitoring, an automatic sound monitoring system with alerting capabilities, currently implemented on the ISS with a hand-held sound level meter and manual crew assays, is foreseen to be developed for interplanetary missions [14].

Water has to be re-used. Water from baths, washes, sweat, urine, and air humidity is going through a catalytic oxidation system [35], which kills bacteria and viruses. The waste is removed from the water, producing purified water. When the water is received it can be used for drinking or for technical needs.

Recently oxygen onboard of the ISS has been received from water by electrolytic process [35] and one of the most challenging issues was the extraction of O_2 from CO_2. USA as well as Russia have been working on the maturation of several technologies that can reach much higher recycling percentages. For example, NASA is working with *pH Matter* on the utilization of the Bosch process, that is able to theoretically provide the highest percentage of oxygen recovery from CO_2, with the objective of reaching >90% of recycling. Also Sabatier-based processes have been studied to increase the oxygen recovery level because of the extraction process

Copyright ©2018 by the International Astronautical Federation (IAF). All rights reserved.

of hydrogen from methane. This process can reach similar recycling percentages with respect to the previously mentioned Bosch process [17]. Recently [1], ESA has finished the Advanced Closed Loop System for the ISS and it successfully works onboard of the station. This solution could be used for the Mars Mission reducing risks of the lack of water and oxygen supply.

Major improvements must be made to reduce at a minimum the required amount of water that has to be carried on the spacecraft during the multi-month interplanetary cruise as well as the mass and power consumption of the entire system. [12] In particular, the water management system must become closed-loop as much as possible, as water accounts for the highest percentage of material by mass.

4.3 EVA requirements

The LSS of the spacesuit controls the atmosphere inside the suit, providing CO_2 removal, oxygen supply, and backup oxygen to restore pressure in case of leak. The correct functioning of the LSS is critical to the suit, since its usual failure mode can lead to a quick death of the astronaut (due to CO_2 build-up or suffocation, for example). Also included in the LSS is water provision for consumption, although this usually consists of a simple bag attached to the astronaut's body [36].

A technology under development that can be considered for application in future Personal/Portable Life Support System include Pressure Swing Adsorption, a process by which CO_2 can be separated from gas more efficiently, and through a repeatable process, as opposed to the current LiOH canisters, that saturate with each use, and are limited to around 8 hours. By regenerating the sorbent during EVA, the size and weight of the sorbent canister can be greatly reduced. Pressure Swing Adsorption accomplishes this by venting CO_2 and water vapor into space [12, 14, 17]

To prevent main hazards regarding the spacesuit it is recommended to ensure all potential sharp edges are controlled before assessing the EVA worksite and tools. Moreover, the spacewalker has a warning in the procedures to avoid contact with the baffle. To eliminate the risk, the tools with sharp edges have to be avoided as much as possible during the EVA. To avoid this risk all repair and inspection tools should be thermally assessed to identify whether their use can reach a thermally hazardous extreme [37].

4.4 Radiation protection

To reduce the radiation effect, medical research suggests that astronauts can be given some medications (for example vitamins A and C, Radiogardase) [3]. Some diet foods such as omega-3-rich fish oils, pectin fiber from fruits and vegetables are countermeasures to damage from long-term radiation exposure. Another solution could be a medication, which can give faster recovery from radiation damage. It stimulates and balances the number of stem cells and progenitor cells in the body as they are impacted by high-energy radiation, but it gives some side-effects on the human body, as hypotension and weakness. [38]

To protect the crew from the outside environment it is important to closely monitor the radiation levels. It is suggested to use dosimeters. Their data can suggest which part of spacecraft is highly shielded and which part is under high radiation [13].

There is no final solution on protection in these cases, but the "Radiation shielding of composite space enclosures" (SIDER) project's researchers have found two different ways to produce light nano-conductive materials [4]. One utilizes nano-conductive materials and another one Tungsten (high in density) foil into a carbon fiber-reinforced plastic.

Composite materials were tested and produced good results. These results were showing such shielding materials are stronger than Aluminum. These materials can reduce whole satellite's weight along with providing enough shielding. In result, it will also reduce total amount of fuel. Some remarkable properties of carbon fiber-reinforced plastic are lightweight, high in strength, enhanced durability, extremely low thermal shrinkage and expansion [4].

Future spacecraft can be made of materials such as polyethylene plastic (RFX1, made of lightweight carbon and hydrogen atoms) and water particles (hydrogen) or a combination of both. According to [39], polyethylene is better at stopping solar particles (50% more) and cosmic galactic particles (15% more) compared to Aluminum. Nevertheless, water is composed of heavy oxygen elements, that make the satellite structure heavier compared to plastic structure. Concisely, light materials are very good at tolerating radiations but they cannot fully eliminate it.

NASA's researchers are working on electrostatic shielding materials that will generate stream of positive and negative charges to deflect the ionization radiation and charging due to plasma in deep space. However, existing shielding will make satellite heavier and costly to launch. Research shows that such shielding can reduce radiation impact till 30 to 35 percent overall, so undoubtedly astronauts will be exposed to 70% radiation in space [38].

4.5 Power system

Nuclear systems have shown attractive mass scaling with power, but require additional radiation shielding to prevent interference with onboard electronics and crew compartments [40, 41].

While solar power is much more commonly used, nuclear power offers great advantages in many areas. Solar cells, although efficient, can only supply energy to

Copyright ©2018 by the International Astronautical Federation (IAF). All rights reserved.

spacecraft in orbits where the solar flux is sufficiently high, such as low Earth orbit and interplanetary destinations close enough to the Sun. Nuclear reactors are especially beneficial in space because of their lower weight-to-capacity ratio than solar cells. Therefore, nuclear power systems take up much less space than solar power systems. Compact spacecraft are easier to orient and direct in space when precision is needed. Estimates of nuclear power, which can power both life support and propulsion systems, suggest that use of these systems can effectively reduce both cost and flight time [40, 41].

As an example, NASA's Kilopower (Kilopower Reactor Using Stirling Technology) project will allow transporting several of the systems on a single landing vehicle [4, 42]. This experimental project, started in October 2015, works on a novel design for nuclear reactors for space travel. The first Kilopower prototype reactor weights 134 kg and contains 28 kg of U235. The space rated 10 kWe Kilopower for Mars is expected to mass 226 kg and contain 43.7 kg of U235 [4, 42].

4.6 Engine Solutions

Engine solutions are widely discussed, but it is important to stress the technological gap behind it. One of the examined solutions is the Variable Specific Impulse Magnetoplasma Rocket (VASIMR) [5]. An electric power source ionizes hydrogen, deuterium, or helium fuel into a plasma by stripping away electrons. Magnetic fields then direct the charged gas in the proper direction to provide thrust. One of advantages of this approach is the mass economy as the spacecraft would not have to lift off carrying all the fuel it needs for the journey. The second relevant advantage is the mission length: the journey to Mars would be much faster [43].

It was shown [44], that a general approach could be feasible, but there are concerns on usage regarding the ethical issues (radiation debris, consequence of accidents etc) and the mission planning approach [43] as for the long-term mission the engine will require a nuclear reactor which will significantly add to the mass of the system.

Nuclear-powered spacecraft could allow human to abort the mission and return to Earth in case of problems [45]. Nuclear thermal propulsion (NTP) has advantages compared to chemical and electric propulsion systems.

Chemical propellants provide a high amount of thrust, but burn up quickly. In this case it is needed to send the fuel for the return trip in advance.

A major advantage of NTP engines is that they can run much longer than chemical rocket engines, such as the Space Shuttle main engines or the Merlin engines on SpaceX Falcon 9 rockets, and still produce significantly more thrust than an electric propulsion system, such as the ion thrusters used on satellites.

An electric propulsion engine—such as a Hall thruster that ionizes xenon gas—has a high specific impulse, as much as 10 times that of NTP. However, t the thrust is very low. This makes the technology good for making minor adjustments to the orbits of satellites. The biggest ion thrusters in space are only about 4 1/2 kW. A much larger, 100-kilowatt ion thruster known as X3 is currently being tested at NASA Glenn as a potential engine for missions to Mars, but even such a big ion thruster has a much lower thrust force than an NTP system.

5. Conclusions

Interplanetary flights can be seen as the future of human space exploration, but the planning of a mission needs a detailed analysis of every step and, at the same time, every element should be analyzed as a part of a whole system. Human is a central element of such complex system and it is a starting point for analysis and this system's disassembling. This strategy allows studying every point systematically taking into account the main reason of system design – human safety.

To realize a Mars mission scientists need to fill many technological gaps and to offer new solutions and approaches, which is not possible to do without global cooperation due to high costs.

This work tried to highlight the most important topics needed to develop as part of a mission and to give a survey of existing solutions or solutions under discussion.

This paper presented astronauts' health issues with regards to astronautical hygiene and psychology, analysis of LSS and resources, reflecting requirements to the closed loop LSS and existing technical opportunities, radiation effects that affects not only the most discussed human or shielding aspects, but also electronic functioning, power system as one of the key systems for the mission and the engine challenge. The work included technical as well as ethical discussions that stress authors' interdisciplinary approach.

The next step of this project will focus on the mission planning phase of a Mars mission. The project group will discuss the topic and propose recommendations.

References

[1] Advanced Closed Loop System, https://www.esa.int/Our_Activities/Human_Spaceflight/Research/Advanced_Closed_Loop_System, (accessed 25.09.2018).

[2] J. Lousada et al., Approaches and solutions for Martian spacesuit design, 68th International Astronautical Congress, Adelaide, Australia, 2017, 25 – 29 September.

Copyright ©2018 by the International Astronautical Federation (IAF). All rights reserved.

[3] Astronaut health protection, https://www.nasa.gov/centers/johnson/pdf/584739main_Wings-ch5d-pgs370-407.pdf, (accessed 23.03.2018).

[4] Space radiation protection, https://cordis.europa.eu/result/rcn/89638_en.html, (accessed 20.05.2018).

[5] Variable-Specific-Impulse Magnetoplasma Rocket, https://www.techbriefs.com/component/content/article/ntb/tech-briefs/mechanics-and-machinery/1768 , (accessed 15.07.2018).

[6] Kilopower Project, https://www.nasa.gov/sites/default/files/atoms/files/kilopower-media-event-charts-final-011618.pdf, (accessed 17.08.2018)

[7] G.R. Kamaletdinova, M.P. Onevsky, S.A. Skvortsov, Human safety in a man-made ecosystem. International Journal of Safety and Security Engineering. Vol. 8(3) (2018) 406-412.

[8] The Human Body in Space. https://www.nasa.gov/hrp/bodyinspace, (accessed 25.05.2017).

[9] David, W., Andre, K., Chiaki, M., Robert T., (2009). Acclimation During Space Flight: Effects on Human Physiology. NCBI. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2696527, (accessed 25.05.2017).

[10] Effects of Gravity on Human Body. http://www.newecologist.ru/ecologs-418-13.html, (accessed 25.05.2017).

[11] N. Kanas. The New Martians, a scientific novel. http://www.springer.com/978-3-319-00974-2, (accessed 15.07.2017)

[12] Life Support Systems. https://www.nasa.gov/content/life-support-systems, (accessed 17.06.2018).

[13] J. Lousada et al., Technical Safety Analysis of a One-Way Human to Mars Mission, 66th International Astronautical Congress, Jerusalem, Israel, 2015, 12 – 16 October.

[14] M. Anderson, Moon/Mars Life Support Systems –How far along are we?, https://pdfs.semanticscholar.org/presentation/9d0b/d749ae9fa94f9470c5f9140f018815cd7c8a.pdf, (accessed 24.05.2017).

[15] Mars' Surface Radiation Environment Measurements with MSL Curiosity Rover, http://www.cosmic-rays.ru/articles/Mars/mars01.pdf, (accessed 15.05.2018).

[16] G. James, Ph.D. G. Chamitoff, Ph.D. D. Barker, M.S., M.A., Resource Utilization and Site Selection for a Self-Sufficient Martian Outpost, NASA/TM-98-206538, National Aeronautics and Space Administration Lyndon B Johnson Space Center Houston, Texas 77058-4406, April, 1998

[17] M. Sargusingh et al., NASA Environmental Control and Life Support Technology Development and Maturation for Exploration: 2017 to 2018 Overview, 48th International Conference on Environmental Systems, Albuquerque, New Mexico (USA), 8-12 July 2018.

[18] NASA Space Technology Roadmaps and Priorities: Restoring NASA's Technological Edge and Paving the Way for a New Era in Space, https://www.nap.edu/download/13354#, (accessed 17.06.2018).

[19] G. Musgrave, A. Larsen, T. Sgobba, Safety Design for space system, Butterworth-Heinemann, 2009.

[20] Volume I. EVA Physiology, https://msis.jsc.nasa.gov/sections/section14.htm#_14.2_EVA_PHYSIOLOGY (accessed 25.04.2018).

[21] D.F. Medina, Solar Radiation and Spacecraft Shielding, in: F. Allahdadi, J. Pelton (eds) Handbook of Cosmic Hazards and Planetary Defense, Springer, Cham, 2014, https://link.springer.com/referenceworkentry/10.1007/978-3-319-02847-7_10-1, (accessed 20.04.2018).

[22] How to Protect Astronauts from Space Radiation on Mars, https://www.nasa.gov/feature/goddard/real-martians-how-to-protect-astronauts-from-space-radiation-on-mars, (accessed 14.05.2018).

[23] Avoiding Cosmic Catastrophes: Designing Rad-Hard Devices, https://www.ecnmag.com/article/2017/01/avoiding-cosmic-catastrophes-designing-rad-hard-devices, (accessed 20.05.2018).

[24] Radiation hardened space electronics, https://www.militaryaerospace.com/articles/print/volume-28/issue-6/technology-focus/radiation-hardened-space-electronics-enter-the-multi-core-era.html, (accessed 18.06.2018).

[25] The perils of nuclear power space flights, https://www.counterpunch.org/2015/06/29/the-perils-of-nuclear-powered-space-flights/, (accessed 03.06.2018).

[26] Challenges for electronic circuits in space application, http://www.analog.com/en/technical-articles/challenges-for-electronic-circuits-in-space-applications.html, (accessed 18.06.2018).

[27] In-Space Propulsion Systems Roadmap https://www.nasa.gov/pdf/501329main_TA02-ID_rev3-NRC-wTASR.pdf, (accessed 23.03.2018).

[28]] L. R. Peterson, O. Liesenfeld, C. W. Woods, et al. Multicentered evaluation of the LightCycler methicillin-resistant Staphylococcus aureus (MRSA) advanced test as a rapid method for detection of MRSA in nasal surveillance swabs, J Clin Microbiol , 2010, vol. 48, pp. 1161-1166.

Copyright ©2018 by the International Astronautical Federation (IAF). All rights reserved.

[29] F. M. Williams, V. Markides, J. Edgeworth, A. J. Williams, Reactivation of coccidioidomycosis in a fit American visitor, Thorax , 1998, vol. 53 (pp. 811-822).

[30] D. F. Johnson, J. D. Druce, C. Birch, A quantitative assessment of the efficacy of surgical and N95 masks to filter influenza virus in patients with acute influenza infection, Clin Infect Dis , 2009, vol. 49, pp. 275-277.

[31] I. Nevo, M. Fitzpatrick, R. E. Thomas, et al., The efficacy of visual cues to improve hand hygiene compliance, Simul Healthc , 2010, vol. 5, pp. 325-331.

[32] A. R. Martineau, et al. High-dose vitamin D(3) during intensive-phase antimicrobial treatment of pulmonary tuberculosis: a double-blind randomised controlled trial, Lancet , 2011, vol. 377, pp. 242-250.

[33] S.R. Zwart SR, et al. Response to vitamin D supplementation during Antarctic winter is related to BMI, and supplementation can mitigate Epstein-Barr virus reactivation, J Nutr , 2011, vol. 14, pp. 692-697.

[34] Manzey, Dietrich, Human missions to Mars: new psychological challenges and research issues. Acta Astronautica, 2004, vol 55.3-9, pp. 781-790.

[35] NIIchimmash, https://www.niichimmash.ru/press/publications/60-let/, (accessed 17.08.2018).

[36] M. C. E. Utrilla et al., Safety Analysis of Spacesuit Design for Martian Surface, 67th International Astronautical Congress (IAC), Guadalajara, Mexico, 2016, 26-30 September.

[37] Space Exploration Beta, https://space.stackexchange.com/questions/13722/is-space-suit-using-algae-for-oxygen-production-possible, (accessed 28.05.2018).

[38] Radiation Educator Guide – NASA, https://www.nasa.gov/pdf/284275main_Radiation_HS_Mod3.pdf, (accessed 21.05.2018).

[39] Mars OXygen ISRU Experiment Project, NASA TechPort. National Aeronautics and Space Administration. Retrieved 19 November, 2015.

[40] Non-power nuclear application, http://www.world-nuclear.org/information-library/non-power-nuclear-applications/transport/nuclear-reactors-for-space.aspx#.UT1x3VpRgz1, (accessed 18.06.2018).

[41] Mars 2019, https://www.nasa.gov/pdf/382362main_40%20-%2020090801.1.mars2019.pdf, (accessed 08.05.2018).

[42] Nuclear Power in Space, http://www.spacedaily.com/reports/Nuclear_Power_In_Space_999.html, (accessed 22.07.2018).

[43] Vasimir Plasma Engine, http://www.spaceflightinsider.com/conferences/humans-to-mars/vasimr-plasma-engine-earth-mars-39-days/#8upAm2ltHD4a2wxY.99, (accessed 10.07.2018).

[44] Multi-MW Closed Cycle MHD Nuclear Space Power Via Nonequilibrium He/Xe Working Plasma, https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20110009914.pdf, (accessed 11.05.2018).

[45] NASA's Nuclear Thermal Engine, https://www.popularmechanics.com/space/moon-mars/a18345717/nasa-ntp-nuclear-engines-mars/, (accessed 23.06.2018).

International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934

Modelling of vapour compression refrigeration system (VCRS) by using alternative refrigerants with TiO_2 and without nano materials

R.S. Mishra

Department of Mechanical & Production Engineering, Delhi Technological University Delhi, India

Abstract

Application of Nano particles in refrigerants has been identified as a better way of enhancing the thermal performance of the vapour compression refrigeration system (VCRS) without modification the system design. When nano particles are dispersed in a refrigerant, they are regarded as nano refrigerants. The improvement in evaporator and condenser heat transfer coefficient are responsible for the enhancement of VCRS performance.

In this paper, the effect of TiO_2 (nano particles) on the first law efficiency in terms of coefficient of performance (COP), exergetic efficiency (second law efficiency0 and system exergy destruction ratio (EDR) based on exergy of fuel / exergy of product) using alternative refrigerants are discussed. It is shown that the application of Nano particles as additives in refrigerant and lubricant in VCRS is favorable and promising. Therefore, Nano refrigerants are expected to be the future refrigerants for improving thermal performances of vapour compression refrigeration systems.

© 2018 ijrei.com. All rights reserved

Keywords: Thermodynamic performance improvements, Mixing Nano materials, Vapour Compression Refrigeration System

1. Introduction

Refrigeration systems have become one of the most important utilities for people's daily lives. With the advancement and technological developments in the field of refrigeration new methods are developed to increase the COP of the systems. Traditional methods for exchanging heat from the system involves increment in the surface area but this leads to the increase in the size of the system, so there was need of some efficient way that can enhance the heat transfer. The inventive idea of using nanofluids, fluids which consist of suspended nanoparticles are used to remove such kind of barriers. The nano refrigerant is used to improve system performance, and it is decreased energy consumption of the system (viz. enhance heat transfer rate in cooling coil). Nano particles of Al_2O_3, TiO_2 and CuO are used due to their higher thermal conductivity to achieve better efficiency. Mishra et al. [1] experimentally evaluated the performance of a vapour compression refrigeration system by using Cu, Al_2O_3,CuO and TiO_2 based nano refrigerants in the primary circuit. The experimental results showed that the C.O.P of the system using Al_2O_3/R134a nano refrigerant was enhanced by 35% which was highest among all other nanorefrigerant. Sabareesh et al. [2] experimentally investigated 17% increase in COP by using 0.01% by volume concentration TiO2 nano fluid in VCRS as a lubricant additive.. Shengshan Bi et al. [3] found performance improvement reduction in power consumption upto 25% by mixed mineral oil TiO2 as the lubricant with refrigerant R600a

Subramani & Prakash [4] found improvement in performance by using Al_2O_3 at 0.06% weight in the mineral oil. Kumar et al. [5] found improvement in COP by 19.6% and reduction in power consumption by 11.5%. Due to enhancement in heat transfer coefficient using nanofluid Al_2O_3 & R600a/mineral oil as working fluid in a domestic refrigerator. Abbas et al. [6] performed the analysis of an air conditioning system by using a concentration of 0.01-0.1wt% of CNT Polyester oil with refrigerant R134a and found 4.2% enhancement in the COP was enhanced by using CNT particles concentration of 0.1% by weight. Hussen [7] showed the conventional

Corresponding author: R. S. Mishra
Email Address: hod.mechanical.rsm@dtu.ac.in

refrigeration system performance improved with nano refrigerant. Compressor work decreases by about 13.3% and system C.O.P. increases by about 12%.

2. Results and Discussion

The numerical computation was carried out for finding thermal performance of vapour compression refrigeration from developed model, the following input data have been assumed.

Condenser temperature=48^{0}C, Evaporator temperature-5^{0}C, mixing of nano particles =5% by volume.

2.1 Effect of refrigerants with mixing TiO_2 nano particles

Table-1 shows the variation of thermal performances with titanium di oxide in the different ecofriendly refrigerants and it was found that by using nano in the R410a gives worst thermal performance in terms of first law efficiency, second law efficiency and exergetic efficiency. The best thermal performance in terms of first law efficiency is found by using R152a. By using hydrocarbons mixed TiO_2 nano the thermal performance of R600a hydrocarbon is lower than by using TiO_2 in the R290a. By comparing R134a and R407C, R404a, R507a and R125, it is observed that R407c gives better performance when TiO_2 is used in these refrigerants. With comparing thermal performances from Table-1, without nano by using R134a, the first law performance (COP) and second law performance (exergetic efficiency) increased significantly. Two types of EDR were computed. EDR computed based on exergy of product i.e. in terms of exergy of fuel in terms of total power utilized for running compressor for finding the components defects (Exergy losses) in the system as shown in Table-1 and effective utilization of cooling load (Q_Eva*(1-(T_Dead_State/T_eva)) in Table-1 respectively.

Table-1: Variation of thermal performances (First law efficiency (COP) and Exergetic efficiency) with ecofriendly refrigerants using TiO_2 nano material mixing

Refrigerants	First law COP	EDR	Exergetic Efficiency
R134a	4.36	1.659	0.3761
R404a	4.303	1.704	0.3698
R407c	4.634	1.458	0.4068
R125	4.044	1.935	0.3407
R507a	4.328	1.685	0.3725
R290	4.826	1.335	0.4282
R600a	4.009	1.969	0.3368
R152a	5.169	1.143	0.4666
R410a	2.138	6.847	0.1274
R502	4.671	1.433	0.4110
R134a without nano	2.978	2.44	0.2904

2.2 Effect of Thermal performances first law Performance (COP) on VCRS using TiO_2 Nano Material

The variation of evaporator temperature with variation of first law thermal performances in terms of COP and second law efficiency (exergetic efficiency) for different ecofriendly refrigerants withTiO_2 nano particles and without nano by using R134a is shown in Table-2 (a) and table-3(a) respectively. It was found that by increasing evaporator temperature the first law performance (COP) increases. And best first law performance (COP) was found by using R152a since R152a and hydrocarbon are flammable, therefore R134a gives better first law performance (COP). Similarly by increasing condenser temperature, the first law thermal performance (COP) decreases as shown in table-2(b) and table-3(b) respectively.

Table-2(a) Variation of COP with Evaporator temperature of Vapour compression using TiO_2 nano materials

T_Evaporator (K)	R134a	R404a	R125	R290	R600a	R410a	R507a	R152a	R502
283	7.932	7.12	6.715	7.802	6.866	2.557	7.09	9.273	7.645
278	6.595	5.942	5.595	6.548	5.751	2.416	5.945	7.734	6.396
273	5.517	5.048	4.746	5.599	4.873	2.282	5.06	6.493	5.448
268	4.36	4.303	4.044	4.826	4.009	2.138	4.328	5.166	4.671
263	2,182	3.525	3.327	4.086	2.588	1.932	3.58	2.70	3.908
258	--------	2.242	2.192	3.043	---------	1.466	2.399	-------	2.769

Table-2(b) Variation of COP with condenser temperature of Vapour compression using TiO_2 nano materials

T_Condenser (K)	R134a	R404a	R125	R290	R600	R410a	R507a	R152a	R502
329	1.992	3.591	3.025	3.768	1.484	1.99	3.249	2.434	3.591
325	3.342	4.118	3.532	4.281	2.323	2.087	3.785	3.99	4.118
321	4.36	4.671	4.044	4.826	3.605	2.138	4.328	5.169	4.671
317	5.229	5.294	4.605	5.444	4.578	2.166	4.924	6.185	5.294
313	6.096	6.035	5.26	6.18	5.446	2.183	5.62	7.207	6.035

Table-3(a) Variation of Exergetic Efficiency with evaporator temperature of Vapour compression using TiO_2 nano materials

T_EVA (K)	R134a	R404a	R125	R290	R600a	R410a	R507c	R152a	R502
283	0.3674	0.3556	0.3029	0.3605	0.3639	0.08253	0.3557	0.4385	0.3522
278	0.4025	0.3707	0.3306	0.3991	0.4137	0.1019	0.3718	0.4845	0.3882
273	0.4136	0.3698	0.3430	0.4211	0.4462	0.1174	0.3725	0.497	0.4073
268	0.3761	0.3360	0.3407	0.4282	0.4488	0.1274	0.3434	0.4666	0.4110
263	0.1573	0.1926	0.3049	0.4107	0.3444	0.124	0.2169	0.2262	0.4387

Table-3(b) Variation of COP with condenser temperature of Vapour compression using TiO_2 nano materials

T_EVA (K)	R134a	R404a	R125	R290	R600a	R410a	R507c	R152a	R502
329	0.1111	0.2469	0.2267	0.3099	0.1593	0.2228	0.2517	0.1606	0.290
325	0.2622	0.3083	0.2834	0.3673	0.2571	0.2337	0.3117	0.3348	0.3491
321	0.3761	0.3648	0.3407	0.4282	0.3368	0.2394	0.3725	0.4666	0.5890
317	0.4734	0.4369	0.4036	0.4975	0.4110	0.2425	0.4392	0.5804	0.4807
313	0.5704	0.5152	0.4769	0.5798	0.490	0.2443	0.4828	0.6948	0.5634

3. Conclusion

Following conclusions were drawn from present investigation.

(i) By using nano in the R410a gives worst thermal performance in terms of first law efficiency, second law efficiency and exergetic efficiency.

(ii) The best thermal performance in terms of first law efficiency is found by using R152a.

(iii) By increasing evaporator temperature the first law performance (COP) increases and best first law performance (COP) was found by using R152a

(iv) R152a and hydrocarbon are flammable, therefore R134a gives better first law performance (COP) and second law performances.

(v) By increasing condenser temperature, the first law thermal performance (COP) decreases.

References

[1] Mishra R S and Jaiswal Rahul Kumar [2015] Thermal Performance Improvements of Vapour Compression Refrigeration System Using Eco-Friendly Based Nanorefrigerants in Primary Circuit, International Journal of Advance Research and Innovation **3,** page-524-535.

[2] Sabareesh R K, Gobinath N, Sajith V, Das S and Sobhan C B 2012 Application of TiO 2 nanoparticles as a lubricant-additive _ for vapor compression refrigeration systems–An experimental investigation International journal of refrigeration 35, page-1989-1996.

[3] Bi Shengshan, [2011] Performance of a domestic refrigerator using TiO 2-R600a nano-refrigerant as working fluid Energy Conversion and Management **52** 733-737

[4] Bandgar M S, Ragit S S, Kolhe K P and Biradar N S 2016 Effect of Nano Lubricant on the Performance of Vapour Compression Refrigeration System: A Review JETIR **3,** page- 2349-5162.

[5] Subramani N and Prakash M J 2011 Experimental studies on a vapour compression system using nanorefrigerants International Journal of Engineering, Scie-nce and Technology ,3, page- 95-102.

[6] Kumar R R, Sridhar K and Narasimha M 2013 Heat transfer enhancement in the domestic refrigerator using *R600a/mineral* oil/nano-Al_2O_3 as working fluid International Journal of Computational Engineering Research **3,** page-42-50.

[7] Abbas M, [2013] Efficient Air—Condition Unit By Using Nano—Refrigerant 1st Engineering undergraduate research catalyst conference.

[8] H.A. Hussen [2014] Experimental investigation for TiO2 nanoparticles as a lubricant- Additive for a compressor of window type air-conditioner system **J.** Eng., 20 (2014), pp. 61-72

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/ACCESS.2018.2870916, IEEE Access

Negative Capacitance Transistor to Address the Fundamental Limitations in Technology Scaling: Processor Performance

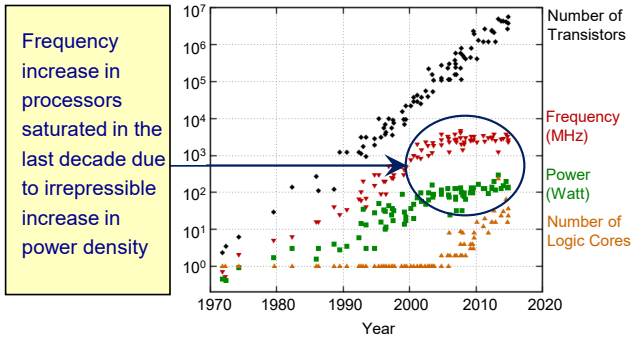
Hussam Amrouch, *Member, IEEE,* Girish Pahwa, *Student Member, IEEE,* Amol D. Gaidhane, *Student Member, IEEE,* Jörg Henkel, *Fellow, IEEE,* and Yogesh S. Chauhan *Senior Member, IEEE*

***Abstract*—Negative Capacitance Field-Effect Transistor (NCFET) addresses one of the key fundamental limits in technology scaling, akin to the non-scalable Boltzmann factor, by offering a sub-threshold swing below 60 mV/decade. In this work, we investigate how NCFET technology can open the doors not only for the continuation of Moore's law, which is approaching its end, but also for reviving Dennard's scaling, which stopped more than a decade ago. We study NCFET for the 7 nm FinFET technology node, from physics to processors and demonstrate that prior trends in processor design with respect to voltage and frequency can be revived with NCFET technology. Our work focuses on answering the following three questions towards drawing the impact of NCFET technology on computing efficiency: In how far NCFET technology will enable processors (i) to operate at higher frequencies without increasing voltage, (ii) to operate at higher frequencies without increasing power density, which is substantial, because maintaining on-chip power densities under tight constraints due to limited cooling capabilities is inevitable, and (iii) to operate at lower voltages, while still fulfilling performance requirements, which is substantial for the emerging Internet of Things (IoTs), in which, available power budgets for such devices are typically very restricted.**

***Index Terms*—Negative capacitance, emerging technology, sub-threshold swing, performance, power, Dennard's scaling**

I. Introduction

In the early 2000s, when the complementary metal-oxide semiconductor (CMOS) technology entered the so-called nano era, in which the feature sizes of transistors became below 90 nm, the prior trend of voltage scaling started to falter. Since then, the well-known Dennard's scaling [2], which states that both the dimensions of transistors and the operating voltage should be scaled by the same factor to maintain a constant electric field, had stopped. As a matter of fact, the operating voltage of processors (V_{DD}) remained almost constant in the last decade, whereas the transistor's dimensions steadily kept scaling down following Moore's law [3]. Consequently, unsustainable heat, that inevitably generated due to excessive on-chip power densities, became a crucial obstacle for technology scaling [4]. To overcome this challenge, the maximum frequency of processors stopped increasing with every new generation, in order to limit the speed of moving electrons inside the chip and, hence, restrict their ability to generate heat [5]. Since then, processor manufacturers have moved from a single-core design to multi-core design, to keep providing processors with better performance with every new generation, despite the enforced limit on the maximum operating frequency. Although this enabled the continuation of Moore's law in the last decade, the latter will unavoidably stop in the near feature as the transistor scaling is approaching its end. This is expected to happen when the dimensions of transistors reach 2-3 nm, where quantum effects become predominant [5]. Moving to 3D architectures, in which several layers are stacked on top of each other, may allow processors to keep moving along the Moore's performance curve, but removing the ever-increasing heat generated by chips is challenging, and this may limit their deployment to memory chips [5] as they involve less power densities than processing units. Fig. 1 demonstrates that around 2005, the frequency increase in processors was discontinued to avoid unsustainable power densities, despite the continuous increase in the number of transistors.



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, C. Batten. New plot and data collected for 2010-2015 by K. Rupp

Fig. 1. The trends in microprocessors based on gathered experimental data. The data and plotting script obtained from [1].

In summary, the prior trends of voltage and frequency scaling stopped since more than a decade in which neither the speed of processors nor the operating voltage has improved. In addition, the current trend of performance improvement in processors may also stop soon, as increasing performance through integrating more CPU cores with every new generation, becomes difficult, if not impossible, because the

H. Amrouch and J. Henkel are with the Department of Computer Science, Karlsruhe Institute of Technology (KIT), Haid-und-Neu Strasse 7, 76133 Karlsruhe, Germany. {amrouch, henkel}@kit.edu.
G. Pahwa, A.D. Gaidhane and Y. Chauhan are with Electrical Engineer Department, Indian Institute of Technology Kanpur (IITK), Kanpur, India. {girish, amold, chauhan}@iitk.ac.in. The work from IITK was partially funded by Department of Science and Technology, Science and Engineering Research Board, Govt. of India and Berkeley Device Modeling Center.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/ACCESS.2018.2870916, IEEE Access

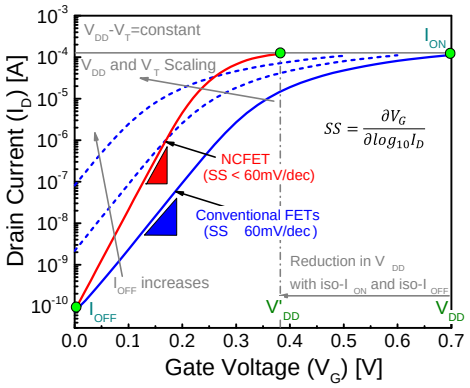


Fig. 2. Impact of supply voltage (V_{DD}) and threshold voltage (V_T) co-scaling on the transfer characteristics of transistor. The overdrive ($V_{DD} - V_T$) is kept constant to maintain the same ON current. With V_T scaling, OFF current (I_{OFF}) starts to exponentially increase due to non-scalable sub-threshold swing (SS), which restricted V_{DD} scaling in the last decade. On the other hand, the steep SS in NCFET technology effectively mitigates this issue and achieves the same I_{ON}, as the conventional FETs at iso-I_{OFF}, but at a lower operating voltage ($V'_{DD} < V_{DD}$). In this experiment a conventional $7\,\mathrm{nm}$ nFinFET is studied as a baseline and then compared to NC-FinFET that has a $4\,\mathrm{nm}$ ferroelectric layer.

process scaling is approaching its end. Therefore, besides the discontinuation of Dennard's scaling, Moore's law may also die in the fear future [6].

A. The fundamental limitation behind voltage scaling

The speed of a processor is essentially governed by the drive current (ON current, I_{ON}) that its constituent transistors can provide. I_{ON} of a transistor is proportional to $(V_{DD} - V_T)$, where V_T denotes threshold voltage. To maintain the same level of I_{ON} with V_{DD} scaling, V_T also must be reduced by the same amount. Reducing V_T, however, exponentially increases the leakage current (OFF current, I_{OFF}) of transistor, which in turn raises the stand-by power consumption of a processor. This is primarily because of the non-scalable sub-threshold swing SS of transistor, which determines the change in gate voltage required to cause a decade change in the current as shown in Fig. 2. As a matter of fact, SS is fundamentally limited to $60\,\mathrm{mV/decade}$ at room temperature due to the Boltzmann distribution of charge carriers at the source of transistor – often termed as the "Boltzmann tyranny" [7], [8].

This fundamental limit restricts the minimum possible V_T to be $300\,\mathrm{mV}$, because in a good design, the I_{ON}/I_{OFF} ratio should be at least 5 orders of magnitudes. To ensure a reliable operation, safety margins need to be included, so that at the end the V_{DD} reaches around $1.0\,\mathrm{V}$ or slightly below. Precisely, this is the main reason behind the discontinuation of V_{DD} scaling, which, in turn, had led to preventing the frequency of processors from increasing to avoid unsustainable power densities and heat as explained earlier.

B. Negative Capacitance FET (NCFET)

NCFET is an emerging transistor technology, that can surmount the fundamental Boltzmann constraint of $60\,\mathrm{mV/decade}$ sub-threshold swing at room temperature, through integration of a ferroelectric layer within the gate stack. The concept of NCFET was first proposed in 2008 by S. Salahuddin and S. Datta [9]. Gradually, it gained popularity due to its remarkable characteristics, such as steep switching, high ON current and excellent short channel immunity, i.e., reduced short channel effects because Drain-Induced Barrier Lowering (DIBL) effect is lower or even negative [9]–[13]. Both theories and experiments have consistently proved its potential at the device level as well as at the circuit level [14]–[19]. In practice, NCFET technology enables the transistor to reach the same ON current, without increasing the OFF current, but at a lower V_{DD}, due to steeper sub-threshold swing (i.e., $SS < 60\,\mathrm{mV/decade}$) as we demonstrate in Fig. 2. As it can be noticed, due to the steeper SS in NCFET technology, the same I_{ON}, as the conventional FETs at iso-I_{OFF}, can be achieved but at a lower operating voltage.

Limitations in NCFET technology process: like any other emerging technology, NCFET also faced some process-related issues and intrinsic ferroelectric limitations. One of the main process challenges is the integration of ferroelectric material in the gate stack with a high quality interface with the silicon. An inflection point occurred in 2012, after the discovery of ferroelectricity in HfO_2 based materials, which made the NCFET technology compatible with the standard CMOS technology process [20]–[22]. Subsequently, several experimental works have demonstrated NCFETs with HfO_2 based ferroelectrics exhibiting sub-$60\,\mathrm{mV/decade}$ swing at room temperature [23]–[27]. A recent work has also proposed a method to reduce interface states between ferroelectric and silicon using a nitrided interfacial oxide layer [28].

Importantly, GlobalFoundries has recently integrated a doped HfO_2 ferroelectric layer in their state-of-the-art industrial $14\,\mathrm{nm}$ FinFET technology demonstrating for the first time how NCFET-based circuits can be successfully manufactured using the standard current CMOS technology [18]. Therefore, NCFET has recently become one of the few emerging technologies that is compatible with the existing CMOS manufacturing process, which may pave the way for massive production of NCFET. Additionally, there have been several other experimental works on NCFETs with the 2D MoS_2 as the active channel material which can provide additional boost over the Si channel based technology [29]–[32].

The other main limitation that is often projected is the polarization damping of ferroelectric dipoles, due to which dipoles may not respond fast enough to the applied voltage during high frequency switching events [33]–[35]. Therefore, it was questionable that whether the switching frequency of NCFETs can actually match the GHz frequency operation in modern processors. However, this concern has recently been discarded. In their experimental work, [18] demonstrated that when NCFET is integrated into $14\,\mathrm{nm}$ FinFET technology, ring-oscillator circuits can operate at frequencies similar to those in the original (i.e., baseline) $14\,\mathrm{nm}$ technology. Another theoretical study has predicted the intrinsic delay due to polarization damping in NCFET to be very small ($\approx 270\,\mathrm{fs}$) in doped HfO_2 ferroelectrics, which can be easily ignored in digital logic applications [36]. This prediction is based on the fact that the actual amount of gate charge to be switched in

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/ACCESS.2018.2870916, IEEE Access

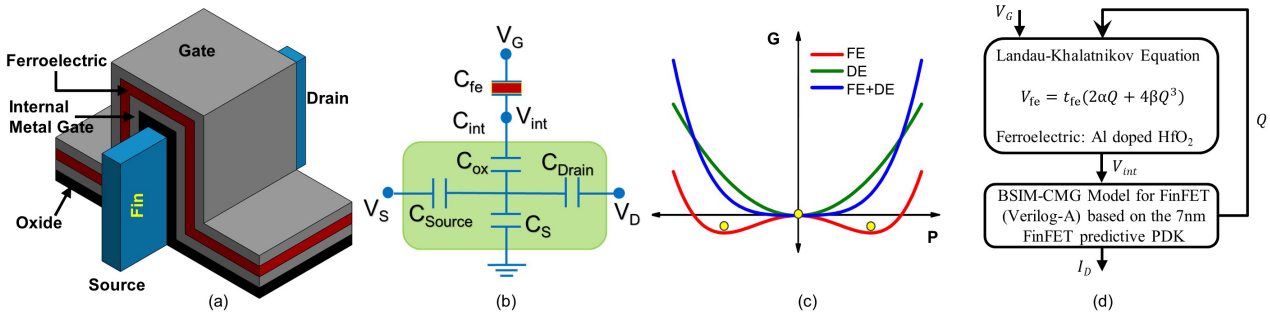


Fig. 3. (a) Schematic of metal-ferroelectric-metal-insulator-semiconductor (MFMIS) NCFET. Presence of intermediate equipotential metal gate enables the partitioning of the full structure into two circuit components: the ferroelectric capacitor and the baseline FinFET for device modeling purpose. (b) Circuit equivalent of MFMIS type NCFET. Shaded region represent gate capacitance of the internal baseline FinFET. (c) Gibbs free energy profiles at zero external electric field as a function of polarization, for ferroelectric (FE), dielectric (DE) and the combined system (FE+DE). In an isolated ferroelectric, either of the two energy minima are occupied. However, addition of a dielectric in series with the ferroelectric produces an energy profile of the system, with a minimum at $P = 0$, where ferroelectric lies in the negative capacitance state. (d) Modeling flowchart for NC-FinFET. The L-K model of ferroelectric capacitor is solved self-consistently within the industrial BSIM-CMG model of FinFET, in order to obtain drain current (I_D) of NC-FinFET.

NCFET is very small (inversion gate charge density in a silicon FET is $\approx$ 1.6 μC/cm^2 [36], much smaller than the remnant polarization of the available ferroelectrics which is generally in the range of tens of μC/cm^2).

C. Key focus areas in this work

In this work, we investigate the implications of NCFET technology, at the 7 nm FinFET node, starting from physics all the way up to the processor level, in the following areas:
(1) Reviving the prior trend of frequency increase in processors without increasing power densities. In other words, under the same power density constraint, how much the frequency of processor can be increased due to NCFET? This, in turn, enables the technology to achieve further improvements in performance when scaling, without exacerbating the heat generation problem.
(2) Reviving the prior trend of voltage decrease in processors without decreasing performance. In other words, under the same performance constraint, how much the voltage of processor can be decreased due to NCFET? This, in turn, enables the technology to achieve further improvements in power when scaling, without degrading performance.
(3) Continuation of Moore's law when reaching the fundamental limit of scaling. In other words, how alterations in the NCFET technology (e.g., increasing the thickness of the ferroelectric) may provide further improvements in the processor's performance without the need to scaling down the transistor dimensions.

II. NCFET Effects from Physics to Processor Level

In the following, we first explain briefly how NCFET technology works and then we describe how their effects at the physical and device levels can be modeled. Afterwards, we discuss the different case studies that our work investigates. Finally, we demonstrate how NCFET-aware cell libraries can be created towards enabling designers to analyze the effects that NCFET technology has on the performance and power of complex circuits like processors.

A. NCFET Basics

NCFET incorporates a ferroelectric material within the gate stack of transistor (see Fig. 3(a)), which exhibits a negative capacitance (NC) under certain constraints. The metal-ferroelectric-metal-insulator-semiconductor (MFMIS) arrangement in NCFET can be equivalently represented by a capacitor divider circuit of the ferroelectric capacitance C_{fe} and the internal baseline FET capacitance C_{int} as shown in Fig. 3(b). A negative value of C_{fe} results in voltage amplification at the internal gate, and can be expressed through the following equation:

$$A_V = \frac{|C_{\text{fe}}|}{|C_{\text{fe}}| - C_{\text{int}}} \tag{1}$$

This internal voltage amplification is responsible for suppressing SS beyond the Boltzmann limit (i.e., achieving a sub-60 mV/decade) [9]. Additionally, this also results in higher ON current in NCFET compared with the baseline FET at the same applied voltage [11]. Thermodynamically, any ferroelectric system in equilibrium can be described by a double well energy profile at zero applied electric field, as shown in Fig. 3(c). The energy profile is characterized by three extrema states; two with positive curvature or positive capacitance minima and one with a negative curvature or negative capacitance (NC) maximum (note that capacitance $= (\partial^2 G/\partial P^2)^{-1}$). Such an energy profile in thermodynamics is generally modeled by a Taylor series expansion of G in even powers of polarization known as the Landau-Khalatnikov (L-K) theory [37], [38]. The system always occupies either of the two states of minimum energy corresponding to a non-zero/remnant polarization at zero bias. The NC state corresponding to the energy maximum is unstable and, therefore, inaccessible in an isolated ferroelectric capacitor.

S. Salahuddin and S. Datta in their pioneering work [9] suggested that an unstable NC state can normally be stabilized by connecting a positive dielectric capacitance in series with the ferroelectric, such that the total capacitance of the system becomes positive. They also illustrated this using the L-K model of ferroelectrics. This method ensures a stable minimum

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/ACCESS.2018.2870916, IEEE Access

in the total energy of the system (at $P = 0$), in which the ferroelectric lies in NC state as shown in Fig. 3(c). In an NCFET structure, the internal baseline FET acts as a positive capacitance ($C_{\rm int}$) (see Fig. 3(b)), which can stabilize the ferroelectric in NC state, if the total gate capacitance $= ({C_{\rm int}}^{-1} + {C_{\rm fe}}^{-1})^{-1}$ is positive. This condition implies that $|C_{\rm fe}|$ must be greater than $C_{\rm int}$ to realize a thermodynamically stable NC state of the ferroelectric in NCFET. Note that in general, both the $|C_{\rm fe}|$ and $C_{\rm int}$ can be function of applied voltages (gate voltage V_G and drain voltage V_D). Violation of the above condition, for any combination of V_G and V_D results in the hysteresis behavior in device's electrical characteristics [39]. To avoid such violation, the thickness of ferroelectric should be limited because $|C_{\rm fe}|$ varies inversely with t_{fe}. Additionally, $|C_{fe}|$ also depends on the remnant polarization (P_r) and coercive field (E_c) values of the ferroelectric. The ferroelectric with a high remnant polarization (P_r) and low coercive field (E_c) also leads to an increase in the value of $|C_{fe}|$ [12]. This ensures that the condition of no hysteresis, i.e. $|C_{\rm fe}| > C_{\rm int}$, is always met in the range of operating voltages. Note that a low t_{fe}, high P_r and low E_c, all reduce capacitance matching between C_{fe} and C_{int}, thus also the ferroelectric gain according to Eq. 1. Thus, their optimum values should be carefully chosen, such that it provides a larger gain (due to NC effect), but does not lead to hysteresis behavior, which is undesirable for the logic applications.

Physical origin of NC: Many recent experimental works employing a resistor-ferroelectric system [40]–[42] have indicated that the NC effect is actually a transient phenomenon which appears due to insufficient screening of bound polarization charges by the free charges delivered from the voltage source during polarization switching [43]. [44] and [45] have explained the emergence of transient negative capacitance in resistor-ferroelectric and ferroelectric-dielectric (FE-DE) systems using the previous condition in conjugation with L-K theory [37]. However, some other works [46]–[48] based on different classical ferroelectric models have provided alternative explanations and argued that the negative capacitance may only be a result of dynamics of ferroelectric domain nucleation and growth during polarization switching and not an intrinsic effect based on the thermodynamic L-K theory. Another work has attributed the total energy minimization in FE-DE systems at $P = 0$ under a quasi-static scenario (Fig. 3(c)) to a strong coupling between FE and DE polarizations [49], which may force the whole system act like a dielectric with a positive capacitance behavior [49]. The real physical origin of NC in ferroelectrics is still under intense research.

B. Physical Level: Model Description

In this work, we consider an NC-FinFET that has configuration of MFMIS as shown in Fig. 3(a). The presence of equipotential intermediate metal between the ferroelectric and the oxide insulator, enables the partitioning of full structure into two circuit components: the ferroelectric capacitor and the internal baseline FinFET. The baseline FinFET part of the device is modeled by the industry standard BSIM-CMG model, which accurately captures both the short channel and quantum mechanical effects in small geometries [50], [51]. We employ the open-source $7\,\rm nm$ predictive Process Design Kit (PDK) [52] for the FinFET model parameters. The ferroelectric part is modeled using L-K equation[1] given by: $V_{\rm fe} = t_{\rm fe}(2\alpha Q + 4\beta Q^3)$, where $V_{\rm fe}$ is the voltage across ferroelectric, Q is the gate terminal charge per unit area, $t_{\rm fe}$ is the ferroelectric thickness and α and β are ferroelectric material specific parameters [9], [37]. We adopt Al doped HfO_2 as the ferroelectric, which is a CMOS technology process compatible material [21]. The experimental remnant polarization $= 0.05$ μC/cm^2 and coercive field $= 1$ MV/cm values of this material, obtained from [21], are used to calculate α and β parameters [53]. Finally, for the full device simulation, the L-K equation is solved in a self-consistent manner inside the BSIM-CMG model [51] of FinFET in Verilog-A code (see Fig. 3(d)), using a commercial SPICE simulator [54].

Note that we do not consider in this work non-idealities related to different SS in forward and reverse sweeps [42], [55] because the employed structure of NCFET here is MFMIS with defects free top and bottom metal electrodes of the ferroelectric with a single domain configuration. Additionally, we have used ferroelectric layers with relatively small thicknesses (i.e., maximum of 4 nm) in which a hysteresis free operation, which is essential for CMOS logic applications, is ensured.

Role of internal metal gate: Theoretically, internal gate may provide a larger ON current than the NCFET without this gate except for very low P_r ferroelectrics as we have shown in [56]. The presence of internal gate also makes the NCFET device less susceptible to hysteresis [56]. However, some other works have shown that the existence of internal metal gate causes some other issues like destabilization of the negative capacitance state in the presence of ferroelectric leakage current [57] or domain formation [58]. Moreover, the presence of internal metal gate can also contribute as an additional source of traps which may cause the device variability of NCFET with the internal gate to be larger than without internal gate [59].

C. Investigated Case Studies

We consider four different NC-FinFETs corresponding to four different ferroelectric thicknesses ($t_{\rm fe}$) of $1\,\rm nm$, $2\,\rm nm$, $3\,\rm nm$ and $4\,\rm nm$; termed as: TFE1, TFE2, TFE3 and TFE4, respectively. Our main focus is to investigate, how different NC-FinFET technologies (i.e., alternations in the NCFET technology itself) can ultimately impact performance and power of the processor compared to the *baseline*, which is the original/conventional $7\,\rm nm$ FinFET, that does not include any ferroelectric material. We refer to the baseline case with TFE0. For fair comparisons, we calibrate the four NC-FinFET devices (TFE1, TFE2, TFE3 and TFE4) through gate work-function tuning in order to acquire iso-$I_{\rm OFF}$ condition i.e. the same OFF current as that of the baseline $7\,\rm nm$ FinFET

[1]Note that the dielectric constant or more precisely the behavior of polarization with the applied electric field takes the higher dielectric constant of the ferroelectric into account.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/ACCESS.2018.2870916, IEEE Access

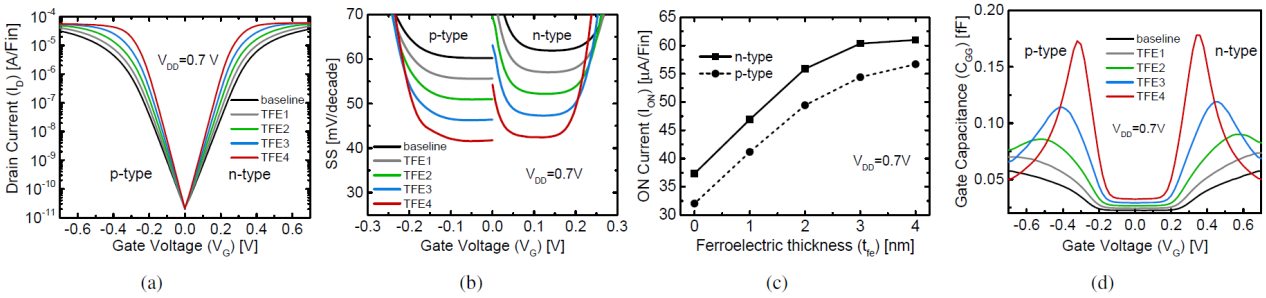


Fig. 4. (a) Transfer characteristics of n-type and p-type NC-FinFET for different ferroelectric thicknesses ($t_{\rm fe}$) at iso-$I_{\rm OFF}$ condition. The $I_{\rm OFF}$ of both types of NC-FinFETs are matched to the corresponding baseline 7nm node FinFETs by work function tuning. (b) sub-threshold swing as a function of gate voltage for different $t_{\rm fe}$ values. (c) ON current variation as a function of $t_{\rm fe}$. (d) Gate terminal capacitance as a function of gate bias for different $t_{\rm fe}$ values.

node (TFE0) at the supply voltage, $V_{DD} = 0.7\,\mathrm{V}$, which is the nominal voltage provided in the PDK.

The electrical characteristics of NC-FinFET and their comparison with those of baseline FinFET are shown in Fig. 4. The voltage amplification due to negative capacitance effect, at the internal gate in NC-FinFET, clearly results in lower sub-threshold swing and higher ON current in comparison to baseline FinFET, as demonstrated in Figs. (4(b) and 4(c)). It can be seen that these quantities further improve with increase in the ferroelectric thickness ($t_{\rm fe}$), due to an improved capacitance matching between the baseline capacitance ($C_{\rm int}$) and ferroelectric capacitance ($C_{\rm fe}$), which consequently, provides higher voltage amplification [10], [39]. However, when the gate voltage (V_G) is such that $C_{\rm fe}$ is negative, the total gate terminal capacitance, $C_{\rm GG}$ exceeds $C_{\rm int}$ (conventional dielectric layer) and exhibits sharp peaks as shown in Fig. 4(d). Also, $C_{\rm GG}$ rises with increase in $t_{\rm fe}$, similar to $I_{\rm ON}$. Hence, in NCFET, there always exists a trade-off between conflicting $I_{\rm ON}$ and $C_{\rm GG}$ requirements, to achieve better performance gain by increasing $t_{\rm fe}$. However, in our analysis, we find that the gain in $I_{\rm ON}$, in general, more than compensates the undesirable increment in $C_{\rm GG}$. The performance of processor typically improves with increase in $t_{\rm fe}$, at least until TFE3 (i.e. $t_{\rm fe} = 3\,\mathrm{nm}$). Above that ferroelectric thickness (i.e. at TFE4), the compensation effect of $C_{\rm GG}$ can diminish the performance improvement at high V_{DD}. Further details and discussion are presented later in Section IV and demonstrated in Fig. 10(a).

D. NCFET-Aware Cell Libraries

After incorporating the physical effects that ferroelectric layer causes in the $7\,\mathrm{nm}$ p-FinFET and n-FinFET devices, we characterize the $7\,\mathrm{nm}$ standard cell library [52] for the five studied cases (i.e., TFE0, TFE1, TFE2, TFE3 and TFE4). Each case is analyzed starting from the nominal operating voltage ($V_{DD} = 0.7\,\mathrm{V}$) to the minimum voltage ($V_{DD} = 0.2\,\mathrm{V}$) using a commercial tool for cell library characterization [60]. The tool takes the post-layout SPICE netlists of standard cells including the parasitics information as input and calculates the delay and the power information of every cell at the targeted scenario (i.e., V_{DD} and $t_{\rm fe}$ values), using a commercial SPICE simulator [54]. For the library characterization process, we employ our developed Verilog-A FinFET model, which takes into account the negative capacitance effect as explained earlier, along with the modified transistor modelcard, which contains the baseline FinFET transistor's parameters as well as the parameters that are additionally required for the ferroelectric layer, e.g., ferroelectric thickness ($t_{\rm fe}$), remnant polarization (P_r) and coercive field (E_c). Every characterized cell library consists of 121 different combinational and sequential cells, covering a wide range of cell complexity, due to the different typologies of cells (i.e., pull-up and pull-down transistor networks inside every cell), starting from an inverter, which has only two transistors to a D-FF cell, which has 26 transistors. Every cell is analyzed under different operating conditions; 7×7 input signal slews and output load capacitances. The selected ranges of operating conditions are obtained from the baseline standard cell library and scaled when considering low V_{DD} scenarios. At the end, all the calculated delay and power information of cells are stored using the standard "liberty" format, which makes the obtained NCFET-aware cell libraries compatible with the existing commercial EDA tools for circuits like logic synthesis and timing/power analysis tool flows [61]. This enables us to integrate our NCFET-aware cell libraries within the standard process flow, to evaluate the overall impact of NCFET on the performance and the power of processors. Figs. 5(b and c) summarize the method of creating NCFET-aware cell libraries and how they can be later employed to evaluate the power and the performance of a processor.

III. Effects of NCFET on Standard Cells

As explained earlier, the NCFET-aware cell libraries contain the delay and power information of every cell under the ferroelectric effects for different operating voltages. This enables us to quantify, how the NCFET technology impacts the delay, power, and capacitance of every gate. In Fig. 6, we demonstrate the delay distribution of gates in the $7\,\mathrm{nm}$ FinFET cell library, for five different analyzed cases: TFE0, TFE1, TFE2, TFE3 and TFE4. It can be noticed that, with increase in the thickness of ferroelectric layer, the delay distribution gets shifted towards the left side. Hence, the delay of gates at the same voltage ($V_{DD} = 0.7\,\mathrm{V}$) becomes smaller, which means, NCFET boosts the delay of gates due to the amplification of gate voltage. To demonstrate the relative delay improvement, we show in Fig. 7, the reduction in delay across the standard

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/ACCESS.2018.2870916, IEEE Access

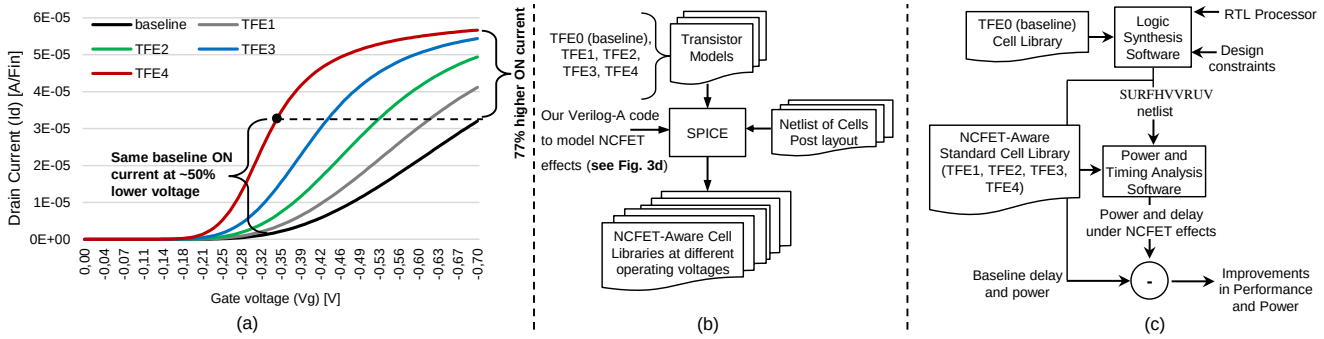


Fig. 5. (a) The I-V characteristics of pFinFET under the effects of negative capacitance at different thickness of ferroelectric. As shown, the same ON current can be reached a lower V_{DD} when NC-FinFET is employed. Increasing the thickness of ferroelectric layer allows further reduction in the operating voltage while still achieving the same $I_{\rm ON}$. (b) A general flowchart demonstrating the cell library characterization for creating NCFET-aware cell libraries at different voltages and different thicknesses of ferroelectric. (c) A general flowchart demonstrating, how the created NCFET-aware cell libraries are later used in order to analyze the performance and power of processor, towards evaluating the overall impact of NCFET technology at the processor level.

cell library for the four NCFET technologies relative to the baseline technology. The delay improvement due to NCFET becomes even better at lower voltages. As shown in Fig. 8, for the case of $V_{DD} = 0.4\,\mathrm{V}$, the NCFET technology can be provide delay improvement by around 40%, 60%, 80% and 90%, on average, for TFE1, TFE2, TFE3 and TFE4, respectively. This improvement is on expected lines, because the improvement in ON current due to NCFET, relative to the baseline, becomes higher at lower voltages, as can be noticed in Fig. 5(a).

It is noteworthy that adding a ferroelectric layer to the gate stack of transistor will directly increase its gate capacitance, since the ferroelectric layer exhibits a negative capacitance (see Fig. 3(b)). Therefore, the total capacitance of gate will increase due to NCFET, as demonstrated in Fig. 4(d). This of course can have a diminishing impact on the performance gain achieved by NCFET, because gates will exhibit larger capacitance to charge. In Fig. 9(a), we present the relative increase in the cells' capacitance due to NCFET, for the different ferroelectric thicknesses. For instance, the average increase in the cells' capacitance reaches around 20% and 70% when $t_{\rm fe}$ is $1\,\mathrm{nm}$ and $4\,\mathrm{nm}$, respectively, for the case of $V_{DD} = 0.7\,\mathrm{V}$. Since, NCFET provides a higher drive current and also leads to increase in the cell's capacitance, the internal power of cell will definitely increase. This is indicated in Fig. 9(b), which demonstrates the average increase in the power of cells due to NCFET for $V_{DD} = 0.7\,\mathrm{V}$ and $V_{DD} = 0.4\,\mathrm{V}$. It is noteworthy that regardless of the increase in the cells' power, at the end of the day, NCFET will still provide a considerable power saving for circuits compared to the baseline. This is because the same baseline performance (i.e., frequency) can be achieved at a lower voltage, which leads to quadratic saving in dynamic power ($P_{dynamic} \propto Freq \times V_{DD}^2$) and exponential saving in stand-by power, thus, compensating the side effect of NCFET with respect to power.

IV. Effects of NCFET on Future Processor Design

Since, the created NCFET-aware cell libraries are compatible with the existing commercial EDA tool flows (e.g., [61]), we can employ these tools directly without any modification for investigating the impact of NCFET technology on the performance and the power of circuits. Because our main focus is to understand how the NCFET technology can reshape the trends in processor design, we analyzed the Berkeley Out-of-Order Machine (BOOM), which is an open-source industry-competitive processor, based on the recent open source RISC-V instruction set architecture (ISA) [62]. The processor is designed to serve as the prototypical baseline processor for future micro-architectural studies for use in education, research, and industry [63], [64]. In our analysis, we employed the default settings of the BOOM processor when generating its register-transfer level (RTL) for the hardware description. As demonstrated in Fig. 5(c), we first synthesize the RTL of processor using the original $7\,\mathrm{nm}$ FinFET cell library. This represents our baseline to compare against. Then, we analyze the performance and the power of processor using the generated gate-level netlist, and our created NCFET-aware cell libraries, for the four different NCFET technologies. This enables us to evaluate, how the gate voltage amplification (originating from the physical level, due to the integration of ferroelectric layer within the gate stack of pFinFET and nFinFET transistors) will finally impact the overall performance and power at the processor level.

A. Impact of NCFET on Computing Efficiency

To investigate the potential impact of NCFET on the processor design trends, we consider the following three scenarios: **(1) Impact of NCFET on increasing the processor's frequency at the same V_{DD} (i.e. performance improvement without voltage increase):** As explained earlier, the NCFET technology provides voltage amplification for the transistor's gate caused by the integrated ferroelectric material, and this leads to higher ON currents in pFinFET and nFinFET transistors at the same voltage as shown in Figs. (4(a) and 5(a)). Since, higher ON current results in higher speed in standard cells, as shown in Figs. (6, 7 and 8), the maximum delay of processor becomes smaller, due to NCFET. In Fig. 10(a), we show the frequency increase relative to the baseline for two different V_{DD}, the nominal voltage ($0.7\,\mathrm{V}$), and a near-threshold voltage ($0.4\,\mathrm{V}$). It can be noticed that, NCFET

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/ACCESS.2018.2870916, IEEE Access

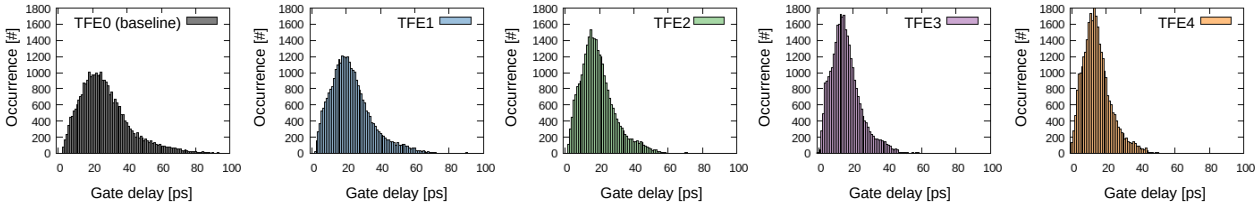


Fig. 6. Evaluating the impact of NCFET on the delay of cells within the 7 nm FinFET standard cell library. Increasing the thickness of the ferroelectric layer provides a larger gate voltage amplification in transistors (i.e., steeper slope and higher ON current). Hence, the delay of cells become smaller.

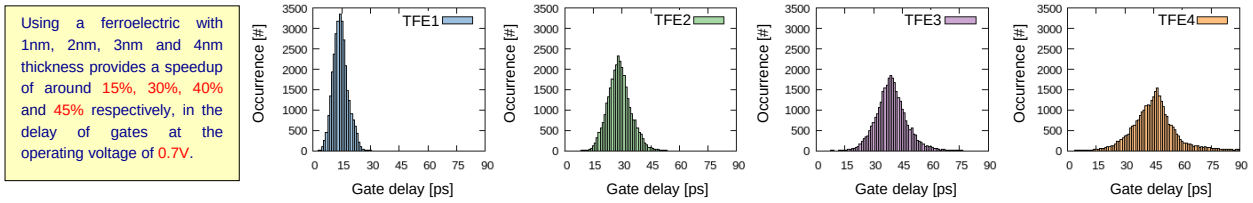


Fig. 7. Quantifying the relative delay decrease/improvement of cells within the 7 nm FinFET standard cell library due to NCFET at $V_{DD} = 0.7\,\mathrm{V}$.

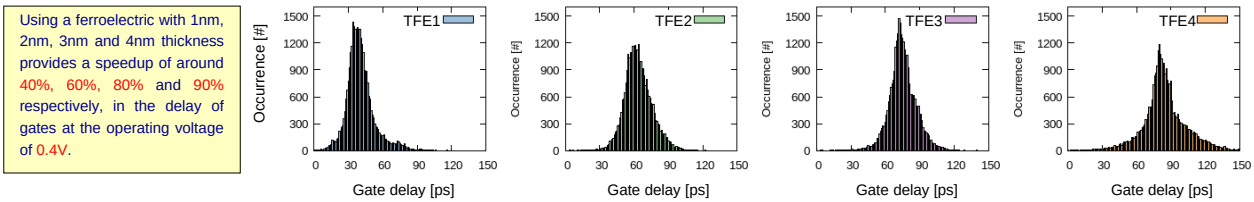


Fig. 8. Quantifying the relative delay decrease/improvement of gates within the 7 nm FinFET standard cell library due to NCFET at $V_{DD} = 0.4\,\mathrm{V}$.

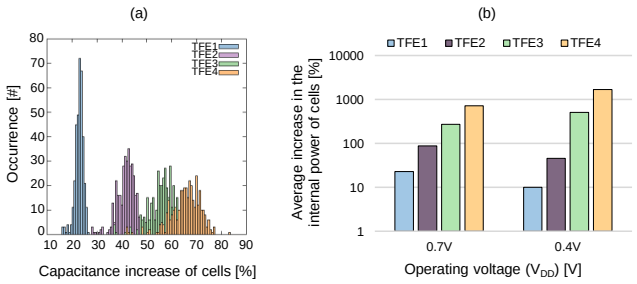


Fig. 9. (a) Evaluating the impact of NCFET on increasing the total capacitance of cells at $V_{DD} = 0.7\,\mathrm{V}$. As expected increasing the thickness of ferroelectric layer leads to an increase in the total cells' capacitance. (b) Evaluating the impact of NCFET on increasing the internal power of cells at $V_{DD} = 0.7\,\mathrm{V}$ and $0.4\,\mathrm{V}$. The average increase among all cells is used in this analysis.

always provides speedup in the processor's performance. The speedup becomes larger with increase in the thickness of ferroelectric layer due to the higher voltage amplification. At lower voltage (i.e., $V_{DD} = 0.4\,\mathrm{V}$), the achieved speedup becomes even higher, which is consistent with the results presented earlier in Fig. 8. It is noteworthy that, unlike the case of $V_{DD} = 0.4\,\mathrm{V}$, increasing thickness of the ferroelectric layer from 3nm to 4nm in the case of $V_{DD} = 0.7\,\mathrm{V}$, does not lead to any further improvement in the performance. This is due to the compensation effect related to the increase in the cells' capacitance caused by the negative capacitance, as discussed earlier in Section II-C and shown in Figs. (4(d) and 9(a)). These trends are also consistent with the previous works on NCFET based digital circuits [65], [66]. Hence, selecting the optimal thickness of ferroelectric layer, such that the performance gain is maximized, is not trivial and it depends on several aspects, including the operating voltage.

(2) Impact of NCFET technology on increasing the processor's frequency within the baseline power density constraint: As explained earlier, in the introduction, due to the non-ideal voltage scaling, the frequency of processors had stopped increasing in the last decade, to avoid unsustainable on-chip power densities. In the following analysis, we investigate whether the NCFET technology will enable processor designers to increase the processor's frequency, without violating the power density constraint. The latter is considered as the baseline power density of the processor, which is at the nominal voltage ($V_{DD} = 0.7\,\mathrm{V}$) when no NCFET is employed. As a matter of fact, NCFET allows the pFinFET and nFinFET to have smaller sub-threshold swing and higher ON current than the baseline at the same voltage. Hence, the same ON current as in the baseline case, can be achieved at a lower voltage (see Fig. 5(a)). Therefore, under the same power density constraint as in the baseline case, there is prospect to operate the processor at a higher frequency. In Fig. 10(b), we demonstrate the speedup in processor performance for the four NCFET technologies. As shown in the figure, by increasing the thickness of ferroelectric, higher speedup can be achieved. The figure also shows, the corresponding operating voltage for every NCFET technology (i.e., TFE1, TFE2, TFE3 and TFE4) where the baseline power density has been met.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/ACCESS.2018.2870916, IEEE Access

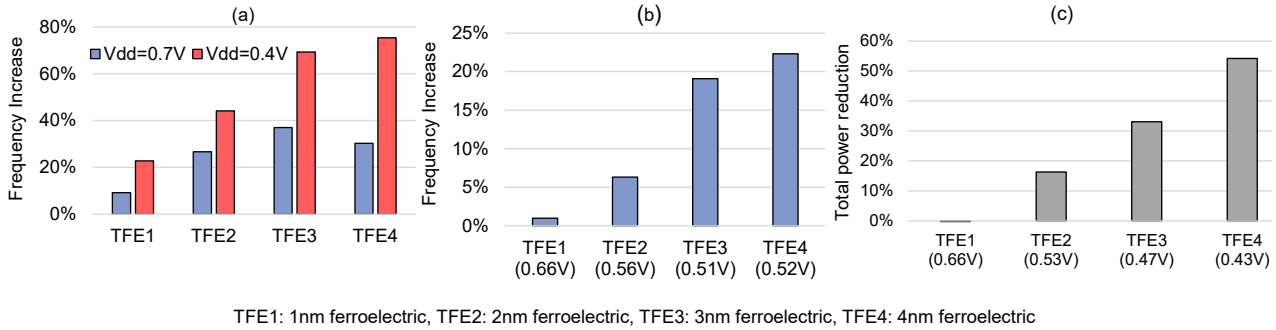


Fig. 10. Analyzing the impact of NCFET technology at the processor level. (a) Addressing the question of what is the frequency increase due to NCFET under the same voltage constraint? (b) Addressing the question of what is the frequency increase under the same (i.e., baseline) power density constraint? The baseline refers to the power density of processor when no NCFET is employed at the nominal voltage ($V_{DD} = 0.7\,\mathrm{V}$). (c) Addressing the question of what is the minimum operating voltage along with the achieved power reduction under the same (i.e., baseline) performance (i.e., frequency) constraint? The baseline refers to the frequency of processor when no NCFET is employed at the nominal voltage of $0.7\,\mathrm{V}$.

(3) Impact of NCFET technology on decreasing the processor's voltage under baseline frequency constraint: Last but not least, we investigate what is the minimum operating voltage at which the baseline performance (i.e., the baseline processor's frequency at nominal voltage when no NCFET is employed) can be achieved. Reduction in V_{DD} directly leads to decrease in the overall power of processor, as both static and dynamic powers get reduced at lower voltages [51]. In Fig. 10(c), we show, how the NCFET provides a noticeable reduction in the processor's power for the same performance. As can be noticed, the reduction reaches up to 54% when the thickness of ferroelectric layer is 4nm. Increasing the ferroelectric thickness leads to higher gate voltage amplification. Hence, the same ON current can be achieved at a lower voltage. As shown in the figure, in case of TFE3 and TFE4, the processor can operate at lower than the nominal voltage (i.e., $0.47\,\mathrm{V}$ and $0.43\,\mathrm{V}$, respectively) while the baseline processor's frequency is still met. This, in turn, leads to 33% and 54% reductions in processor's power, respectively. Note that from this analysis, it becomes apparent that NCFET can still provide a considerable power reduction despite the increase in power of cells shown in Fig. 9(b), because it allows the processor to operate at a lower voltage, while still achieving the same baseline performance.

B. Rewarding Effects of NCFET

In order to put the achieved gains due to NCFET in context, now we discuss, how reductions in power density due to NCFET will lead to reductions in temperature as well as the role that NCFET may play in the Internet of Things (IoTs) domain, where power budgets are very restricted.

(1) Temperature Reduction: Since, processors nowadays are thermally constrained [5], we analyze further the positive consequences of reducing the total power of processor due to NCFET (see Fig. 10(c)) on temperature. Since, the processor area is the same in all the analyzed cases, the reduction in total power due to NCFET is directly related to the reduction in power density of processor. Hence, NCFET will, in principle, lead to lower operating temperatures. To evaluate this, we use a commercial multiphysics simulation software [67], in order to accurately calculate the maximum temperature of chip for the five different power densities, resulting from the four studied NCFET cases beside the original (i.e., baseline) case. Fig. 11(a) shows an overview of the simulated system. It consists of a chip on top of a printed circuit board (PCB) and under a metal heatsink. Fig. 11(b) summarizes the resulting maximum temperature for the different cases. As shown in the figure, NCFET with ferroelectric thickness more than $1\,\mathrm{nm}$ leads to a noticeable temperature reduction, due to the decrease in the on-chip power density. For the case of TFE4, the temperature reduction reaches 62%.

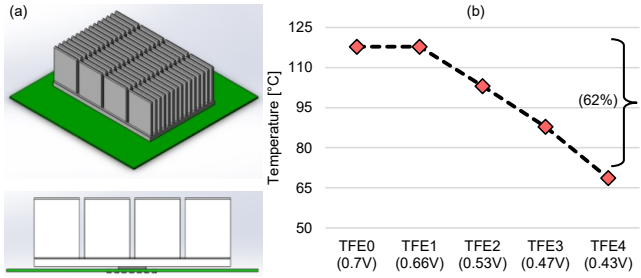


Fig. 11. Evaluating the corresponding impact of power density reduction due to NCFET on reducing the maximum temperature of chip. (a) shows the built system in a multiphysics simulation software to conduct the required experiments. The system consists of a silicon chip that is located between a PCB and a heatsink. (b) shows the resulting reduction in temperature for every studied NCFET technology case.

Implementation details: In the aforementioned analysis, the size of heatsink is $7.8 \times 5.2\,\mathrm{cm}$ with a height of $2.7\,\mathrm{cm}$. The heatsink consists of several fins providing a total surface area of $64133.7\mathrm{mm}^2$. The heatsink material is aluminum with a thermal conductivity of $237.5\,\mathrm{W/mK}$. We consider a thermal transfer coefficient of $70\mathrm{W/m}^2\mathrm{K}$, as in the forced convection of air [68], which provides a temperature below $120°\mathrm{C}$ at the maximum simulated power density (i.e., the baseline power density). The material of simulated chip is silicon anisotropic, with thermal conductivity of $124\mathrm{W/mK}$, and the thermal conductivity of PCB is $1.4\mathrm{W/mK}$. The properties of all used materials have been obtained from the engineering data, which is available in the used multiphysics software [67]. The

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/ACCESS.2018.2870916, IEEE Access

ambient temperature of 27°C was considered in all analyzed cases. It is noteworthy that operation of circuits at lower voltage along with lower temperature can have a considerable impact on their lifetime. This is because the majority of the underlying mechanisms behind reliability degradation, e.g. Bias Temperature Instability (BTI) aging phenomenon, are stimulated by higher voltage and temperature [69]. Therefore, it is also expected that traditional aging mechanisms (e.g., BTI) will be significantly mitigated in NCFET, due to the reduction in voltage and temperature.

(2) Enabling Internet-of-Thing (IoT) Devices: IoTs has recently emerged as an attractive paradigm in which everything and everyone might be connected to internet. IoTs aim at providing variety of services such as sensing and processing data, exchanging information with other IoT devices as well as with cloud for making decisions. One of the key challenges to enable IoTs is that the vast majority of IoT devices may not be possibly powered by batteries, due to the difficulty of replacing them. Therefore, power sources through harvesting energy such as body heat, could be one of the attractive alternatives – especially for wearable medical IoT devices. However, the harvested energy from such resources is typically very low and, hence, not sufficient to operate the processor at a suitable frequency to meet required performance constraints.

Since, NCFET technology is a good candidate for ultra-low power applications, we now investigate, how much the NCFET technology will provide performance improvement for the processor, when it is operated under restricted power budgets which are harvested from a human body heat. For this purpose, we built, using a commercial multiphysics simulation software [67], a thermoelectric generator (TEG) that consists of 128 N-P couples within an area of $3\times3\,\mathrm{cm}$. The TEG device converts the difference in temperature into electrical power due to the Seebeck effect [70]. To harvest energy from the body heat, the TEG with a small heatsink above it can be located on top of the wrist of a person. This TEG converts the temperature difference between the body skin and the ambient temperature into power. In our analysis, we consider four different ambient temperatures ranging from 0°C to 25°C. In every case, we simulate the heat flux that corresponds to the generated heat from human body at that ambient temperature. The heat flux values haven been obtained from the experimental work in [71]. From the multiphysics simulation, we extract the output voltage of the TEG for the open circuit (V_{OC}). Then, the power generated by connecting a load to the TEG, is calculated using: $P_L = \dfrac{V_{OC}^2 R_L}{(R_{TEG}+R_L)^2}$. Here, R_{TEG} is the internal resistance of TEG and R_L is the resistance of load. The maximum power point occurs when R_L and R_{TEG} are equal. In Fig. 12(a), we show the maximum generated power for different ambient temperatures. For instance, when the ambient temperature is 15°C, the total power generated by the studied $3\times3\,\mathrm{cm}$ TEG device is $232\,\mu\mathrm{W}$.

In Fig. 12(b), the improvement in the processor's performance due to NCFET is shown for every power budget that corresponds to the maximum generated power in every ambient temperature case. Unlike the previous analysis in Fig. 10 in which a high-performance processor was studied,

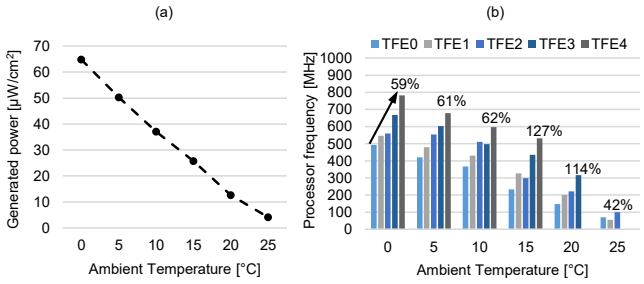


Fig. 12. (a) Maximum harvested power from body heat using a thermoelectric generator (TEG) at different ambient temperatures. Experiments are done using a multiphysics simulation software. The TEG's area is $9\,\mathrm{cm}^2$. (b) The impact of NCFET technology on improving the performance of an IoT processor, at different power budgets that correspond to the harvested energy from the body heat at different ambient temperatures.

a much simpler processor, which suits more the purpose of IoT devices, has to be considered in this analysis. To this end, we analyze the open-source 32-bit PULPino processor, which has been developed to satisfy the demands of IoT applications [72]. This processor is based on the Instruction Set Architecture of RISC-V, similar to the BOOM processor. As shown in Fig. 12(a), higher the ambient temperature is, smaller the available power budget becomes. Hence, the processor needs to operate at a low V_{DD} to fulfill the ultra-low power budget, which, in turn, leads to considerable performance loss, due to the very low operating frequency. However, for all different ambient temperatures and thus power budgets, NCFET provides a noticeable improvement in the processor's frequency. For instance, when the ambient temperature is 10°C, the NCFET with $4\,\mathrm{nm}$ ferroelectric thickness, can boost the processor frequency by 62%. For a smaller power budget, when the ambient temperature is 20°C, the NCFET with a $3\,\mathrm{nm}$ ferroelectric layer can increase the processor's performance by more than 110%. In addition, assuming that the IoT device has a performance requirement of $500\,\mathrm{MHz}$, the baseline technology fails to meet such a constraint even at the lowest ambient temperature (0°C), where the harvested power is at maximum. On the other hand, the NCFET technology allows the IoT processor to meet this frequency constraint until the ambient temperature reaches near 15°C. Note that when the power budgets become very small (e.g., when the ambient temperature goes above 15°C), the NCFET technology even with thicker ferroelectric layer (i.e., TFE3, TFE4) fails to fulfill such power budgets. This is mainly due to the increase in the cells' capacitance when the ferroelectric thickness becomes large, as shown in Figs. (4(d), 9(a) and 9(b)). This, in turn, leads to increasing the power consumption of processor beyond the small available power budgets. Note that efficiency loss due to connected circuitries to TEG (e.g., DC-DC converter) is not taken into account in this analysis. However, such loss might be compensated by generating more power through e.g., increasing the size of TEG or adding a second TEG device on the bottom side of the wrist of a person.

Implementation details: The simulated TEG device has a Seebeck coefficient of $200\,\mathrm{V}/^\circ\mathrm{C}$ [73] and a total internal resistance of $4.2\,\Omega$. The size of heatsink is $3\times3\,\mathrm{cm}$ matching

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/ACCESS.2018.2870916, IEEE Access

the TEG surface with a thickness of merely $0.1\,\mathrm{cm}$. The heatsink material is aluminum with a thermal conductivity of $237.5\,\mathrm{W/mK}$. We consider the thermal transfer coefficient to be $5\,\mathrm{W/m^2K}$, as in the natural free convection of air [68]. The heatsink has 10 fins, which provides a total surface area of $6528\,\mathrm{mm}^2$. The thickness of the human skin is $1.3\,\mathrm{mm}$ and it thermal conductivity is $0.209\,\mathrm{W/mK}$, based on [71], [74].

V. Conclusion

In this work, we have analyzed the impact of NCFET technology on processor performance, based on the $7\,\mathrm{nm}$ FinFET. We have demonstrated that the NCFET can provide a quantum leap in computing efficiency, because it enables processors to operate at higher frequencies, without increasing the operating voltage. Additionally, NCFET can revive the prior trend of frequency, which stopped a decade ago due to the irrepressible increase in power densities, because it allows the processor to operate at a higher frequency compared to the conventional $7\,\mathrm{nm}$ FinFET, without increasing its power density. We have also shown, how NCFET can lead to a considerable reduction in power density, and hence, temperature, as it allows the processor to operate at a lower voltage while still be clocked at the same frequency as in the the conventional FinFET. Finally, for ultra-low power applications, such as in IoTs, NCFET can play a major role. We have shown that under very small power budgets harvested from body heat, NCFET can noticeably boost the the processor's frequency compared to the conventional FinFET technology.

Acknowledgment

Authors at KIT would like to thank Hammam Kattan for his valuable efforts in performing the multiphysics simulations.

References

[1] "Karl Rupp: 40 Years of Microprocessor Trend Data," https://www.karlrupp.net/tag/processors/.

[2] R. H. Dennard, F. H. Gaensslen *et al.*, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, 1974.

[3] G. Moore, "Cramming more components onto integrated circuits," *Electronics Magazine*, vol. 38, no. 8, 1965.

[4] L. B. Kish, "End of moore's law: thermal (noise) death of integration in micro and nano electronics," *Physics Letters A*, vol. 305, no. 3, pp. 144–149, 2002.

[5] M. M. Waldrop, "The chips are down for Moore's law." *Nature*, vol. 530, no. 7589, pp. 144–147, 2016.

[6] A. Huang, "Moore's law is dying (and that could be good)," *IEEE Spectrum*, vol. 52, no. 4, pp. 43–47, April 2015.

[7] J. D. Meindl, Q. Chen *et al.*, "Limits on silicon nanoelectronics for terascale integration," *Science*, vol. 293, no. 5537, pp. 2044–2049, 2001.

[8] V. V. Zhirnov and R. K. Cavin, "Nanoelectronics: Negative capacitance to the rescue?" *Nature Nanotechnology*, vol. 3, no. 2, pp. 77–78, 2008.

[9] S. Salahuddin and S. Datta, "Use of negative capacitance to provide voltage amplification for low power nanoscale devices," *Nano Letters*, vol. 8, no. 2, pp. 405–410, 2008.

[10] A. Khan, C. Yeung *et al.*, "Ferroelectric negative capacitance MOSFET: Capacitance tuning and antiferroelectric operation," in *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2011, pp. 11.3.1–11.3.4.

[11] G. Pahwa, T. Dutta *et al.*, "Analysis and compact modeling of negative capacitance transistor with high ON-current and negative output differential resistance – Part II: Model validation," *IEEE Transactions on Electron Devices*, vol. 63, no. 12, pp. 4986–4992, Dec. 2016.

[12] G. Pahwa, T. Dutta *et al.*, "Designing energy efficient and hysteresis free negative capacitance FinFET with negative DIBL and 3.5X I_{ON} using compact modeling approach," in *European Solid-State Device Research Conference (ESSDERC)*, Sept 2016, pp. 41–46.

[13] H. Ota, T. Ikegami *et al.*, "Fully coupled 3-D device simulation of negative capacitance finfets for sub 10 nm integration," in *2016 IEEE International Electron Devices Meeting (IEDM)*, Dec 2016, pp. 12.4.1–12.4.4.

[14] S. K. Samal, S. Khandelwal *et al.*, "Full chip power benefits with negative capacitance FETs," in *Low Power Electronics and Design (ISLPED, 2017 IEEE/ACM International Symposium on*, 2017, pp. 1–6.

[15] T. Dutta, G. Pahwa *et al.*, "Performance evaluation of 7-nm node negative capacitance FinFET-based SRAM," *IEEE Electron Device Letters*, vol. 38, no. 8, pp. 1161–1164, Aug. 2017.

[16] X. Li, J. Sampson *et al.*, "Enabling energy-efficient nonvolatile computing with negative capacitance FET," *IEEE Transactions on Electron Devices*, vol. PP, no. 99, pp. 1–7, Jun. 2017.

[17] J. Zhou, G. Han *et al.*, "Ferroelectric $HfZrO_x$ Ge and GeSn PMOSFETs with sub-60 mv/decade subthreshold swing, negligible hysteresis, and improved I_{DS}," in *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2016, pp. 12.2.1–12.2.4.

[18] Z. Krivokapic, U. Rana1 *et al.*, "14nm ferroelectric FinFET technology with steep subthreshold slope for ultra low power applications," in *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2017, pp. 15.1.1–15.1.4.

[19] D. Kwon, K. Chatterjee *et al.*, "Improved subthreshold swing and short channel effect in FDSOI n-channel negative capacitance field effect transistors," *IEEE Electron Device Letters*, vol. 39, no. 2, pp. 300–303, Feb 2018.

[20] T. S. Boscke, J. Mller *et al.*, "Ferroelectricity in hafnium oxide: CMOS compatible ferroelectric field effect transistors," in *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2011, pp. 24.5.1–24.5.4.

[21] S. Mueller, J. Mueller *et al.*, "Incipient ferroelectricity in Al-doped HfO_2 thin films," *Advanced Functional Materials*, vol. 22, no. 11, pp. 2412–2417, 2012.

[22] J. Mueller, T. S. Boscke *et al.*, "Ferroelectricity in simple binary ZrO_2 and HfO_2," *Nano Letters*, vol. 12, no. 8, pp. 4318–4323, 2012, pMID: 22812909.

[23] K.-S. Li, P.-G. Chen *et al.*, "Sub-60 mv swing negative-capacitance FinFET without hysteresis," in *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2015, pp. 22.6.1–22.6.4.

[24] M. H. Lee, S.-T. Fan *et al.*, "Physical thickness 1.x nm ferroelectric HfZrOx negative capacitance FETs," in *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2016, pp. 12.1.1–12.1.4.

[25] M. H. Lee, P.-G. Chen *et al.*, "Ferroelectric Al:HfO2 negative capacitance FETs," in *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2017, pp. 23.3.1–23.3.4.

[26] C.-J. Su, T.-C. Hong *et al.*, "Ge nanowire FETs with HfZrOx ferroelectric gate stack exhibiting SS of sub-60 mv/dec and biasing effects on ferroelectric reliability," in *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2017, pp. 15.4.1–15.4.4.

[27] C.-C. Fan, C.-H. Cheng *et al.*, "Energy-efficient HfAlOx NCFET: Using gate strain and defect passivation to realize nearly hysteresis- free sub-25mv/dec switch with ultralow leakage," in *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2017, pp. 23.2.1–23.2.4.

[28] A. J. Tan, A. K. Yadav *et al.*, "A nitrided interfacial oxide for interface state improvement in hafnium zirconium oxide-based ferroelectric transistor technology," *IEEE Electron Device Letters*, vol. 39, no. 1, pp. 95–98, Jan 2018.

[29] A. Nourbakhsh, A. Zubair *et al.*, "Subthreshold swing improvement in MoS_2 transistors by the negative-capacitance effect in a ferroelectric Al-doped-HfO_2/HfO_2 gate dielectric stack," *Nanoscale*, vol. 9, pp. 6122–6127, 2017.

[30] F. A. McGuire, Y.-C. Lin *et al.*, "Sustained sub-60 mv/decade switching via the negative capacitance effect in MoS_2 transistors," *Nano Letters*, vol. 17, no. 8, pp. 4801–4806, 2017, pMID: 28691824.

[31] Z. Yu, H. Wang *et al.*, "Negative capacitance 2D MoS_2 transistors with sub-60mv/dec subthreshold swing over 6 orders, 250 μa/ μm current density, and nearly-hysteresis-free," in *IEEE International Electron Devices Meeting (IEDM)*, Dec 2017, pp. 23.6.1–23.6.4.

[32] M. Si, C.-J. Su *et al.*, "Steep-slope hysteresis-free negative capacitance MoS_2 transistors," *Nature Nanotechnology*, vol. 13, pp. 24–28, 2018.

[33] M. Kobayashi, N. Ueyama *et al.*, "Experimental study on polarization-limited operation speed of negative capacitance FET with ferroelectric HfO_2," in *IEEE International Electron Devices Meeting (IEDM)*, Dec 2016, pp. 12.3.1–12.3.4.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/ACCESS.2018.2870916, IEEE Access

[34] Y. Li, K. Yao *et al.*, "Effect of ferroelectric damping on dynamic characteristics of negative capacitance ferroelectric MOSFET," *IEEE Transactions on Electron Devices*, vol. 63, no. 9, pp. 3636–3641, Sept 2016.

[35] Y. Li, K. Yao *et al.*, "Delay and power evaluation of negative capacitance ferroelectric MOSFET based on SPICE model," *IEEE Transactions on Electron Devices*, vol. 64, no. 5, pp. 2403–2408, May 2017.

[36] K. Chatterjee, A. J. Rosner *et al.*, "Intrinsic speed limit of negative capacitance transistors," *IEEE Electron Device Letters*, vol. 38, no. 9, pp. 1328–1330, Sept 2017.

[37] L. D. Landau and I. M. Khalatnikov, "On the anomalous absorption of sound near a second order phase transition point," *Dokladii Academii Nauk*, vol. 96, pp. 469–472, 1954.

[38] K. M. Rabe, C. H. Ahn *et al.*, *Physics of ferroelectrics: a modern perspective*. Springer Science & Business Media, 2007, vol. 105.

[39] G. Pahwa, T. Dutta *et al.*, "Analysis and compact modeling of negative capacitance transistor with high ON-current and negative output differential resistance – Part I: Model description," *IEEE Transactions on Electron Devices*, vol. 63, no. 12, pp. 4981–4985, Dec. 2016.

[40] A. I. Khan, K. Chatterjee *et al.*, "Negative capacitance in a ferroelectric capacitor," *Nature materials*, vol. 14, no. 2, pp. 182–186, 2015.

[41] M. Hoffmann, M. Pei *et al.*, "Direct observation of negative capacitance in polycrystalline ferroelectric HfO_2," *Advanced Functional Materials*, vol. 26, no. 47, pp. 8643–8649, 2016.

[42] P. Sharma, J. Zhang *et al.*, "Time-resolved measurement of negative capacitance," *IEEE Electron Device Letters*, vol. 39, no. 2, pp. 272–275, Feb 2018.

[43] G. Catalan, D. Jiménez *et al.*, "Ferroelectrics: Negative capacitance detected," *Nature materials*, vol. 14, no. 2, pp. 137–139, 2015.

[44] S.-C. Chang, U. E. Avci *et al.*, "Physical origin of transient negative capacitance in a ferroelectric capacitor," *Phys. Rev. Applied*, vol. 9, p. 014010, Jan 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevApplied.9.014010

[45] S. C. Chang, U. E. Avci *et al.*, "Inversion charge-boost and transient steep-slope induced by free-charge-polarization mismatch in a ferroelectric-metal-oxide-semiconductor capacitor," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, 2018.

[46] S. J. Song, Y. J. Kim *et al.*, "Alternative interpretations for decreasing voltage with increasing charge in ferroelectric capacitors," *Scientific Reports*, vol. 6, no. 1, feb 2016. [Online]. Available: https://doi.org/10.1038%2Fsrep20825

[47] B. Obradovic, T. Rakshit *et al.*, "Ferroelectric switching delay as cause of negative capacitance and the implications to NCFETs," in *Symposium on VLSI Technology*, 2018, pp. 51–52.

[48] A. K. Saha, S. Datta *et al.*, "Negative capacitance in resistor-ferroelectric and ferroelectric-dielectric networks: Apparent or intrinsic?" *Journal of Applied Physics*, vol. 123, no. 10, p. 105102, mar 2018. [Online]. Available: https://doi.org/10.1063%2F1.5016152

[49] J. A. Kittl, B. Obradovic *et al.*, "On the validity and applicability of models of negative capacitance and implications for MOS applications," *Applied Physics Letters*, vol. 113, no. 4, p. 042904, 2018. [Online]. Available: https://doi.org/10.1063/1.5036984

[50] Y. S. Chauhan, D. D. Lu *et al.*, *FinFET Modeling for IC Simulation and Design: Using the BSIM-CMG Standard*. Academic Press, 2015.

[51] "BSIM-CMG Technical Manual." [Online]. Available: http://www-device.eecs.berkeley.edu/bsim/?page=BSIMCMG

[52] L. T. Clark, V. Vashishtha *et al.*, "ASAP7: A 7-nm FinFET predictive process design kit," *Microelectronics Journal*, vol. 53, pp. 105–115, 2016.

[53] G. Vizdrik, S. Ducharme *et al.*, "Kinetics of ferroelectric switching in ultrathin films," *Physical Review B*, vol. 68, no. 9, p. 094113, 2003.

[54] "Synopsys HSPICE simulator," https://www.synopsys.com/verification/ams-verification/circuit-simulation/hspice.html.

[55] P. Sharma, K. Tapily *et al.*, "Impact of total and partial dipole switching on the switching slope of gate-last negative capacitance FETs with ferroelectric hafnium zirconium oxide gate stack," in *Symposium on VLSI Technology*, June 2017, pp. T154–T155.

[56] G. Pahwa, T. Dutta *et al.*, "Physical insights on negative capacitance transistors in non-hysteresis and hysteresis regimes: MFMIS vs MFIS structures," *IEEE Transactions on Electron Devices*, 2017 (in press).

[57] A. I. Khan, U. Radhakrishna *et al.*, "Negative capacitance behavior in a leaky ferroelectric," *IEEE Transactions on Electron Devices*, vol. 63, no. 11, pp. 4416–4422, Oct. 2016.

[58] M. Hoffmann, M. Pei *et al.*, "Modeling and design considerations for negative capacitance field-effect transistors," in *Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSOI-ULIS)*, Apr. 2017, pp. 1–4.

[59] Z. Dong and J. Guo, "A simple model of negative capacitance FET with electrostatic short channel effects," *IEEE Transactions on Electron Devices*, vol. 64, no. 7, pp. 2927–2934, July 2017.

[60] "Synopsys SiliconSmart for cell library characterization," https://www.synopsys.com/implementation-and-signoff/signoff/siliconsmart.html.

[61] "Synopsys EDA Tool Flows," https://www.synopsys.com/.

[62] "The RISC-V Foundation's Web Site," https://riscv.org/.

[63] C. Celio, D. A. Patterson *et al.*, "The berkeley out-of-order machine (boom): An industry-competitive, synthesizable, parameterized risc-v processor," EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2015-167, Jun 2015.

[64] "Github link for the Berkeley Out-of-Order Machine (BOOM) project," https://github.com/ucb-bar/riscv-boom.

[65] G. Pahwa, T. Dutta *et al.*, "Energy-delay tradeoffs in negative capacitance finfet based cmos circuits," in *2016 3rd International Conference on Emerging Electronics (ICEE)*, Dec 2016, pp. 1–4.

[66] S. Gupta, M. Steiner *et al.*, "Device-circuit analysis of ferroelectric FETs for low-power logic," *IEEE Transactions on Electron Devices*, vol. 64, no. 8, pp. 3092–3100, Aug 2017.

[67] "ANSYS Multiphysics Simulation Software," http://www.ansys.com/products/platform/multiphysics-simulation.

[68] C. J. Lasance and R. E. Simons, "Advances in high-performance cooling for electronics," *Electronics Cooling*, vol. 11, no. 4, pp. 22–39, 2005.

[69] S. Mishra, H. Y. Wong *et al.*, "Predictive tcad for nbti stress-recovery in various device architectures and channel materials," in *IEEE International Reliability Physics Symposium (IRPS)*, 2017, pp. 6A–3.

[70] L. E. Bell, "Cooling, heating, generating power, and recovering waste heat with thermoelectric systems," *Science*, vol. 321, no. 5895, pp. 1457–1461, 2008.

[71] V. Leonov, "Energy harvesting for self-powered wearable devices," in *Wearable Monitoring Systems*. Springer, 2011, pp. 27–49.

[72] "The PULP's Web Site," http://www.pulp-platform.org/.

[73] "Thermoelectric Generator Module," http://www.txlgroup.com/images/TXL-127-03L_data_sheet.pdf.

[74] "Thermal Conductivity data for Humans," http://users.ece.utexas.edu/~valvano/research/Thermal.pdf.

Hussam Amrouch (M'15) is a Research Group Leader at the Chair for Embedded Systems (CES), Karlsruhe Institute of Technology (KIT), Germany. He is leading of the Dependable Hardware research group. He received his Ph.D. (Summa cum laude) degree from KIT (2015). His main research interests are design for reliability, thermal-aware VLSI design, modeling and mitigating aging effects at the device/circuit levels. He holds six HiPEAC Paper Awards. He has three best paper nominations at DAC'16, DAC'17 and DATE'17 for his work on reliability. He currently serves as Associate Editor at Integration, the VLSI Journal. He is a member of the IEEE. ORCID iD 0000-0002-5649-3102

Girish Pahwa received the B.Tech. degree in electronics and communication engineering from Delhi Technological University, Delhi, India, in 2014. He is currently pursuing the Ph.D. degree with IIT Kanpur, Kanpur, India. His current research interests include physics and modeling of negative capacitance transistors.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/ACCESS.2018.2870916, IEEE Access

Amol D. Gaidhane received the M.Tech. degree in VLSI from the Center for VLSI and Nanotechnology, Visvesvaraya National Institute of Technology, Nagpur, India, in 2016. He is currently pursuing the Ph.D. degree with IIT Kanpur, Kanpur, India. His current research interests include simulation, modeling, and fabrication of negative capacitance transistors.

Jörg Henkel (M'95-SM'01-F'15) is the Chair Professor for Embedded Systems at Karlsruhe Institute of Technology. Before that he was a research staff member at NEC Laboratories in Princeton, NJ. He received his diploma and Ph.D. (Summa cum laude) from the Technical University of Braunschweig. His research work is focused on co-design for embedded hardware/software systems with respect to power, thermal and reliability aspects. He has received six best paper awards throughout his career from, among others, ICCAD, ESWeek and DATE. For two consecutive terms he served as the Editor-in-Chief for the ACM Transactions on Embedded Computing Systems. He is currently the Editor-in-Chief of the IEEE Design&Test Magazine and is/has been an Associate Editor for major ACM and IEEE Journals. He has led several conferences as a General Chair incl. ICCAD, ESWeek and serves as a Steering Committee chair/member for leading conferences and journals for embedded and cyber-physical systems. Prof. Henkel coordinates the DFG program SPP 1500 "Dependable Embedded Systems" and is a site coordinator of the DFG TR89 collaborative research center on "Invasive Computing". He is the chairman of the IEEE Computer Society, Germany Chapter, and a Fellow of the IEEE.

Yogesh Singh Chauhan (SM'12) is an associate professor at Indian Institute of Technology Kanpur (IITK), India. He was with Semiconductor Research & Development Center at IBM Bangalore during 2007 2010; Tokyo Institute of Technology in 2010; University of California Berkeley during 2010-2012; and ST Microelectronics during 2003-2004. He is the developer of industry standard BSIM-BULK (formerly BSIM6) model for bulk MOSFETs and ASM-HEMT model for GaN HEMTs. His group is also involved in developing compact models for FinFET, Nanosheet/Gate-All-Around FET, FDSOI transistors, Negative Capacitance FETs and 2D FETs. He is the Editor of IEEE Transactions on Electron Devices and Distinguished Lecturer of the IEEE Electron Devices Society. He is the member of IEEE-EDS Compact Modeling Committee and fellow of Indian National Young Academy of Science (INYAS). He is the founding chairperson of IEEE Electron Devices Society U.P. chapter and Vice-chairperson of IEEE U.P. section. He has published more than 200 papers in international journals and conferences. He received Ramanujan fellowship in 2012, IBM faculty award in 2013 and P. K. Kelkar fellowship in 2015, CNR Rao faculty award and Humboldt fellowship in 2018. His research interests are characterization, modeling, and simulation of semiconductor devices. He has served in the technical program committees of IEEE International Electron Devices Meeting (IEDM), IEEE International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), IEEE European Solid-State Device Research Conference (ESSDERC), IEEE Electron Devices Technology and Manufacturing (EDTM), and IEEE International Conference on VLSI Design and International Conference on Embedded Systems.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

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/TCE.2018.2873181, IEEE Transactions on Consumer Electronics

NFC Secure Element-based Mutual Authentication and Attestation for IoT access

Divyashikha Sethia, Daya Gupta and Huzur Saran

***Abstract*—Certain resourceful and powered Internet of Things can become victims to launch cyber attacks. Near Field Communication (NFC) can be used for their secure on-demand access. In this paper, we present a novel security framework for NFC Secure Element-based Mutual Authentication and Attestation for IoT access with a user device such as a mobile device using NFC-based Host Card Emulation (HCE) mode for the first time. HCE is robust as compared to the other NFC modes. A cloud-based Trusted Certified Authority (TCA) manages all cryptographic credentials and stores them in the tamper-resistant Secure Element (SE) and TPM (Trusted Platform Module)-based attestation modules on the devices. It uses a newly proposed NFC SE-based Mutual Authentication and Attestation (NSE-AA) protocol for proof-of-locality, end-to-end anonymous mutual authentication between the SEs and an associated remote attestation for trust. NSE-AA is robust and lightweight as compared to the existing schemes. We provide its informal and formal security analysis using the ROR model. A simulation on the AVISPA (Automated Validation of Internet Security Protocols and Applications) tool proves it's safety. We also briefly present the details of a prototype with a commercial mid-range priced mobile device and an SBC (Single Board Computers)-based IoT device.**

***Index Terms*—NFC, Internet of Things, Secure Element, Mutual Authentication, Attestation**

I. Introduction

CERTAIN Internet of Things (IoT) devices have sufficient power, resources and high computational capabilities which we term as **Advanced IoT devices (AdIoT)** such as surveillance cameras, powered access systems, heavy duty consumer home appliances like washing machine, refrigerator and family car. They can become victims to initiate cyber attacks such as the Distributed Denial of Service (DDoS) attack in October 2016 [1] caused due to malware and known factory default credentials on insecure storage of connected surveillance cameras. Federal Bureau of Investigation (FBI) has provided recommendations for measures to mitigate such potential threats and compromise of IoT devices in the future [1]. The IoT security survey by Granjal et al. [2] further suggests the importance of security, privacy and trust.

Near Field Communication (NFC) [3] is a low energy wireless technology operating at 13.56 MHz frequency, with few centimetres of access distance and a maximum throughput of 424 kbit/s. It can assist for a secure on-demand connecting, controlling and commissioning of such AdIoT devices through proximity to assure that only the two devices are communicating. It can help in secure device configuration, firmware update, cryptographic keys setup and access to logs. NFC has the advantage that due to proximity, it is hard to perform eavesdropping and man in the middle attacks. However, it uses an untrusted communication channel and does not ensure the authenticity, authorization and trustful state of the devices [4].

Two-way communication is essential for mutual security handshake and trust between the devices. Traditional NFC access of IoT device comprises of a user's device (typically a mobile device, special cards or Key fob) to access the IoT device through NFC tags which uses insecure NFC Data Exchange Format (NDEF) messages [5] or Peer-to-Peer mode [6] which is not open to developers for two-way communication. Hence there is a need to explore other NFC modes which are secure and open for developers. On user devices such as certain commercially available mobile devices, NFC uses a component Secure Element (SE) [7] which is a secure tamper-resistant smart card microchip connected to the NFC controller of the mobile device. It can be used for hardware card emulation of a contactless smart card on a mobile device with limited storage [3] as well as for tamper-resistant secure storage. In this paper, we suggest usage of Host Card Emulation (HCE) [7][8] mode on an NFC-based user device for a practically feasible secure IoT access. It provides emulation of a smart card using ISO 7816 standards [6] at a software level and can be accessed by a traditional card reader or an HCE-reader application on another user device. HCE has the advantages of providing an open platform for developers, support for two-way communication and large storage as compared to the other NFC modes [9]. It is typically used for financial applications, and this paper proposes a novel application using HCE for a secure IoT access for the first time to the best of our knowledge. However, since HCE is software-based, it is vulnerable to threats and requires mechanisms such as cloud-based SE, internal SE or Trusted Execution Environment (TEE) to secure the interaction [10]. Since TEE is less secure and cloud-SE may not be 24/7 available, we propose usage of SE to secure an HCE card.

More recently researchers have also been looking into the matter of trust with attestation for IoT access [11][12]. Malware can compromise the devices and make them victims to cause a cyber attack [1]. Remote attestation is a mechanism through which a device can prove it's trustful state to a remote device. Trusted computing can be in various forms such as hardware-based Trusted Platform Module (TPM) [11], or software-based Trusted Execution Environment (TEE) [12].

Divyashikha Sethia and Daya Gupta are faculty members in the Department of Computer Science and Engineering, Delhi Technological University, Delhi 110042, India. (email: sethiadivya@gmail.com, dgupta@dce.ac.in)

Huzur Saran is Professor in the Department of Computer Science, Indian Institute of Technology, New Delhi, 110016, India. (email: saran@cse.iitd.ernet.in)

Manuscript received March 23, 2018; revised September 27, 2018.

0098-3063 (c) 2018 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.

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/TCE.2018.2873181, IEEE Transactions on Consumer Electronics

Although TEE is lightweight, it is less secure and prone to physical attacks as compared to TPM [11][12].

TPM resides on the motherboard of the computing platform and is resistant to physical attacks. Before any component takes control of the CPU, it saves measurement of its characteristic code and configuration into TPM's Platform Configuration Registers (PCR). For every measurement of the system state, it stores it in a Stored Measurement Log (SML). The PCR values are later used with the SML to attest the device state to a remote party using an Attestation Identity Key (AIK) to sign and authenticate these values. The remote machine compares the signed values with the reference values to determine if the device is in a trustful state. The TPM also has an Endorsement Key (EK) which is generated by the manufacturer to uniquely identify it and obtain the AIK key from a trusted server.

HCE can provide a practical application of NFC with proof-of-locality for a secure IoT access with support for bidirectional communication for mutual authentication and attestation. However, these issues are not addressed together in any NFC-based security schemes in the previous work [13][14][11][15][16][17][18] to the best of our knowledge. All further references to the IoT and user device in the paper are for AdIoT device and user's mobile device respectively.

A. Case Study and Security Requirements

Internet-connected surveillance cameras are AdIoT devices which can provide remote monitoring for cities, offices and home. Compromise of such devices has made them victims to cause the DDoS attack in 2016 [1]. The devices were used to form an army of zombies to attack and make specific services and websites unavailable. Hence it is crucial to configure such devices with updated firmware, configuration and credentials for secure access. NFC can be used for secure access to such devices by an authorised technician with his user device for control, configuration, setup of keys and access to critical logs. Another example is a Portable patient's HCE-based health card on a mobile device that can retain dispersed health records to provide readily available health history to seek timely diagnosis and treatment when required [19]. An authorized medical professional must be able to selectively access it based on the role and permissions for a direct yet secure read and write access over an NFC tap with his mobile device. In both use cases, NFC can provide proof-of locality and make man in the middle attacks difficult.

Security Requirements: According to the survey by Granjal et al. [2], it is critical to access IoT devices with security, privacy and trust. With the recent DDoS attacks and recommendations for overcoming threats [1], the security framework must fulfil the security requirements for Confidentiality, Integrity, Availability, Authentication, Trust, Privacy, User Anonymity and Secure Storage. Although NFC assures security through proximity, it is vulnerable to various threats [4]. The NFC-interaction must assist bidirectional communication to ensure the validity of devices through mutual authentication and trustful states through attestation. Any malware on the user device or the IoT device can harm the configuration, setup, key installation process or cause theft of critical data. Secure storage must retain all cryptographic credentials. All stored, as well as communicated data must be encrypted to provide confidentiality and preserve user anonymity. Privacy and anonymity for the user must prevent any disclosure such as ailments like mental health, urinary diseases for a healthcard. The framework must also secure the smart cards [20] on devices from attacks such as Dos, Replay, Collusion, Man in the middle (MTM), Parallel session, Forgery and Platform impersonation attacks.

B. Contribution

- We propose a novel framework for NFC Secure Element-based Mutual Authentication and Attestation for IoT access with an HCE card application on a user device over an NFC tap. It is supported by a cloud-based Trusted Certified Authority (TCA) for management of cryptographic credentials which are stored in the tamper-resistant internal SEs [7] and TPM[11] attestation modules on devices.
- We propose a novel NFC SE-based Authentication and Attestation (NSE-AA) protocol for an end-to-end anonymous lightweight mutual authentication with limited use of symmetric encryption between the two SEs and TPM-based Attestation for security and trust over HCE.
- We present a detailed security analysis of the NSE-AA protocol with an informal security proof, a formal security proof using the ROR model [21] and a detailed comparison of its computation and communication overheads with the existing schemes. We also prove that it is safe using the Automated Validation of Internet Security Protocols and Applications (AVISPA) tool[22].
- We demonstrate the details of a prototype for the proposed framework consisting of commercial mid-range priced Single Board Computers (SBC)-based IoT device being accessed by a mobile device using NFC-based HCE card.

C. Roadmap

In the rest of the paper, section II presents the Related Work. In Section III we describe the proposed framework design and details of the newly proposed NSE-AA security protocol. Section IV explains the Security Analysis for informal analysis, formal proof using ROR model and formal verification through simulation on the AVISPA tool. In Section V we present the Performance Analysis and Comparison of the NSE-AA protocol with the related schemes qualitatively as well as quantitatively. It is followed by Section VI which discusses the details of the prototype for the framework in brief. We finally conclude the paper in Section VII with the conclusion and future work.

II. Related Work

Previous papers for NFC-based security focused on the issues of authentication and attestation separately using NFC modes other than HCE which has several advantages as discussed in the introduction. In this paper, we look into the HCE mode for a feasible alternative for bidirectional communication. It helps in a robust security handshake comprising of lightweight anonymous mutual authentication, attestation and protection from several threats.

0098-3063 (c) 2018 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.

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/TCE.2018.2873181, IEEE Transactions on Consumer Electronics

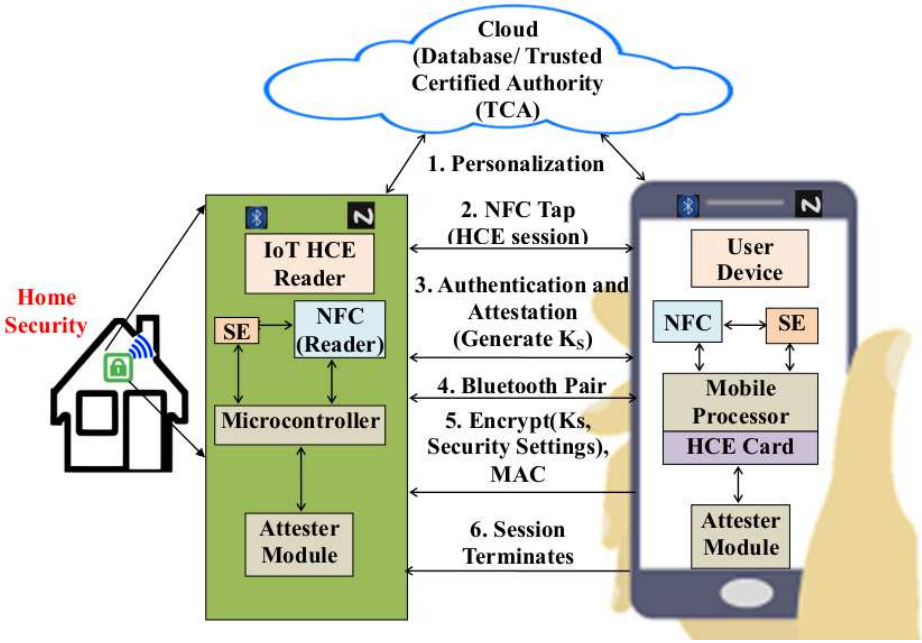


Fig. 1: Block diagram for the proposed framework

Authentication Ceipidor et al. [13] proposed a mutual authentication protocol Kernees based on Needham Schroeder protocol. However, it fails to achieve message authenticity and is prone to Brute Force attacks. Thammarat et al. [14] proposed a lightweight mutual authentication protocol based on limited use of session key and prevented Brute Force attack by having a set of sessions keys. However, it lacks support for user anonymity and attestation. There have been several authentication protocols using NFC-based Peer-to-Peer mode such as [16][18][15][23]. However, none of them addresses the issues of mutual attestation, secure storage, user anonymity and protection to threats like MTM and DoS attacks. The NFC Peer-to-Peer mode for two-way communication uses Simple NDEF Exchange Protocol (SNEP) service [6] which is not open for developers on unrooted Android devices as compared to the HCE mode.

Attestation Toegl et al. [11] proposed a scheme to use an NFC-based mobile device as a Mobile Attestation Token (MA) to access a TPM-based kiosk. In this scheme, the mobile device generates a random nonce N_A and sends it as part of a challenge to the kiosk which passes it internally to its TPM module. The TPM signs its attestation report and nonce N_A with the private AIK key and sends it to a trusted Virtual server (*VS*), which validates it and prepares a validation ticket for the user. However, the mobile and kiosk use insecure NDEF messages, and there is no validation of the mobile device which may also be prone to malware, especially on rooted user devices and cause a breach of trust [7][8]. Aziz et al. [24] provided an extension of TLS (Transport Layer Security) for asymmetric encryption-based mutual authentication and TPM-based mutual attestation over TCP/IP. Each host with identifier *idhost* generates a unique random nonce N_H and exchanges them in the TLS session to generate a session key K_S. It then generates the AIK certificate for the validity of its AIK identifier id_{AIKH}:$h(idhost,N_H)$ and public AIK key Pk_{AIKH}. In the attestation phase, the host sends its attestation report consisting of the PCR and SML values signed by its AIK private key Sk_{AIKH} along with its AIK certificate $Cert_{AIKH}$ to the remote host to verify its trustful state. However, the scheme has drawbacks of being computationally expensive due to asymmetric encryption, no support for user anonymity, no proof-of-locality and secrecy of PCR and SML since the host sends an unencrypted $Cert_{AIKH}$.

The NSE-AA protocol proposed in this paper improves Aziz et al.'s [24] work with secure storage on SE and TPM, an end-to-end lightweight anonymous mutual authentication using symmetric encryption between the SEs, a lightweight AIK certificate generation, secrecy of PCR and SML and proof-of-locality using NFC. The NSE-AA mutual authentication phase improves the authentication scheme presented by Thammarat et al. [14] by introducing user and device anonymity with symmetric encryption, random nonces and password known only to the user and device owner. It provides integrity by using hash functions. It is also secure based on the Real-Or-Random (ROR) model [21]. There are other provably secure lightweight schemes such as presented by Chang et al. [25]. However, it is based on xor operations which are less secure as compared to symmetric encryption. Also, the scheme provides only user anonymity and no device anonymity. Tables V, VI and VI present the comparison of our protocol with the related schemes in section V.

III. Proposed Framework

We propose a novel framework for NFC Secure Element-based Mutual Authentication and Attestation for IoT access from a user device over HCE. Fig 1 illustrates the block diagram for the proposed framework with the use case of a user's mobile device accessing the Home security IoT system over NFC for secure configurations. The novelty of the framework is that it uses NFC-based HCE mode which has advantages over the other NFC modes as discussed in

0098-3063 (c) 2018 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.

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/TCE.2018.2873181, IEEE Transactions on Consumer Electronics

the introduction. A user device emulates an HCE card which interacts with the reader application on the NFC module of the IoT device over an NFC tap. It uses a novel HCE library for two-way communication and exchange of large data through fragmentation. A cloud-based Trusted Certified Authority (TCA) manages device identities and cryptographic credentials that are stored in the internal SEs and TPM attestation modules on the devices. The framework assumes that the SE is accessible to special applications on the processor as well as a TPM-based attestation module. The framework uses a novel end-to-end protocol between the SEs and TPMs of the devices for lightweight anonymous mutual authentication and mutual attestation for a secure IoT access.

Figure 1 illustrates the sequence of steps which are explained as follows:

1) Both devices register with a Trusted Certified Authority (TCA) on a secure cloud-based service and initialize cryptographic credentials on their SE and TPM.
2) The devices tap to communicate over the NFC interface. The HCE reader on the NFC module of the IoT device accesses the HCE card on the user's mobile device.
3) Both devices initiate the NFC SE-based mutual Authentication and Attestation (NSE-AA) Protocol for an anonymous mutual authentication phase followed by mutual attestation phase as discussed in the following subsection. It generates a one-time symmetric session key K_S to encrypt all further communication. The attestation module can alternatively comprise of TEE [12].
4) The HCE tap further automates Bluetooth pairing for the exchange of large-sized data for higher throughput as compared to NFC.
5) The user device sends all device communication such as security settings using encryption and generation of its Message Authentication Code (MAC) using the session key K_S.
6) The IoT access terminates on exiting the HCE application or removing the devices apart.

A. NSE-AA Protocol

We present the details of the NSE protocol in this section. It improves the mutual authentication scheme suggested by Thammarat et al. [14] and associates it with a TPM-based attestation phase. The attestation phase improves the attestation scheme proposed by Aziz et al. [24] for a secure NFC-based IoT access. It comprises of the following phases:

- Mutual authentication which generates the session key K_S.
- Each device generates an Attestation Identity (AIK) Certificate by using nonces and K_S in the previous phase.
- Each device attests the software and memory status using the AIK private keys to the remote device.

1) Registration and Personalisation: Table I provides the notations used for the NSE-AA protocol. During the personalization phase, each host H(IoT/User device) initializes its SE and TPM with credentials issued by the TCA. Each host has a unique actual identity Id_{AH} and a secure symmetric key K_{HS} shared between itself and the TCA. Every user device and IoT device also retain a unique symmetric key K_{UD} shared between them. All credentials are stored on the SE of the devices. Each host H generates a password pwd_H and a random number b_H and uses them to generate pwb_H:$h\ (pwd_H \| b_H)$ and R_H:$h(Id_{AH} \| K_{HS})$. It stores id_{AH}, b_H and R_H on its local SE and the TCA. These values are later used to generate the virtual identity id_{VH} in the mutual authentication phase for verification along with user anonymity. The Endorsement certificate $Cert_{EKH}$ for attestation credentials resides on the TPM. The Endorsement public key Pk_{EKH} uniquely identifies the host's TPM and presents it to the TCA for identification and generation of the AIK certificate.

TABLE I: Notations Used

Symbol	Meaning
Pk_{TCA}	Public verification key of the Trusted CA
A	Adversary
H	Host (IoT device D /User device U)
id_{AH}	Unique actual host identifier
pwd_H	Password for host H identity
Id_{VH}	Virtual identity/ Pseudonym for host H
N_H	Non-predictable nonce of host H
KDF	Key Generation function
K_{HS} (K_{US}/K_{DS})	Symmetric key between host and server
K_{UD}	Symmetric key between IoT and user devices
K_S	Symmetric session key generated in NSE-AA
id_{AIKH}	AIK identifier for host
$Cert_{AIKH}$	Certificate AIK for host H
Pk_{AIKH},Sk_{AIKH}	Public/Private AIK keys for host H
$Cert_{EKH}$	Endorsement certificate for host H
S_{PCR}	Selection of PCR values to verify
I_V	Request to indicate if TPM version is required
$TPMinfo_H$	Version/revision information for host H's TPM
SML_H	Stored Measurement Log (SML) on TPM
$BT\text{-}Addr$	Bluetooth MAC Address
$sign(M)K$	Sign message M with private key K
$E(K,M)$	Encrypt message M with key K
$D(K,M)$	Decrypt message M with key K
$h(M)$	Hash over message M
$IoT\text{-}SE$<->$User\text{-}SE$	End-to-end communication between SEs

2) Mutual Authentication: Table II illustrates the steps for the NSE-AA mutual authentication phase. All valid user and IoT devices must securely identify and mutually authenticate each other to ensure that only registered devices are interacting. The SEs compute all the cryptographic operations. Special internal applications on the device access these computations from the applets on the SE and send them to the SE of the remote device over the HCE interface. Hence mutual authentication is end-to-end between the two SEs and does not allow a snooper to gain any valuable information from the communication.

Each device generates its virtual identity id_{VH} using a random nonce N_H. Before initiating access, a host H must enter password pwd'_H and the actual identity id'_{AH}. It computes pwb'_H and R'_H and compares them as well as id'_{AH} with the stored values on the SE. It then generates id_{VH}:<C_H,D_H,B_H> and sends it to the TCA for verification, where C_H:$h(pwb_H \| N_H \| R'_H)$, D_H:$xor(h(Id_{AH} \| pwb_H), N_H)$, B_H:$E(K_{HS}, (Id_{AH} \| pwb_H \| N_H))$. On receipt of id_{VH}, the TCA verifies it by decrypting B_H to extract id'_{AH}, pwb'_H and N'_H. It then computes R'_H:$h(Id'_{AH} \| K_{HS})$, N'_H:$xor(D_H, h(Id'_{AH} \| pwb'_H))$ and C'_H:$h\ (pwb'_H \| N'_H \| R'_H)$. If N'_H and C'_H are same as the received values, then it validates the host H. The details of the mutual authentication phase are:

0098-3063 (c) 2018 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.

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/TCE.2018.2873181, IEEE Transactions on Consumer Electronics

TABLE II: Mutual Authentication

Messages	Description
IoT-SE<->User-SE	1.M1:Select AID
User-SE:	2.Generate N_U, $T1{:}xor(N_U,K_{UD})$, Id_{VU}, $O{:}E(K_{US},(N_U))$,$Q{:}h(N_U\|K_{UD})$
IoT-SE<-User-SE:	3.$M2{:}Id_{VU}\|T1\|E(K_{UD},O)\|Q$
IoT-SE:	4.$N_U{:}xor(T1,K_{UD})$; Decrypt to extract O; Verify N_U in Q; Generate N_D,Id_{VD}
IoT-SE->Server:	5.$M3{:}Id_{VD}\|Id_{VU}\|E(K_{DS},(N_D))\|O$
Server:	6.Verify Id_{VD},Id_{VU}; Decrypt and extract N'_U, N'_D; Generate $T3{:}xor(pwb'_U,K_{DS})$,$T4{:}xor(pwb'_D,K_{US})$, $R{:}h(pwb'_U\|N'_U\|N'_D\|K_{US})$, $X{:}h(pwb'_D\|N'_U\|N'_D\|K_{DS})$
IoT-SE<-Server:	7.$M4{:}T3\|T4\|X\|R$
IoT-SE:	8.$pwb'_U{:}xor(T3,K_{DS})$; Verify X; Generate $K_S{:}$ $KDF(pwb'_U\|pwb_D\|N_U\|N_D\|K_{UD})$,$T2{:}xor(N_D,K_{UD})$, $Y{:}h(T2\|R\|K_{UD})$
IoT-SE->User-SE:	9.$M5{:}T2\|T4\|Y$
User-SE:	10.$N'_D{:}xor(T2,\ K_{UD})$, $pwb'_D{:}xor(T4,K_{US})$, Verify nonces in Y and authenticate IoT-SE; Generate $K_S{:}KDF(pwb_U\|pwb'_D\|N_U\|N'_D\|K_{UD})$; $Z{:}h(N'_D\|K_{UD}\|K_S)$
IoT-SE<-User-SE:	11.$M6{:}Z$
IoT-SE:	12.Verify Z and authenticate user U
IoT<->User:	13.$E(K_S,Data)$

TABLE III: NSE-AA AIK Certificate Generation

Messages	Description
Host:	1.Get N_H;$id_{AIKH}{:}h(id_{VH}\|N_H)$;TPM Load Key ($Sk_{AIKH}$)
Host:	2.SigTPM:TPMMakeIdentity:$sign(h(id_{AIKH}\|Pk_{TCA}))Sk_{EKH}$
Host->TCA:	3.$M7{:}E(K_{HS},(Pk_{AIKH}\|N_H\|SigTPM\|Pk_{EKH}))\|id_{VH}$
TCA:	4.Decrypt $M7$ using K_{HS};Verify Pk_{EKH},id_{VH}; Generate $id'_{AIKH}{:}h(id_{VH}\|N_H)$;Verify SigTPM;
TCA:	5.Generate $Cert_{AIKH}{:}sign(Pk_{AIKH},\ id_{AIKH})Sk_{TCA}$;
TCA:	6.Generate K_C,$C1{:}E(K_{HS},(K_C\|h(Pk_{AIKH}))$, $C2{:}E(K_C,Cert_{AIKH})$
Host<-TCA:	7.$M8{:}C1\|C2$
Host:	8.'.Decrypt $C1$ to extract K'_C,Generate $h'(Pk_{AIKH})$; Verify it with $h(Pk_{AIKH})$;Decrypt $C2$ to extract $Cert_{AIKH}$.

TABLE IV: Mutual Attestation

Messages	Description
IoT<-User:	1.M9:N_U,I_V,S_{PCR} (Send challenge)
IoT:	2.Get SML_D;loadKey(Sk_{AIKD});$SigTPM{:}TPM\text{-}Quote{:}$ $sign(PCR_D,SML_D,N_U,TPMInfo)Sk_{AIKD}$; Q:SigTPM|TPMInfo|E($K_S$, $Cert_{AIKD}$)
IoT->User:	3. $M10{:}Q,MAC_{KS}(Q)$
User:	4.Decrypt to extract Cer_{AIKD};Verify $Cert_{AIKD}$ using Pk_{TCA}; Verify SigTPM using Pk_{AIKD},N_U and TPMInfo; Generate $Id'_{AIKD}{:}h(Id_{VD}\|N_D)$ and compare with Id_{AIKD} in $Cert_{AIKD}$

Step 1: IoT-SE/Reader selects applet on the HCE card on the User-SE using the Application identifier(AID).

Step 2-3: User-SE generates a random nonce N_U,$T1{:}xor(N_U,K_{UD})$, virtual identity $Id_{VU}{:}<C_U,D_U,B_U>$, $O{:}E(K_{US},\ (N_U))$, $Q{:}h(N_U\|K_{UD})$ to assure the integrity of N_U to the IoT device; sends a message $M2{:}Id_{VU}\|T1\|E(K_{UD},O)\|Q$ to the IoT device.

Step 4-5: IoT-SE extracts $N'_U{:}xor(T1,K_{UD})$, decrypts to extract O; verifies nonce N'_U in Q; generates a random nonce N_D,virtual identity $Id_{VD}{:}<C_D,D_D,B_D>$; sends a message $M3{:}Id_{VD}\|Id_{VU}\|E(K_{DS},N_D)\|O$ to the server.

Step 6-7: The server decrypts and extracts N'_U and N'_D; verifies the virtual identities Id_{VU} and Id_{VD}, extracts pwb'_U and pwb'_D, validates the nonces N'_U and N'_D from the virtual identities; Generates $T3{:}xor(pwb'_U,K_{DS})$, $T4{:}xor(pwb'_D,K_{US})$, $R{:}h(pwb'_D\|N'_U\|N'_D\|K_{US})$ to assure the integrity of nonces sent to the User-SE; generates $X{:}h(pwb'_U\|N'_U\|N'_D\|K_{DS})$ to assure the integrity of nonces sent to the IoT-SE; sends a message $M4{:}T3\|T4\|X\|R$ to the IoT-SE.

Step 8-9: IoT-SE extracts $pwb'_U{:}\ xor(T3,K_{DS})$; verifies nonce N_D and N_U in X; generates $T2{:}xor(N_D,K_{UD})$, $Y{:}h(T2\|R\|K_{UD})$ and session key $K_S{:}KDF(pwb'_U\|pwb_D\|N'_U\|N_D\|K_{UD})$; sends a message $M5{:}T2\|T4\|Y$ to the User-SE.

Step 10-11: User-SE extracts $N'_D{:}xor(T2,K_{UD})$, $pwb_D{:}xor(T4,K_{US})$; validates N'_D by verifying Y; authenticates IoT-SE by validating N_U and K_{US} in R; generates the session key $K_S{:}KDF(pwb_U\|pwb'_D\|N_U\|N'_D\|K_{UD})$ and $Z{:}h(N'_D\|K_{UD}\|K_S)$ to authenticate itself to the IoT-SE; sends a message $M6{:}Z$ to the IoT-SE.

Step 12-13: IoT-SE verifies N_D in Z and authenticates User-SE. The session key K_S secures all further communication.

3) Mutual Attestation: Both devices must prove their trustful states. We assume that the devices have an inbuilt attester module for remote attestation. The MAC with session key K_S encrypts all attestation messages, to maintain confidentiality and integrity. The following are the detailed phases:

AIK certificate generation protocol: Table III shows the certificate generation for host H(IoT/user device) with a virtual identity id_{VH} and a random nonce N_H generated prior. The details of the steps are:

Step 1: Host H generates the AIK identity $id_{AIKH}{:}h(id_{VH}\|N_H)$ to prevent replay and collusion attacks.

Step 2: Host TPM verifies itself to the TCA with its Endorsement Key (EK) credentials and generates an asymmetric attestation key pair Pk_{AIKH}, Sk_{AIKH} and $SigTPM{:}sign\ (h(id_{AIKH}\|Pk_{TCA}))Sk_{EKH}$.

Step 3: Host H sends a message $M7{:}E(K_{HS},(Pk_{AIKH}\|N_H\|SigTPM\|Pk_{EKH}))\|id_{VH}$ to the TCA for generation of the AIK certificate after validation of its id_{VH} and Pk_{EKH}.

Step 4: TCA generates id'_{AIKH} to verify SigTPM;

Step 5: TCA generates the AIK certificate $Cert_{AIKH}$;

Step 6: TCA creates a session key K_C; $C1{:}E(K_{HS},(K_C\|h(P_{AIKH}))$ and $C2{:}E(K_C,Cert_{AIKH})$;

Step 7: TCA sends a message $M8{:}C1\|C2$ to the host H.

Step 8: Host H decrypts $C1$ to extract key K'_C, generates the hash of Pk_{AIKH} and compares it with the received hash. It uses K'_C to decrypt $C2$ and extract and store $Cert_{AIKH}$.

We improve the Aziz et al.'s scheme [24] by reducing the computations of the certificate generation phase with symmetric encryption in messages $M7$ and $M8$.

TPM-based Attestation Protocol: Table IV shows the steps for attestation of an IoT device by a user device. The TPM-based attestation scheme improves the integrity report protocol by Aziz et al. [24]. It comprises of a TPM challenge-response authentication and uses the nonces and virtual identities exchanged in the NSE-AA mutual authentication phase to prevent a replay attack. The details of the steps are:

Step 1: User device sends nonce N_U as a challenge, I_V to

0098-3063 (c) 2018 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.

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/TCE.2018.2873181, IEEE Transactions on Consumer Electronics

request TPM version number and S_{PCR} to specify the selection of PCR values in message *M9*.

Step 2: IoT device uses the nonce N_U received and its private AIK key to generate a signed quote $SigTPM_D{:}sign(PCR_D \| SML_D \| N_U \| TPMInfo_D)Sk_{AIKD}$ to preserve the secrecy of PCR_D and SML_D. It then generates $Q{:}SigTPM \| TPMInfo \| E(K_S, Cert_{AIKD})$ and its Message Authentication Code $MAC_{KS}(Q)$ using the session key K_S.

Step 3: IoT device sends a message $M10{:}Q \| MAC_{KS}(Q)$ to the user device.

Step 4: User device verifies $MAC_{KS}(Q)$ and the AIK certificate with the stored TCA's public key Pk_{TCA}. It further verifies the attestation data to ascertain the trustful state of the IoT device.

Similarly, the IoT device also remotely attests the user device. It also preserves the secrecy of the PCR and SML for attestation by encrypting $Cert_{AIK}$ with key K_S in parameter Q of message *M10* which is sent unencrypted by Aziz et al.'s scheme [24].

IV. Security Analysis

A. Informal Proof

The NSE-AA satisfies the following security requirements identified in Section I.A:

- **S1: Confidentiality** The communication of encrypted data along with storage of the encrypted data stored on the tamper-resistant SEs ensures confidentiality.
- **S2: Integrity** Communication of encrypted data along with its MAC ensures the integrity of data.
- **S3: Mutual Authentication and Attestation** The NSE-AA protocol provides lightweight anonymous mutual authentication along with attestation for secure identification and establishment of trust over HCE. The mutual authentication phase improves the Thammarat et al. [14] scheme by providing user and device anonymity with symmetric encryption, random nonces and password known only to the device owners, integrity by using hash functions and being secure based on the Real-Or-Random (ROR) model [21]. The mutual attestation phase improves the Aziz et al.'s [24] scheme with reduced overhead since it uses symmetric encryption and hash functions in the mutual authentication phase and preserves the secrecy of the PCR and SML in the attestation phase.
- **S4: Privacy** Encryption of data with credentials stored on the secure tamper-resistant SEs prevents an intruder to seek any critical user data.
- **S5: User Anonymity** A host H sends its virtual identity id_{VH} to the TCA for verification. $id_{VH}{:}(C_H \| D_H \| B_H)$. It is generated using a password known only to the host H and random nonce N_H along with other parameters stored on the SE of the devices. Hence, it prevents replay attacks and also prevents the adversary from finding the actual identities of the devices.
- **S6: Proof-of-locality** NFC assures proof-of-locality for a secure IoT access. The device proximity for NFC makes man-in-the-middle attack, and eavesdropping difficult.
- **S7: Secure Storage** Tamper-resistant SE ad TPM provide secure storage and cryptographic computations for an end-to-end NSE-AA protocol

The framework also prevents the following attacks:

- **T1: Denial of Service (Dos) attack** An attacker cannot pass through the verification of the tamper-resistant SE and hence prevents the DoS attack.
- **T2: Replay attack** Unique nonces in NSE-AA prevent an intruder to resend the messages and hence prevent a replay attack.
- **T3: Collusion attack** For a collusion attack, a host H provides another host V's attestation data: Vs SML and PCR information signed by V's AIK private key Sk_{AIKV} and V's AIK certificate $Cert_{AIKV}$. However, it has previously shared id_{VH} and N_H with the remote host in the mutual authentication phase. The remote device calculates $id_{AIKH}{:}h(id_{VH} \| N_H)$ and finds it different from the id_{AIKV} in $Cert_{AIKV}$ and hence the attack fails.
- **T4: Parallel session attack** Each session uses tamper-resistant SE for cryptographic computations and secure storage and also uses unique nonces to prevent replay for parallel sessions.
- **T5: Forgery attack** It is difficult to forge since credentials reside on tamper-resistant SE and TPM. Also, the virtual identity components C_H, D_H and B_H require pwb_H, id_{AH} and the secret symmetric key K_{HS} which is known only to the user.
- **T6: Platform impersonation attack** Virtual identity and messages exchanged in the communication comprises of symmetric keys shared between devices and the TCA or the private component of the asymmetric keys of TCA and AIK. Hence it is difficult for a malicious server to replace a valid TCA server. The cryptographic credentials are stored on the SE of the devices and known only to the valid server.
- **T7: Man-in-the-middle (MTM) attack** Both devices register with the TCA. Also, tamper-resistant SE stores unique identities and certificates and executes cryptographic computations for the end-to-end security handshake in the NSE-AA protocol. Hence, any unregistered user cannot eavesdrop, spoof, decrypt and relay messages.
- **T8: Insider attack impersonation of the device** End-to-end mutual authentication between the two SEs and unique keys for each user prevents any user to capture messages of another user and impersonate it.

B. Formal Security proof using ROR model

We prove the security of the session key K_S generated in the Mutual Authentication phase of the NSE-AAA protocol using the Real-Or-Random (ROR) model [21]. The protocol comprises of entities user U, IoT device D and Trusted Certified Authority *TCA*.

1) **ROR Model**

- **Participants** Let I_U^{t1}, I_D^{t2} *and* I_{TCA}^{t3} be the instances *t1, t2* and *t3* of *U, D* and *TCA* and are called oracles.
- **Accepted State** An instance I^t moves to an Accepted state after receiving the final protocol message. The ordered concatenation of all communications messages sent and received by I^t is called the session identification (sid) of I^t for the current session.

0098-3063 (c) 2018 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.

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/TCE.2018.2873181, IEEE Transactions on Consumer Electronics

- **Partnering** Instances I^{t1} and I^{t2} are called partners if: 1) both are in the Accepted state, 2) both mutually authenticate each other and share the same sid, and 3) both are mutual partners of each other.
- **Freshness** If the session key K_S between U and D is not revealed through a reveal query as defined below, I_U^{t1} and I_D^{t2} are considered as fresh.
- **Adversary** An adversary A is modelled using the Dolev-Yao (DY) model [26] and can eavesdrop, modify, delete or inject the messages transmitted between the entities involved during the communication with the help of the following queries:
  Execute(I^{t1},I^{t2}): The query models the eavesdropping attack and allows an adversary to eavesdrop the messages communicated among *U, D* and *TCA*.
  Send(I^t,msg): The query models an adversary to transmit a message *msg* to I^t and receives a response from it. It is further modelled as an active attack.
  Reveal(I^t): The query reveals the session key K_S between I^t and its partner to the adversary in the current session.
  CMD(I_U^{t1}): Under this query, an adversary can fetch all the sensitive secret credentials stored from the lost or stolen mobile device's memory and represents an active attack.
  Test(I^t): At the beginning of a game, an unbiased coin *c* is flipped. The adversary executes this test query and if *c:1* and the session key K_S between *U* and *D* is fresh, I^t returns K_S else if *c:0* it returns a random number else it returns a null value.
  We assume that the adversary can access only a limited number of *CMD(I_U^{t1})* queries, while an unlimited number of Test(I^t) queries are accessible.

2) **Security Proof** We prove that the NSE-AA protocol is safe.
Theorem 1: For an adversary running in polynomial time *t* against the NSE-AA protocol in the random oracle model, and *D* a uniformly distributed password dictionary, if no device has been compromised by the adversary then:

$$Adv_{\text{NSE-AA}}^{\text{ake}} <= \frac{q_h^2}{|Hash|} + \frac{2qsend}{|D|}$$

where q_h: number of Hash queries, *qsend*: number of Send queries, |*Hash*|: range of the hash function, and |*D*|: size of *D*.
Proof: We follow proof similar as presented by Chang et al. [25]. Let *Gi*, where *i:[0,3]* be the sequence of games and $Succ_i$ be an event when the adversary succeeds in guessing the bit b in the game G_i.
Game G0: The game *G0* presents the real attack by the adversary against the NSE-AA protocol. The bit *b* is chosen at random at the beginning of this game, and hence we have:

$$Adv_{\text{NSE-AA}}^{\text{ake}} = 2Pr[Succ_0] - 1 \quad (1)$$

Game G1: This game simulates the adversary's eavesdropping attacks with the *Execute* query. In the end, the adversary sends the *Test* query, whose output decides if the adversary obtains a real session key K_S or a random number. The key K_S*:KDF($pwb_U \| pwb_D \| N_U \| N_D \| K_{UD}$);$pwb_U$:h($pwd_U$,$b_U$), pwb_D:h(pwd_D,b_D)*. An adversary cannot compute K_S unless it has the SEs which has the secret key K_{UD} and the random numbers b_U and b_D and the nonces which are exchanged using xor with the symmetric key K_{UD} in the terms *T1* and *T2* in messages *M2* and *M5* respectively. The adversary also needs the passwords of the user and IoT device besides the random numbers b_U and b_D to generate terms pwb_U and pwb_D. These are exchanged in terms *T3* and *T4* in message *M4* and hence cannot be obtained by the adversary without the knowledge of symmetric keys K_{DS} and K_{US} respectively stored in the SEs of the devices. Hence the chances of winning for the adversary do not increase when an adversary eavesdrops.

$$Pr[Succ_0] = Pr[Succ_1]. \quad (2)$$

Game G2: The game *G2* modifies the game *G1* by addition of the *Send* and the *Hash* queries to model an active attack. In this attack, the adversary tries to deceive a participant to accept an illegal message. The adversary sends repeated *Hash* queries to find collisions. Since each of the messages is associated with the identity of a participant or a unique session symmetric key between the entities which are generated based on the actual identities, the *Send* query does not cause any hash collisions. As per the birthday paradox, we have

$$|Pr[Succ_1] - Pr[Succ_2]| \leq \frac{q_h^2}{2|Hash|} \quad (3)$$

Game G3: This game *G3* modifies the game *G2* by addition of the *CorruptSC* queries to simulate the loss of a smart card or SEs (lost attack). Even though the details of the SE are known it is important to know the user and IoT device passwords pwd_U and pwd_D to find the session key K_S. Since the password has a low-entropy, the adversary can try to guess the password using the online dictionary attacks. However, if the system limits the number of wrong password guesses then:

$$|Pr[Succ_2] - Pr[Succ_3]| \leq \frac{q_{\text{send}}}{|D|} \quad (4)$$

After the last game *G3*, all queries have been simulated. If the adversary has failed to break the security of the NSE-AA protocol it finally tries to win the game through the *Test()* query to guess the bit *b*. Hence, *Pr[$Succ_3$]* is the same as the probability of guessing the bit *b*.

$$Pr[Succ_3] \leq \frac{1}{2} \quad (5)$$

From equations 1-5 we have:

$$Adv_{\text{NSE-AA}}^{\text{ake}} <= \frac{q_h^2}{|Hash|} + \frac{2qsend}{|D|}$$

Hence, the NSE-AA protocol is secure when the range of the hash function and the size of the password dictionary are large. The adversary may query *CorruptSC* oracle to simulate the stolen SE and simulate smart card breach attacks. Still, the adversary cannot generate the session key K_S due to the requirement of the user's password. Hence the adversary can compromise the NSE-AA protocol only through an online guessing attack. However, since

0098-3063 (c) 2018 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.

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/TCE.2018.2873181, IEEE Transactions on Consumer Electronics

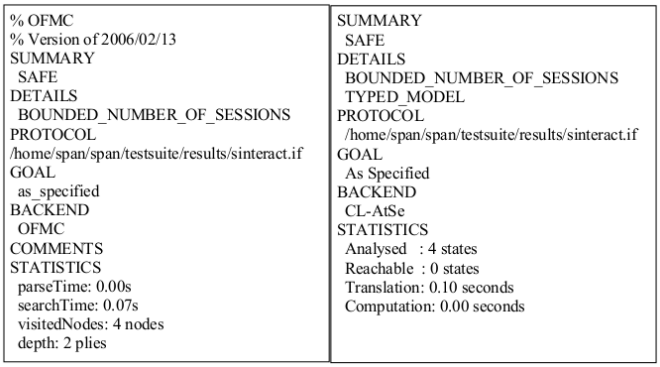


Fig. 2: AVISPA Output

the authentication system normally limits the number of failed logins which is much smaller than |D|, the online password guessing attack is not possible.

C. Simulation For Formal Verification Using AVISPA

The NSE-AA protocol has been simulated using the verification tool AVISPA [22] to prove that it is secure under the Doley Yao model [26]. The protocol is written using High-Level Protocols Specification Language(HLPSL) which allows representation of different independent roles and is simulated using SPAN (Security Protocol ANimator for AVISPA). The tool provides four different backends for the analysis of the protocols: On-the-fly Model-Checker (OFMC), Constraint-Logic-based Attack Searcher (CL-ATSE), SAT-based Model-Checker (SATMC), and Tree Automata based on Automatic Approximations for the Analysis of Security Protocols (TA4SP). The script provides roles of agents U, D and S for the user device, IoT device and TCA respectively along with definitions for session and environment. The detailed script is available in the appendix of our technical report [9]. The script follows all phases of the NSE-AA protocol as given in tables II and IV. The intruder (i) and the communication channels are based on the Dolev Yao model [26]. The first phase consists of mutual authentication as per the steps in table II. The role U generates a virtual identity Id_{VU} and a random N_U and sends a challenge to role D. Role D further generates a virtual identity Id_{VD} and a random nonce N_D, and forwards the challenge to S. Role S verifies the virtual identities and sends tickets for roles D and U, which helps them to mutually authenticate each other and generate a session key K_S to encrypt further communication. The generation of all cryptographic operations is on an SE of the devices. The script has been executed using all the backends of AVISPA. The widely-accepted OFMC and CL-AtSe backends specify if, for a replay attack protection, the valid users can execute the security protocol in the presence of a passive intruder. They also check if for the Dolev-Yao model there is any man-in-the-middle attack possible by the intruder. The protocol script assures the following security goals:

secrecy_of sec_ks: Secret Ks between device U and D.
secrecy_of sec_ua: Actual identity IAu is known to the user device and TCA
secrecy_of sec_da: Actual identity IdAd is known to the IoT and TCA
authentication_on auth_1: Authentication of role D
authentication_on auth_2: Authentication of role U
secrecy_of sec_smlu: Secrecy of logs for role U
secrecy_of sec_smld: Secrecy of logs for role D
secrecy_of sec_pcru: Secrecy of PCR for role U
secrecy_of sec_pcrd: secrecy of PCR for role D
authentication_on smlu_verify: Verify h(SMLU)
authentication_on smld_verify: Verify h(SMLD)

1) Analysis of Simulation Output: We have successfully simulated and prooved the safety of the NSE-AA protocol and the AVISPA tools. Fig 2 shows the results for the OFMC and CL-AtSe backends to illustrate that NSE-AA protocol is secure against man-in-the-middle and replay attacks and is safe. The outputs of SATMC and TA4SP are INCONCLUSIVE since they cannot verify NSE-AA that uses algebraic properties of XOR. The random nonces for mutual authentication are unique for each session, which prevents a replay attack. The session key K_S: *KeyGen(PwbU.PwbD.Nu.Nd.Kud)* requires knowledge of the secret K_{UD}, nonces as well as the user's virtual identity which requires the password known only to the user. The intruder cannot replay and generate a session key. The proximity over NFC as well as difficulty for an adversary to guess the session key makes the MTM attacks difficult. Since the intruder does not know the secrets between the devices as well as the session key which encrypts all further messages including those for attestation, it prevents the MTM attacks. Both devices share their PCR and SML signed by the AIK private key. The remote device verifies the AIK certificate, PCR and SML values and the *IdAikH* for a host H.

V. Performance Analysis and Comparison

Table V compares the NSE-AA protocol with the different attestation schemes for the security and threats requirements. NSE-AA satisfies all security requirements and protects from various threats. Aziz et al. [24] presents an extension of TLS with TPM-based mutual attestation for remote access but lacks proof-of-locality. The MAT protocol by Toegl et al. [11] attests only the remote device. None of these schemes provides user anonymity, secure storage and protection from DDoS and MTM attacks. NSE-AA is the only protocol which supports two-way communication over the NFC interface for security and trust handshake.

Tables VI and VII illustrate the comparison of computation and communication overheads for the different schemes. This paper considers RSA and AES encryption algorithms for asymmetric and symmetric encryption respectively since they are commonly available on the commercially available SEs. We consider the following parameter sizes, nonce: 64 bit, Identity: 64 bit, Stored Measurement Log (SML): 2k bits, PCR: 160 bits [27], RSA Certificate size: 2048 bit, RSA key length 1024 bits, AES key length 1024 bits. For comparison, we assume the following cryptographic overheads: SHA operation time T_H: 7.81* 10^{-4} msec, Symmetric key AES encryption and decryption time T_S: 10.5*10^{-4} msec and Asymmetric RSA time using Chinese remainder theorem T_{AS} (RSA) 12.06 msec as considered by Gope et al. [28]. We assume that the key derivation function time *Tkdf* time is approximately the same as T_H. Time for RSA encryption is incremental as per the input blocks and also pads plaintext if

0098-3063 (c) 2018 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.

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/TCE.2018.2873181, IEEE Transactions on Consumer Electronics

TABLE V: Comparison For Security Requirements

Requirement	[11]	[24]	NSE-AA
S1:Confidentiality	Y	Y	Y
S2:Integrity	Y	Y	Y
S3:Mutual Auth and Attes	N	Y	Y
S4:Privacy	Y	Y	Y
S5:User Anonymity	N	N	Y
S6:NFC Proof-of-locality	Y	N	Y
S7:Secure Storage	N	N	Y
T1:DoS	N	N	Y
T2:Replay	Y	Y	Y
T3:Collusion	N	Y	Y
T4:Parallel session	N	N	Y
T5:Forgery	Y	Y	Y
T6:Platform Impersonation	Y	N	Y
T7:MTM	N	N	Y
T8:Insider Attack	N	N	Y

TABLE VI: Comparison Of Computational Overhead

Phase	[11]	[24]	NSE-AA
Mut. Auth.	0	$2T_H+6T_{AS}$	$22T_H+18T_S$
AIK Cert.	0	$14T_H+64T_S+36T_{AS}$	$20T_H+176T_S+14T_{AS}$
Attest.	$19T_{AS}$	$4T_H+6T_{AS}$	$4T_H+64T_S+12T_{AS}$
Net Comp.	$19T_{AS}$	$20T_H+64T_S+48T_{AS}$	$46T_H+ 258T_S+26T_{AS}$
Time (ms)	229.14	579.00	313.87

TABLE VII: Comparison of Communication Overhead

Phase	[11]	[24]	NSE-AA
Messages	6	10	10
Communication bits	12352	35904	27936

it is less than 1024 bits. The MAT protocol by Toegl et al. [11] has lower overheads as compared to the other schemes. However, it satisfies only limited security requirements. The NSE-AA takes around $46T_H + 258T_S + 26T_{AS}$ computations which is around 313.87 ms as per timing assumptions in Gope et al. [28]. It requires 10 messages with around 27936 bits for communication. The overheads of the NSE-AA protocol are lower as compared to Aziz et al's. [24] scheme since it uses symmetric encryption. Even though the overheads are higher as compared to the Toegl et al.'s scheme, they are acceptable since NSE-AA provides better security as compared to the Toegl et al.'s scheme. The overheads are also acceptable as compared to the overheads of the Toegl et al.'s [11] scheme since it provides better security as compared to the Toegl et al.' s scheme

VI. Prototype for the Framework

We have developed a prototype for the proposed framework using commercial mid-ranged priced SBC (Single Board Computer) for an IoT device with an inbuilt NFC controller that is accessed by a user's mobile device over an NFC tap. The mobile device emulates a software-based HCE card which interacts with the HCE reader application on the NFC controller. Both devices retain secure commercial microSD Card-based SE and TPM-based attestation modules. The SBC operating system resides on the insecure region of the microSD card. It is connected to the NFC controller with a Future

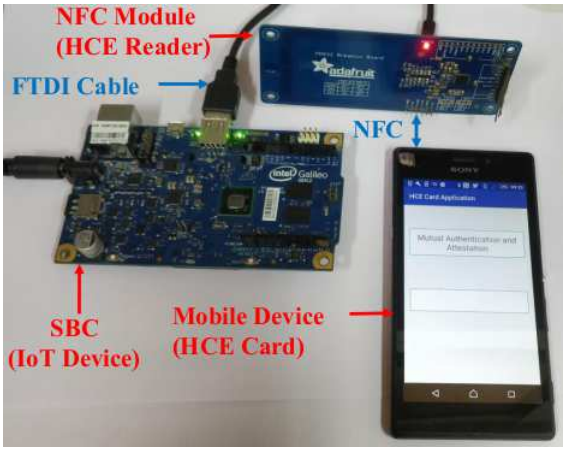


Fig. 3: System Protoctype

Technology Devices International (FTDI) cable. Both devices use the APDU packets to initiate the phases of the NSE-AA protocol as discussed in the previous sections. The devices use the HCE mode to establish automated Bluetooth pairing through the exchange of the Bluetooth address over HCE and direct communication using Radio frequency communication(RFCOMM) sockets. The computational overheads for the prototype are comparable to the overheads given in tables VI and VII.

In the practical world, the consumer IoT devices such as remote surveillance cameras can have a similar inbuilt NFC controller to interface with a user's handheld device over an HCE tap for secure setup of configurations, key management and control of logs. A proximity-based IoT access can ensure proof of locality so that the devices are not prone to become victims to cause cyber attacks such as DDoS attack[1]. The benefits of using HCE with two-way communication and open platform for development makes such an application feasible in the practical work. The further details of the prototype given in our technical report [9].

VII. Conclusion And Future Work

The newly framework for NFC Secure Element-based Mutual Authentication and Attestation for IoT access provides a novel on-demand communication and control of IoT devices with security, privacy, trust and proof-of-locality using the NFC-based HCE mode and secure tamper-resistant SE and TPM modules. HCE provides a feasible and open platform for developers for a bidirectional communication as compared to the other NFC modes used in previous NFC-based security schemes [15], [23], [16], [18], [11]. The NSE-AA protocol presents a combined anonymous mutual authentication and attestation over HCE with improvement in the schemes proposed by Thammarat et al.[14] and Aziz et al.[24].It is proved secure under the ROR model[21] as well as through simulation on the AVISPA tool. It provides better security with satisfactory overheads as compared to the previous schemes by Aziz et al. [24] and Toegl et al. [11]. We can incorporate biometric-based attributes in our mutual authentication phase as suggested by Ghosh et al.[15] and Majumdar et al. [23]. In the future, we can use runtime attestation to verify the dynamic device state such as Control-Flow Integrity (CFI) [29]. Alternately

0098-3063 (c) 2018 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.

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/TCE.2018.2873181, IEEE Transactions on Consumer Electronics

TEE can provide a lightweight and fast attestation [10]. We can also use modifications of lightweight provably secure mutual authentication schemes such as suggested by Chang et al. [25]. Alternate High-speed SEs with extended APDU commands can reduce computations time for NSE-AA. With the improving computational capabilities on mobile devices the framework can be used to securely access simple IoT devices besides the AdIoT devices such as a portable mobile-based healthcard[19] and a mobile-based pickup of a child from school[30].

REFERENCES

[1] "FBI Cyber Bulletin: Distributed Denial of Service Attack Against DNS Host Highlights Vulnerability of Internet of Things Devices," Last Accessed: Aug. 2018. [Online]. Available: https://info.publicintelligence.net/FBI-IoT-DDoS.pdf

[2] J. Granjal, E. Monteiro, and J. S. Silva, "Security for the Internet of Things: A Survey of Existing Protocols and Open Research Issues," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1294–1312, Jan. 2015.

[3] V. Coskun, B. Ozdenizci, and K. Ok, "A Survey on Near Field Communication (NFC) Technology," *Wireless Pers. Commun.*, vol. 71, no. 3, pp. 2259–2294, Aug. 2013.

[4] G. Madlmayr, J. Langer, C. Kantner, and J. Scharinger, "NFC Devices: Security and Privacy," in *Proc. IEEE Int. Conf. Availability, Reliability and Security*, Mar. 2008.

[5] M. Roland, J. Langer, and J. Scharinger, "Security Vulnerabilities of the NDEF Signature Record Type," in *Proc. IEEE Int. Workshop on Near Field Communication*, Feb. 2011, pp. 65–70.

[6] M. Roland and J. Langer, "Comparison of the usability and security of NFCs different operating modes in mobile devices," *Springer Journal Elektrotechnik und Informationstechnik*, vol. 130, no. 7, pp. 201–206, Nov. 2013.

[7] "Secure Element Deployment and Host Card Emulation v10," Last Accessed: Aug. 2018. [Online]. Available: http://simalliance.org/wp-content/uploads/2015/03/Secure-Element-Deployment-Host-Card-Emulation-v1.0.pdf

[8] M. Alattar and M. Achemlal, "Host-based Card Emulation: development, security, and ecosystem impact analysis," in *Proc. IEEE Int. Conf. High Performance Computing and Communications*, Aug. 2014.

[9] "Technical report for implementation of secure NFC-based IoT prototype using mobile devices," Last Accessed: Aug. 2018. [Online]. Available: https://sites.google.com/site/divyashikhasethia/home/secure-nfc-based-iot-access/IoTPrototypeTechReport.pdf

[10] B. Yang, K. Yang, Y. Qin, Z. Zhang, and D. Feng, "DAA-TZ: An Efficient DAA Scheme for Mobile Devices using ARM Trustzone," in *Proc. Inter. Conf. on Trust and Trustworthy Computing*, Aug. 2015.

[11] R. Toegl and M. Hutter, "An approach to introducing locality in remote attestation using near field communications," *The Journal of Supercomputing*, vol. 55, no. 2, pp. 207–227, 2011.

[12] M. A. Bouazzouni, E. Conchond, and F. Peyrard, "Trusted mobile computing: An overview of existing solutions," *Future Gener. Comput. Syst.*, vol. 80, pp. 596–612, Mar. 2018.

[13] U. B. Ceipidor, C. M. Medaglia, A. Marino, S. Sposato, and A. Moroni, "KerNees A protocol for mutual authentication between NFC phones and POS terminals for secure payment transactions," in *IEEE Proc. Inter. ISC Conf. on Information Security and Cryptology*, 2012, pp. 115–120.

[14] C. Thammarat, R. Chokngamwong, C. Techapanupreeda, and S. Kungpisdan, "A secure lightweight protocol for NFC communications with mutual authentication based on limited-use of session keys," in *Proc. IEEE Inter. Conf. on Information Networking (ICOIN)*, Jan. 2015, pp. 133–138.

[15] S. Ghosh, A. Majumder, J. Goswami, A. Kumar, S. P. Mohanty, and B. K. Bhattacharyya, "Swing-pay: One Card Meets All User Payment and Identity Needs: A Digital Card Module using NFC and Biometric Authentication for Peer-to-Peer Payment," *IEEE Consum. Electron. Mag.*, vol. 6, pp. 82–93, 2017.

[16] V. Odelu, A. K. Das, and A. Goswami, "SEAP: Secure and Efficient Authentication Protocol for NFC Applications Using Pseudonyms," *IEEE Trans. Consum. Electron.*, vol. 62, no. 1, pp. 30–38, Apr. 2016.

[17] C.-H. Hung, Y.-W. Bai, and J.-H. Ren, "Design and implementation of a door lock control based on a Near Field Communication of a smartphone," in *IEEE Inter. Conf. Consum. Electr.*, Jun. 2015.

[18] J. Xu, K. Xue, Q. Yang, and P. Hong, "PSAP: Pseudonym-Based Secure Authentication Protocol for NFC Applications," *IEEE Trans. Consum. Electron.*, vol. 64, pp. 83–91, Feb. 2018.

[19] D. Sethia, D. Gupta, and H. Saran, "Security framework for portable NFC mobile based health record system," in *IEEE Proc. Inter. Conf. Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct. 2016.

[20] L. Yang, J.-F. Ma, and Q. Jiang, "Mutual Authentication Scheme with Smartcards and Password under Trusted Computing," *Inter. Journal of Network Security*, vol. 14, pp. 155–162, 2012.

[21] M. Abdalla, P.-A. Fouque, and D. Pointcheval, "Password-Based Authenticated Key Exchange in the Three-Party Setting," in *Inter. Workshop on Public Key Cryptography*, Jan. 2005, pp. 65–84.

[22] A. Armando, D. Basin, Y. Boichut, Y. C. L. Compagna, J. Cuellar, P. H. Drielsma, P. C. Heám, O. Kouchnarenko, J. Mantovani, S. Mödersheim, D. von Oheimb, M. Rusinowitch, J. Santiago, M. T. Viganò, and L. Vigneron, "The AVISPA Tool for the Automated Validation of Internet Security protocols and Applications," in *Inter. Conf. on Computer Aided Verification*, Jul. 2005, pp. 281–285.

[23] A. Majumder, J. Goswami, S. Ghosh, R. Shrivastawa, S. P. Mohanty, and B. K. Bhattacharyya, "Pay-Cloak: A Biometric Back Cover for Smartphones: Facilitating secure contactless payments and identity virtualization at low cost to end users." *IEEE Consum. Elect. Mag.*, vol. 6, no. 2, pp. 78–88, Apr 2017.

[24] N. Aziz, N. Udzir, and R. Mahmod, "Extending TLS with Mutual Attestation for Platform Integrity Assurance," *J. Commun.*, vol. 9, pp. 63–72, Jan. 2014.

[25] C.-C. Chang and H.-D. Le, "A Provably Secure, Efficient, and Flexible Authentication Scheme for Adhoc Wireless Sensor Networks," *IEEE Trans. Wireless Communications*, vol. 15, pp. 357–366, 2016.

[26] D. Dolev and A. Yao, "On the security of public key protocols," *IEEE Trans. Information Theory*, vol. 29, no. 2, pp. 198–208, Mar. 1983.

[27] S. Delaune, S. Kremer, M. D. Ryan, and G. Steel, "Formal analysis of protocols based on TPM state registers," in *IEEE Proc. Computer Security Foundations Symposium*, Aug. 2011.

[28] P. Gope and T. Hwang, "Lightweight and energy-efficient mutual authentication and key agreement scheme with user anonymity for secure communication in global mobility networks," *IEEE Syst. J.*, vol. 10, no. 4, pp. 1370–1379, Apr. 2015.

[29] T. Abera, N. Asokan, L. Davi, F. Koushanfar, A. Paverd, A.-R. Sadeghi, and G. Tsudik, "Things, trouble, trust: On building trust in iot systems," in *Proc. IEEE Conf. Design Automation Conference*, Jun. 2016.

[30] Y.-W. Bai, C.-N. Fu, and J.-H. Yang, "Using NFC tags and smartphones to design a reliable mechanism to pick a child up from school," in *IEEE Inter. Conf. Consum. Electr.*, Jan. 2018.

Divyashikha Sethia received B.Tech degree in Computer Science and Engineering from the Maharaja Sayajirao University of Baroda in 1997 and M.Tech degree in Computer Science and Engineering from Indian Institute of Technology (IIT) Delhi in 2006. She is an Assistant Professor in Department of Computer Science and Engineering at Delhi Technological University(DTU), Delhi. Her research interests are Distributed systems, mobile computing, telemedicine and security.

Daya Gupta received M.Sc.(Computer Science) Degree at Indian Institute of Technology(IIT) Delhi, and PhD in Computer Science from the University of Delhi. She is Professor in Department of Computer Science and Engineering at Delhi Technological University(DTU). Her research interests are Information Security, Requirements Engineering, Swarm Intelligence, Nature Inspired Algorithm, Data Warehouse and Data Mining.

Huzur Saran received B.Tech degree in Electrical Eng., Indian Institute of Technology(IIT) Delhi in 1983 and PhD in the Computer Science University of California, Berkeley in 1990. He is currently a professor in Department of Computer Science at IIT Delhi and Heads, the Amar Nath and Shashi Khosla School of Information Technology, at IIT Delhi. His research is in the areas of Computer Systems, Security and Algorithms.

0098-3063 (c) 2018 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.

Vacuum 159 (2019) 173–181

Contents lists available at ScienceDirect

Vacuum

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

Numerical simulations: Toward the design of 18.6% efficient and stable perovskite solar cell using reduced cerium oxide based ETL

Rahul Pandey[a,b], Anand Prakash Saini[b], Rishu Chaujar[b,*]

[a] Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
[b] Department of Applied Physics, Delhi Technological University, New Delhi, India

ARTICLE INFO

Keywords:
Cerium oxide
Polaron
Perovskite
Solar cell
Simulations
Efficiency

ABSTRACT

This work presents an extensive study of one of the potential alternatives in electron transport material (ETM) for perovskite solar cell (PSC). Reduced cerium oxide (CeO_x) as an electron transport layer (ETL) is studied for different composition of oxygen and analysis with [6,6]-phenyl-C61-butyric acid methyl ester (PCBM) interface layer thickness if stacked in between CeO_x ETL and perovskite layer has also been done. A detailed investigation has been carried out, incorporating the mobility of CeO_x due to small polaron hopping mechanism at the different composition of x, the effect of variation in doping concentration for CeO_x and PCBM have also been studied and the optimum efficiency 18.2% is obtained for PSC. Furthermore, to improve the shortcomings in terms of hysteresis and moisture stability, a device is proposed incorporating carbon nanotube (CNT) layer stacked between perovskite and hole transport layer (HTL) which shows 18.6% conversion efficiency.

1. Introduction

In order to find an alternative to the conventional source of electricity supply and for the conservation of environmental degradation, climate change and depletion of fossil fuels, an efficient way to harness the immense potential of solar energy has to be discovered. The most researched material in the history of solar cells is the crystalline silicon (c-Si) material which is covering about 90% of the current market share in the photovoltaic industry [1]. However, in the past 15 years, the efficiency of the c-Si-based solar cell has shown only marginal progress and according to the studies, c-Si-based solar cell has the potential to achieve the theoretical efficiency of 29.4% [2,3]; whereas, current record conversion efficiency of about 26.3% is obtained experimentally [4]. To further increase the efficiencies keeping module cost low, cost-effective energy efficient modules are required. Among the existing materials for solar cell devices such as c-Si, amorphous silicon hydrogenated (a-Si: H), copper indium gallium selenide (CIGS), cadmium telluride (CdTe), organic and perovskite, the organic-inorganic hybrid perovskite-based solar cell is emerged as the rapidly growing solar cell with improved efficiency in a very short period of time and expected to exceed the efficiency of c-Si in the near future. The power conversion efficiency (PCE) of the PSCs is escalated from 3.8% in 2009 to more than 22% in 2017 [5,6], owing to its special features like high absorption, low production cost, ease of fabrication and high PCE [7–9] and hence, it has become commercially attractive and also a promising scope of research in the field of solar cells. Conventionally, to achieve the optimum performance, the PSCs consists the absorber layer, i.e., perovskite layer stacked between ETL and HTL for fast charge transport [10]. In addition to that, for more smooth electron injection after the charge separation process, the hole blocking layer and electron blocking layer is used at ETL and HTL respectively [11–14]. The characteristics of ETL involves proper band alignment, high transmittance and ETL is mostly affected by various factors such as charge mobility, trap states, morphology, and interface properties [15,16]. The electron transport material is overwhelmingly dominated by TiO_2 as the electron injection rate from the perovskite absorber layer to TiO_2 (ETL) is very rapid, thereby producing the high efficiency. However, TiO_2 is typically manufactured at a very high temperature, and electrons recombine with holes at a high rate [5,6,17,18]. Thus, in order to solve these problems and obtain a low temperature processed, low manufacturing cost, and high PCE material, various other ETM were introduced such as ZnO, SnO_2, SiO_2, ZrO_2, and WO_x, etc. [15].

ZnO, owing to its relatively high electron mobility, more efficient charge extraction, high transparency, and versatile nanostructure is used as an ETL for low cost and easily manufactured PSCs [19,20]. Mesoporous TiO_2 nanostructures in conventional PSCs is also replaced by ZnO nanorods [21,22]. Recently, Shang-Hsuan Wu et al. showed that the fluorine doped tin oxide (FTO)/Zinc oxide (ZnO) and FTO/ZnO/

* Corresponding author.
E-mail address: Chaujar.rishu@dtu.ac.in (R. Chaujar).

https://doi.org/10.1016/j.vacuum.2018.10.033
Received 15 December 2017; Received in revised form 22 August 2018; Accepted 15 October 2018
Available online 18 October 2018
0042-207X/ © 2018 Elsevier Ltd. All rights reserved.

AZO (Al-Doped ZnO) based PSC have achieved PCE of 12.25% and 16.07% respectively [23]. However, along with chemical instability, ZnO as an ETL based PSC does not exhibit long-term stability [24]. Another emerging ETM is SnO_2 which has several advantages over TiO_2 and ZnO such as anti-reflection and deeper conduction band, that facilitates the charge transport from absorber layer to ETL and reduces charge accumulation nevertheless hysteresis is observed in these devices [25]. Therefore for relatively greater stability, lesser hysteresis and high PCE device should be proposed.

To solve aforesaid limitations, a CeO_x based ETM is introduced possessing various advantages over other ETMs such as high ionic conductivity [26], variable wide band gap (3.0–3.6 eV), excellent optical and dielectric properties [27], which makes it a promising alternative for ETL in PSCs. From the past research [28,29], it is observed that CeO_x exhibits small polaron hopping mechanism, the n-type conductivity takes place by small polaron transport. A small polaron is defined by Tuller et al. [28] as a defect created due to the displacement of adjacent atoms or ions when an electronic carrier is trapped at a given site. The entire defect (carrier plus distortion) then migrates by an activated hopping mechanism [29]. This paper incorporates the CeO_x mobility due to polaron hopping for different composition of oxygen, i.e., for different values of x in CeO_x at 300 K, which gave our simulation a realistic approach for the overall analysis of the device in technology computer-aided design (TCAD) Silvaco ATLAS. To the best of our knowledge this type of investigation, incorporating mobility variation due to polaron hopping for a CeO_x and PCBM layer ETL based PSC and variation of other aspects having an impact on photovoltaic parameters has never been reported before. However, the similar experimental work is reported by Xin Wang et al. [18] which is calibrated and simulated in this work for further study and optimization, to achieve an improved result.

2. Overview

PSC as reported in Fig. 1 consists of n-type CeO_x (ETL) with doping density of $5 \times 10^{19}\,cm^{-3}$, n-type PCBM (Interface layer for stability) with doping density of $1 \times 10^{17}\,cm^{-3}$, p-type perovskite layer (active layer) with doping density of $2.14 \times 10^{17}\,cm^{-3}$ and p-type Spiro-MeOTAD (HTL) with doping density of $3 \times 10^{18}\,cm^{-3}$ the corresponding thickness is 60 nm, 10 nm, 250 nm and 150 nm respectively. The simulated device is identical to the device reported in the literature [18]. Silvaco ATLAS device simulator is used to study the optoelectronic behavior of the device. Various literature has shown that CeO_x, which is a low mobility material exhibits small polaron hopping mechanism for carrier transport [28,29]. In the transportation of a small polaron from one site to another, the hopping of both electron and the polarized atomic configuration are required. When single Ce^{3+} ions are formed, lowering in energy due to relaxation of the adjacent O^{2-} ions takes place, which leads to the polarization, that in turn serves to trap

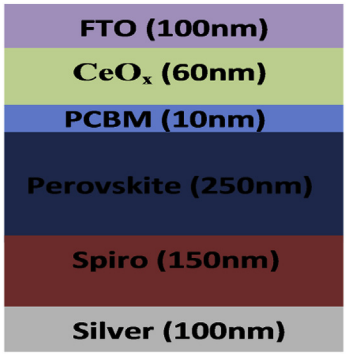


Fig. 1. Simulated device structure similar to experimental device as reported in Ref. [18].

the electron in the localized configuration constituting the small polaron [29]. Therefore, a quantitative simulation study is required to explain all the aspects of the changes due to the composition which is not described in any literature to the best of our knowledge as far as the study of solar cells is considered. Hence, CeO_x mobility due to polaron hopping for different composition of oxygen at 300 K is incorporated, that gave our simulation a realistic approach for the overall investigation of the device. Mobility in CeO_x for each value of x is calculated at 300K using the formulas as reported by Tuller et al. [28] and shown in Equation (5). Material parameters used in the analysis are indicated in Table 1, whereas models parameters are identical as published in our previous work [30]. The optical properties, i.e., imaginary index (k) of CeOx, PCBM and perovskite are indicated in Fig. 2 (a), Fig. 2b and Fig. 2(c) respectively and the real index (n) for CeO_x, Perovskite, and SpiroMeOTAD is reported in Fig. 2d, e and f respectively. All the reported optical properties of different materials are obtained from literature [31–34]. Results reported in this work covers calibration, composition-wise photovoltaic parameters, the impact of PCBM thickness, doping variation in ETL along with PCBM, device optimization and projection of CNT based optimized PSC which consist of CeO_X as ETL layer.

3. Equations

To calculate the mobility at each value of x, i.e., for each composition of oxygen in CeO_x. The following equations are used. The small polaron in CeO_x is expressed as $[Ce'_{Ce}]$, i.e., the concentration of Ce^{3+} ions cm^{-3} and the electronic conductivity is given by:

$$\sigma_e \equiv [Ce'_{ce}]e\mu \tag{1}$$

The drift mobility takes the form as provided in Refs. [28,29]:

$$\mu = (1 - C)ea^2\Gamma/kT \tag{2}$$

Where 'C' is the fraction of the site which contains an electron and given by:

$$C = n/N \tag{3}$$

Jump rate of polaron from one site to another, Γ

$$\Gamma = P\vartheta_0 \exp\left(-\frac{E_h}{kT}\right) \tag{4}$$

Finally, the mobility equation takes the form as shown below

$$\mu = \left(\frac{B}{T}\right)\exp\left(\frac{-E_h}{kT}\right) \tag{5}$$

Where, μ = drift mobility, e = electronic charge, a = lattice parameter, c = fraction of site which contains an electron, N = number of available sites per unit volume, n = number of electrons, Γ = jump rate of polaron from one site to another, B = same meaning as in Ref. [28], P = Probability of electron transfer and ϑ_o = Phonon frequency.

4. Results

4.1. Impact of oxygen composition in CeO_x on photovoltaic performance

Result section includes the calibration and analysis of oxygen composition along with the impact of PCBM thickness, doping variation in ETL and PCBM, optimization and projection of CNT based PSC for moisture stability. Initially influence of oxygen composition, x has been carried out on the photovoltaic performance of the device. As discussed in previous section mobility in CeO_x is calculated using Eqs. (1)–(5) as suggested by Tuller et al. [28], and the same is reported in Table 2. Data reported in Table 2 shows, reducing oxygen composition from 1.992 to 1.960 results in an increment in mobility, whereas decreasing x beyond 1.960 shows significant degradation in mobility of CeO_x. The charge carriers in CeO_x, are proposed to be small-polaron formed by

Table 1
Material parameters used in simulations.

Material		$CeO_{1.96}$ (ETL) [18,31]	PCBM [35]	$CH_3NH_3PbI_3$ [30,36]	Spiro (HTL) [30,36]	CNT [37]
Dielectric Constant		23	3	18 [38]	3	3.4
Band Gap (eV)		3.5	2.1	1.55	2.91	1.1
Electron Affinity (eV)		4.6	4.3	3.93	2.2	4.27
Thickness (nm)		60	5	250	150	30
Electron, Hole mobility ($cm^2/V{\cdot}s$)		Table 2 [28] and 1	1×10^{-3}, 1×10^{-4}	8.2, 8.2 [39]	1×10^{-4}, 1×10^{-4}	8×10^4, 2×10^3
Conduction band density of states (DOS) (cm^{-3})		1×10^{20}	1×10^{22}	2.5×10^{20}	2.5×10^{20}	5×10^{16}
Valence band density of states (DOS) (cm^{-3})		2×10^{21}	1×10^{22}	2.5×10^{20}	2.5×10^{20}	6×10^{17}
Conduction tail states	$N^{c\text{-}tail}$ A (cm^{-3})	10^{14}	10^{10}	10^{15}	10^{14}	N/A
	$E^{c\text{-}tail}$ A (eV)	0.01	0.015	0.01	0.01	N/A
Valence tail states	$N^{v\text{-}tail}$ D (cm^{-3})	10^{14}	10^{10}	10^{15}	10^{14}	N/A
	$E^{v\text{-}tail}$ D (eV)	0.01	0.015	0.01	0.01	N/A
Acceptor-like (A) dangling bond states	N^{db} A (cm^{-3})	10^{17}	10^{14}	10^{16}	10^{16}	N/A
	E^{db} A (eV)	1.1	1.2	1.1	1.1	N/A
	σ^{db} A (eV)	0.1	0.1	0.1	0.1	N/A
Donor-like (D) dangling bond states	N^{db} D (cm^{-3})	10^{17}	10^{14}	10^{16}	10^{16}	N/A
	E^{db} D (eV)	1.1	1.2	1.1	1.1	N/A
	σ^{db} D (eV)	0.1	0.1	0.1	0.1	N/A
Capture cross-section for electrons and holes in the tail distribution of acceptor-like states (cm^2)		10^{-17}, 10^{-15}	10^{-10}, 10^{-9}	10^{-18}, 10^{-16}	10^{-17}, 10^{-15}	N/A
Capture cross-section for electrons and holes in the tail distribution of donor-like states (cm^2)		10^{-15}, 10^{-17}	10^{-9}, 10^{-10}	10^{-16}, 10^{-18}	10^{-15}, 10^{-17}	N/A

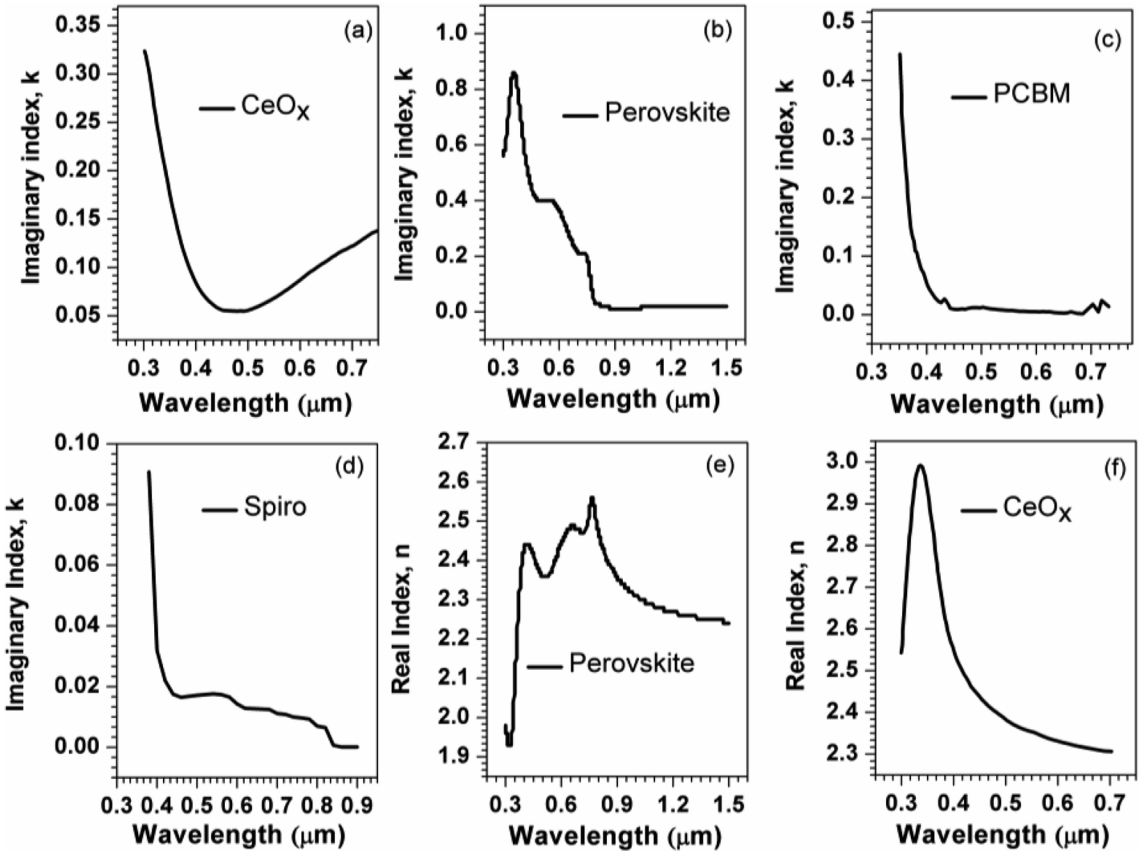


Fig. 2. Optical properties used in simulations, k values for CeO_x [31], perovskite [33,34], PCBM [35], and Spiro-MeOTAD [30,34] is reported in (a), (b), (c) and (d) respectively. Real index, n for perovskite [33,34] and CeO_x [31] is reported in (e) and (f) respectively.

localization of electrons at cerium sites and the charge transport process is proposed to occur by a hopping mechanism. The hopping mobility is activated at hopping activation energy or activation energy (E_h). Reducing the composition of oxygen in CeO_x increases the activation energy for hopping and hence E_h increases with the decrease of x in CeO_x [28,40].

Moreover, photovoltaic parameter of the device is obtained with a different composition based CeO_x ETL and same is reported in Fig. 3a-e. Fig. 3a shows the illuminated J-V curve, and PV parameters such as J_{SC}, V_{OC}, FF, and PCE is depicted in Fig. 3b-e, respectively. The result shows significant improvement in terms of fill factor (FF) for $CeO_{1.960}$ (ETL) based device compared to $CeO_{1.880}$ (ETL) based device; this is attributed to the improvement in the mobility as reported in Table 2. Moreover, open circuit voltage (V_{OC}) and short circuit current density

Table 2
Mobility of CeO_2 with different oxygen composition, x [28].

x in CeO_x	E_h (eV)	B (K cm^2/V.s)	$\mu(10^{-7})$ ($cm^2.V^{-1}s^{-1}$) at 300K
1.992	0.40	393.6	2.5136
1.960	0.38	266.8	3.6925
1.910	0.42	237.8	0.7008
1.880	0.44	273.6	0.3720
1.810	0.46	204.6	0.1283
1.800	0.48	240.0	0.0694

(J_{SC}) remains unaffected. FF of the device is mainly affected by series resistance (R_s), shunt resistance (R_{sh}), the recombination current in the depletion region of the cell, and reverse saturation current of the junction. The R_s is composed of the resistance of the different semiconductor layers (series resistance of semiconductor depends on carrier mobility and doping [41]) used in the cell in addition to the metal-semiconductor contacts (Schottky or Ohmic), and the shunt resistance arises because of the leakage across the junction. The main impact of R_s is a reduction in the FF, although excessively high values may also reduce the short-circuit current [41]. R_s does not affect the solar cell at V_{oc} since the overall current flow through the solar cell is zero. Moreover, higher recombination in the depletion region may overweight the recombination in the regions outside the depletion thus reducing the ideality factor of the device and consequently reducing the squareness of the J-V characteristics and hence the FF. The last parameter, i.e., the reverse saturation current decreases the squareness of the J-V curve and hence it is anticipated to have low reverse saturation current in order to increase the FF of the device.

The main function of ETL is to collect the electrons from the electron-hole (e-h) pairs approaching the interface of CeO_x/PCBM/$CH_3NH_3PbI_3$. Thus, to perform the electron collection function, ETL need to be a good electron conductor. In addition, it must prevent stacking up of the electrons (which occurs due to low electron mobility) near the interface. Otherwise, free electrons will recombine with free holes on both sides of the interface which eventually reduce the collected current. Another challenging effect caused by the low electron mobility is increased resistance that consequently drops a large voltage for the same cell current. Thus, leads to a decrease in FF, due to the increased series resistance of the device. Fig. 3a and Fig. 3d shows decreasing x values from 1.960 reflects a significant reduction in FF which is attributed to increased series resistance as discussed above. Superior photovoltaic parameters: J_{SC} (24.9 mA cm^{-2}), V_{OC} (993 mV), FF (67%) and PCE (16.5%) is obtained with $CeO_{1.960}$ (ETL) compared to any other x values as shown in Fig. 3a-e. Moreover, contour representation along with data plot of recombination rates has also been examined inside the perovskite layer for two devices-one with $CeO_{1.96}$ and another with $CeO_{1.88}$ is depicted in Fig. 4. $CeO_{1.88}$ based device has comparatively low mobility which results in higher recombination rate near the electron collecting interface (depletion region) and hence decreases the ideality factor, i.e., the squareness of the J-V curve reported in Fig. 3a.

Additionally, energy band diagram along with contour as well as data plot of the potential profile is also analyzed and is reported in Fig. 5a-e. Conduction band (E_C) along with electron Fermi energy (E_{fn}) is reported in Fig. 5a and Fig. 5b for $CeO_{1.96}$ and $CeO_{1.88}$ based ETL respectively. Potential profile near electron collecting interface for $CeO_{1.96}$ and $CeO_{1.88}$ based ETLs are reported in Fig. 5c, whereas, the contour representation of the same is reported in Fig. 5d and e respectively. The E_{fn} and potential profile reported in Fig. 5b and Fig. 5c, respectively shows large gradient due to low mobility in case of $CeO_{1.88}$ based device and hence it is difficult for electrons to make transitions from perovskite to ETL, whereas, E_{fn} and potential profile of $CeO_{1.96}$ based device shows, it is easier for the electrons to make a transition from perovskite/ETL interface to ETL as shown in Fig. 5a and Fig. 5c, respectively.

4.2. Impact of PCBM thickness and doping on photovoltaic performance

The device structure as reported in Fig. 1 contains PCBM as an interface layer between CeO_x and perovskite layer. Literature shows that stacking PCBM between ETL and perovskite significantly suppresses the hysteresis losses [18,42] and does not hamper the electron collection efficiency of ETL layer. Moreover, the presence of PCBM substantially reduces trap-assisted charge recombination at the perovskite/ETL interfaces as well as in the perovskite absorber layer [18]. Results reported in the previous literature show superior electron transport layer when PCBM is doped with fluorene-based polyelectrolytes [43] and Graphdiyne [44]. This section is dedicated to the study of PCBM, where the impact of PCBM thickness and doping on photovoltaic performance is examined. Initially, a comprehensive study has been carried out to examine the effect of PCBM thickness, and n-type PCBM with doping 1×10^{17} cm^{-3} is used for the same. Fig. 6a-d shows increasing the thickness of PCBM from 5 nm to 70 nm significantly degrades the conversion efficiency from 16.7% to 11.6%. The degradation of efficiency is mainly attributed to the reduction in FF from 67% to 48%. It is noted that used doping is 1×10^{17} cm^{-3} which do not provide a significant improvement in conductivity of PCBM and hence thicker PCBM

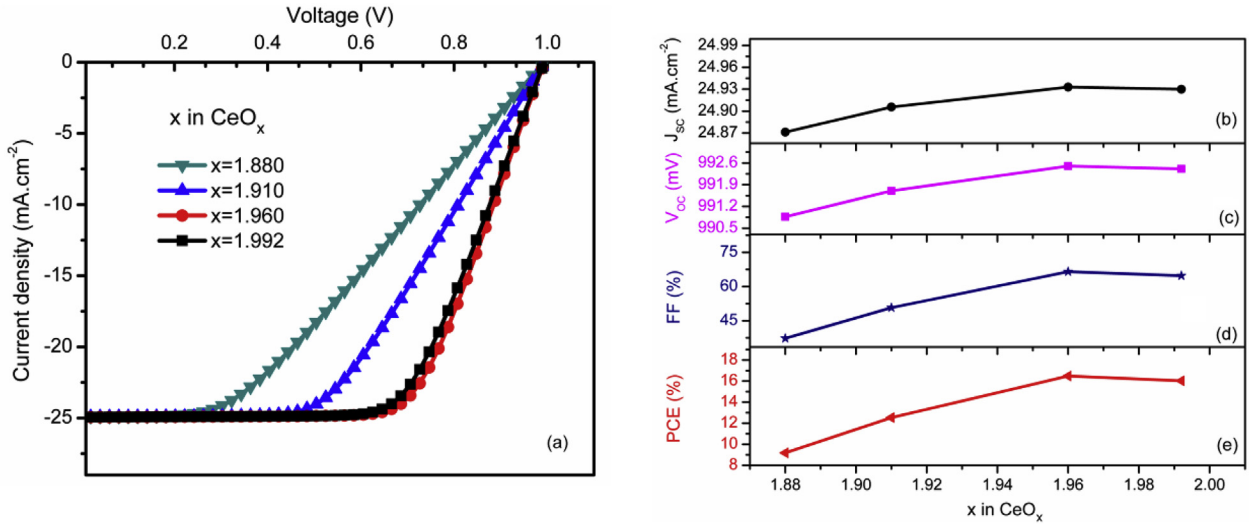


Fig. 3. (a) Current density-voltage (J–V) curves (b) J_{SC}, (c)V_{OC}, (d) FF, and (e) PCE with different oxygen composition in CeO_x (ETL).

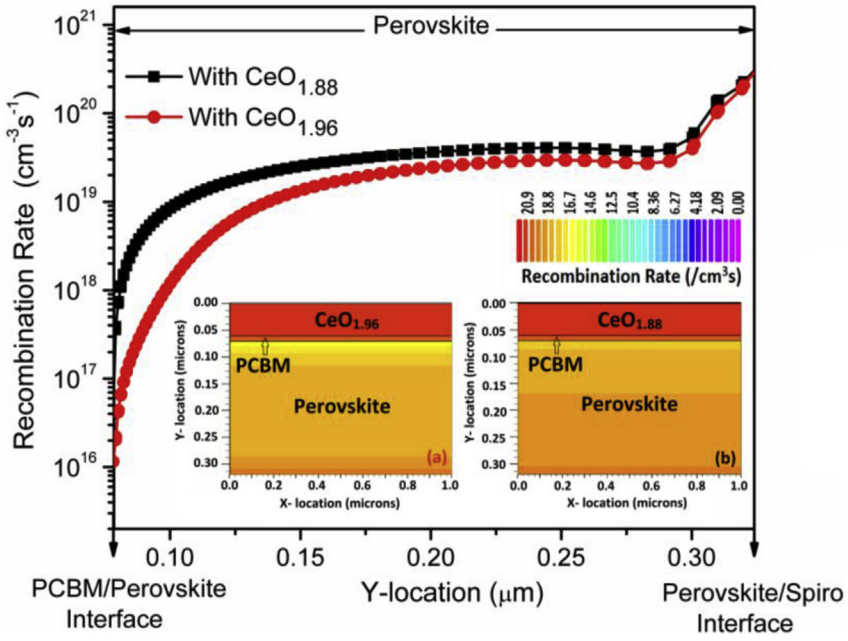


Fig. 4. Recombination rate near the electron collecting interface. *INSET* shows the contour representation of recombination rate (a) with $CeO_{1.96}$ and (b) with $CeO_{1.88}$. The analysis is carried out under short circuit with illumination ON.

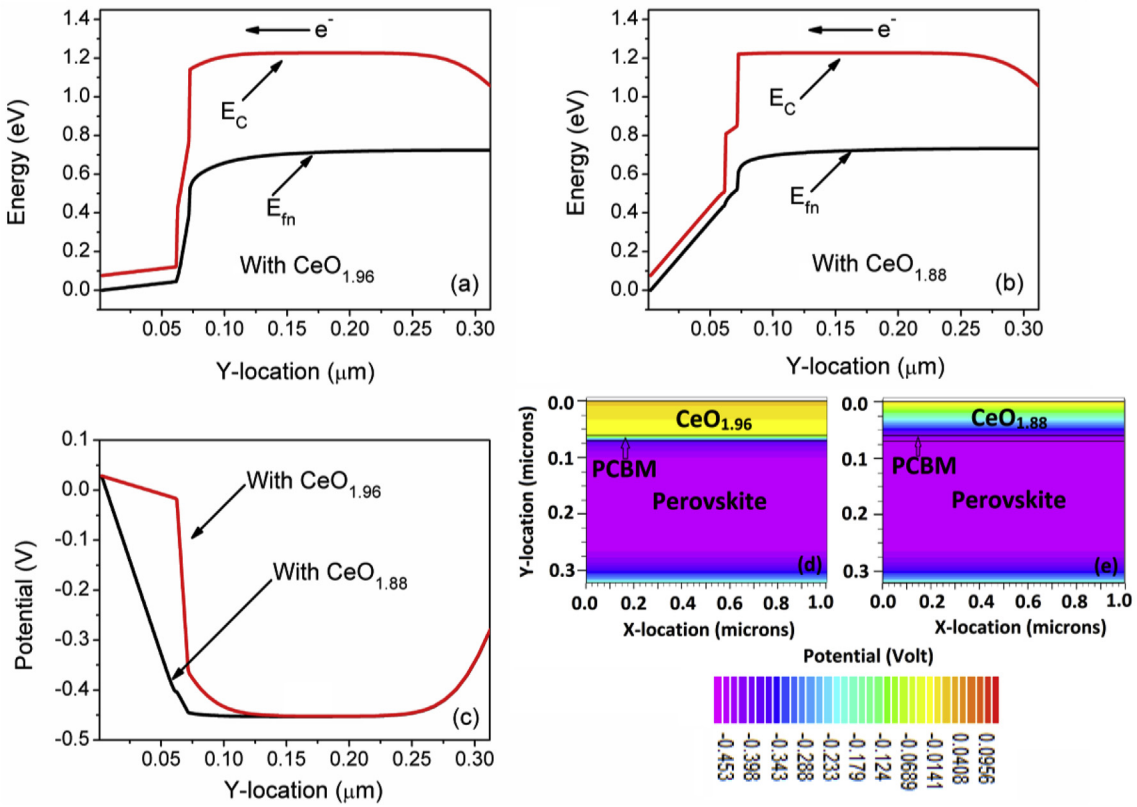


Fig. 5. Energy band diagram (a) $CeO_{1.96}$/PCBM/perovskite interface, (b) $CeO_{1.88}$/PCBM/perovskite interface, and (c) potential profile of respective devices. (d) and (e) shows contour representation of the potential, to obtain the results small region near CeO_x/PCBM/Perovskite is exaggerated. The analysis is carried out under short circuit with illumination ON.

layer would lead to higher series resistance [45] and undesired recombination in perovskite, shown in Fig. 6e. Moreover, thicker PCBM can also be employed, provided its conductivity is high. Fig. 6e and its inset show a reduction in the squareness of J-V curve for 70 nm thick PCBM with doping density $1 \times 10^{17}\,cm^{-3}$. However, when doping density is increased to $5 \times 10^{17}\,cm^{-3}$, the recombination rate and J-V curve of 70 nm thick PCBM based device is close to 5 nm thick PCBM based device. This improvement in J-V curve and recombination is attributed to the increase in the conductivity of PCBM, which eventually increases the FF of the device.

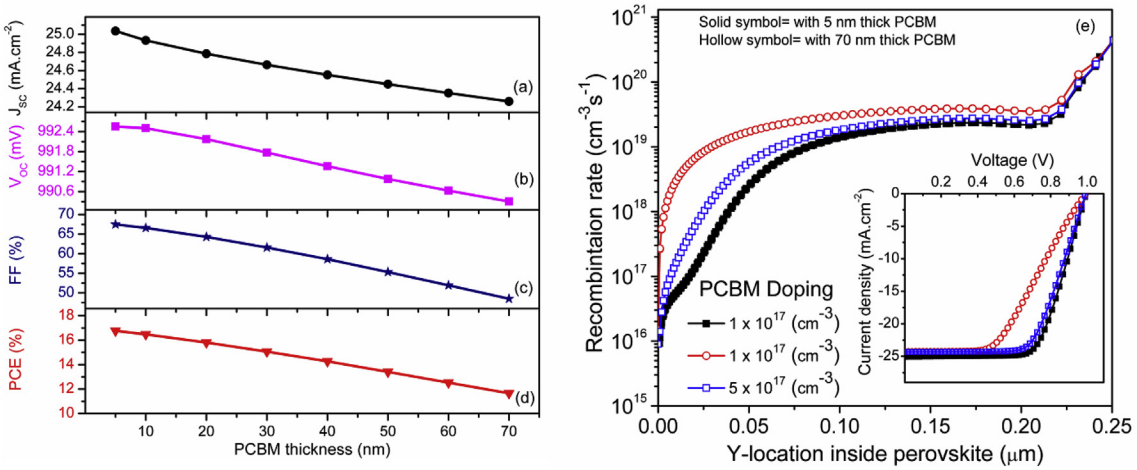


Fig. 6. Effect of PCBM thickness on photovoltaic parameters: (a) J_{SC}, (b)V_{OC}, (c)FF and (d)PCE. (e) Recombination rate inside the perovskite layer with different doping of 5 nm and 70 nm thick PCBM layer. *INSET* shows the J-V curve of the device with different doping of 5 nm and 70 nm thick PCBM layer.

Moreover, Fig. 6e shows the recombination rate is lower at PCBM/perovskite interface, increases in the bulk region of perovskite and become maximum near perovskite/HTL interface. Under short circuit condition, electrons from generated e-h pairs are collected at the interface of ETL/perovskite (due to the associated electric field at the junction). The aforesaid collection is completed due to the effect of interface field. However, e-h pairs that are generated outside the ETL/perovskite interface will be collected via diffusion (due to neutral region) towards the field region of ETL/perovskite interface. This attributes to the lower recombination rate near the ETL/perovskite interface which increases outside the ETL/perovskite interface. Results reported in this section conclude that a thin PCBM layer is desired for undoped or lightly doped PCBM based device, and thicker PCBM can also be employed by increasing the doping.

4.3. Influence of doping variation in $CeO_{1.96}$

This subsection presents a detailed analysis of doping density in $CeO_{1.96}$ and its impact on photovoltaic performance. Performance of the device with different $CeO_{1.96}$ doping is examined and reported in Fig. 7a-e. Donor doping density in $CeO_{1.96}$ is varied from $1 \times 10^{17}\,cm^{-3}$ to $1 \times 10^{20}\,cm^{-3}$, and the corresponding J-V curve is reported in Fig. 7a, whereas J_{SC}, V_{OC}, FF, and PCE is reported in Fig. (b–e), respectively. Theoretical mobility as calculated and reported in Table 2 shows very low mobility, i.e., $3.69 \times 10^{-7}\,cm^2\,V^{-1}\,s^{-1}$, and hence doping density plays a significant role to enhance the conductivity of $CeO_{1.96}$. Fig. 7a shows significant degradation in J-V curve as the doping density decreases below $2.5 \times 10^{19}\,cm^{-3}$. This is attributed to higher series resistance due to low mobility and doping level below $2.5 \times 10^{19}\,cm^{-3}$ is not able to mitigate the resistive losses. Fig. 7b-e shows, conversion efficiency falls from 16.8% to 2.2% as doping density is varied from $5 \times 10^{19}\,cm^{-3}$ to $1 \times 10^{17}\,cm^{-3}$. Moreover, till doping density of $5 \times 10^{18}\,cm^{-3}$, only FF is reduced, nevertheless, reducing the doping density below $5 \times 10^{18}\,cm^{-3}$ results in degradation of FF as well as collected current density, as shown in Fig. 7a and Fig. 7b&d. The conductivity (which directly depends on mobility and doping density of the layer) of ETL has an appreciable effect on the performance of the device. The basic role of ETL is to provide a low resistive path to electrons, in order to collect the electrons. Also, it must not allow the stack up of the electrons near the interface. However, if this occurs due to low conductivity, free electrons recombine with free holes on both sides of the interface and thus lead to a reduced collected current. In addition to this, having low conductivity results in increased series resistance of the device which leads to decrease in the FF. Impact of low conductivity is twofold, i.e., it decreases the both FF and current density as shown in Fig. 7a.

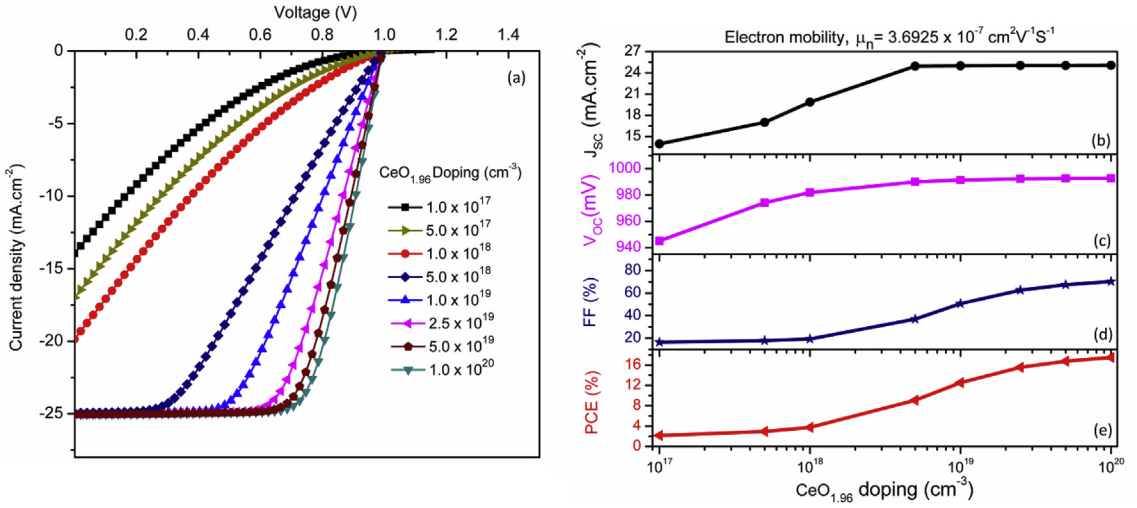


Fig. 7. (a) J-V curve of the device with different doping density in $CeO_{1.96}$ (ETL) layer. Photovoltaic parameters: (b) J_{SC}, (c) V_{OC}, (d) FF and (e) PCE with different doping density. The analysis is done with $CeO_{1.96}$ mobility $3.6925 \times 10^{-7}\,cm^2\,V^{-1}.\,s^{-1}$.

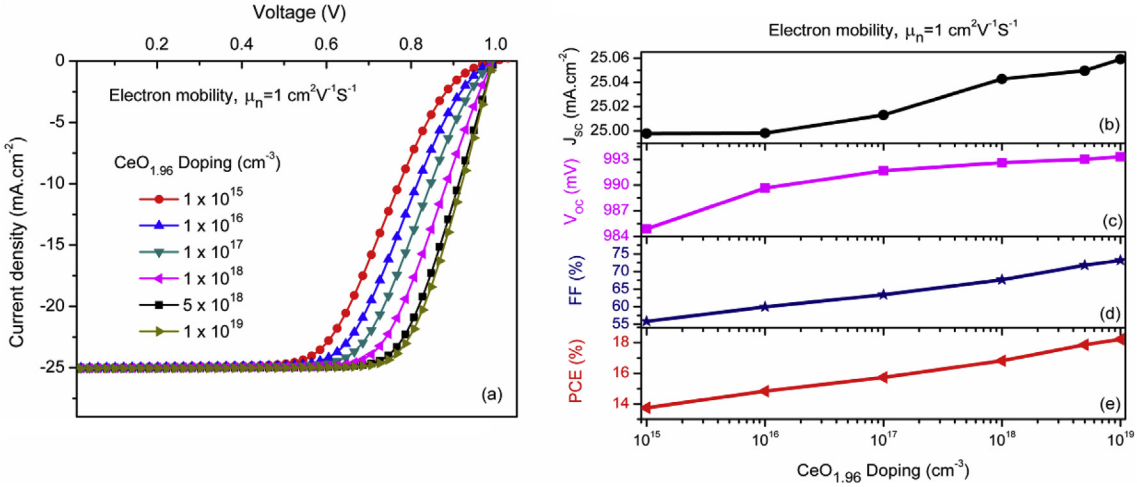


Fig. 8. (a) J-V curve of the device with different doping density in $CeO_{1.96}$ (ETL) layer. Photovoltaic parameters: (b) J_{SC}, (c) V_{OC}, (d) FF and (e) PCE with different doping density. The analysis is done with $CeO_{1.96}$ mobility 1 $cm^2 V^{-1} s^{-1}$.

Moreover, theoretical mobility as reported in Table 2 is very low, and hence heavily doped $CeO_{1.96}$ is needed to perform the function of electron collection. However, most of the recent work shows that it is possible to achieve higher mobility and superior electron collection using high-quality CeO_x film [18,45] and hierarchically mesostructured doped CeO_2 [46]. CeO_x material is reported with promising electrical properties such as high electron mobility of (10^3 $cm^{-2}kV^{-1}s^{-1}$) and suitable energy level for efficient electron extraction in perovskite solar cells [18,46,47]. This section deals with the comprehensive study of high mobility $CeO_{1.96}$ (ETL) based perovskite solar cell. Fig. 8a shows the J-V curve of the device with different doping density and the PV parameters such as J_{SC}, V_{OC}, FF, and PCE is summarised in Fig. 8b-e, respectively. Comparing to the efficiency (9.1%) of the previous section at doping density 5×10^{18} cm^{-3}, the conversion efficiency 17.8% is achieved at the same doping density. The improvement in performance is due to increase in FF, i.e., from 37% (as shown in Fig. 7d) to 72% (as shown in Fig. 8d). Higher mobility in $CeO_{1.96}$ leads to improved conductivity in ETL which eventually results in enhanced FF of the device at lower doping density level compared to the previous section. Moreover, at ETL doping density of 1×10^{18} cm^{-3}, the conversion efficiency (16.8%) of the device is in good agreement with already published experimental results [18,45], same is reported in Table 2.

4.4. Optimised device with stacked CNT between perovskite and spiro-MeoTAD for moisture stability

All the previous analysis presented in this work deal with the optimization of the device, whereas, current section is devoted to the design of CNT stacked HTL based perovskite solar cell. Literature shows, the presence of CNT between perovskite and Spiro-MeOTAD provides moisture stability improvements, attributed to hydrophobic nature of CNTs [48,49]. CNTs have exceptional charge carrier properties, chemically and mechanically stable along with hydrophobic nature, longer carrier lifetime, low resistivity and high flexibility. CNTs are simulated using thin, i.e., 30 nm thick CNT layer whose parameters are reported in Table 1. 30 nm thick CNT layer is stacked between perovskite and HTL layer of the previously discussed device as reported in Fig. 1. Device design presented in this section is projected for both hysteresis and moisture stability due to the presence of PCBM between ETL and perovskite and CNT between perovskite and HTL. The J-V curve of the device is reported in Fig. 9, the result shows the presence of CNT does not provide a significant improvement in the conversion efficiency though could provide moisture stability and hence will have environment sustainability and robustness. It gives the scope for future development of hysteresis and moisture free device. Concluding, this subsection proposes the 18.6% efficient device design which will be free from stability issues. Additionally, overall results are also summarized in Table 3.

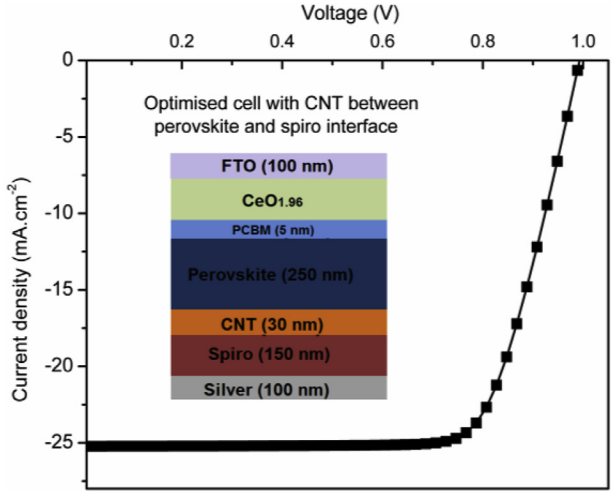


Fig. 9. J–V curve of CNT based optimized PSC. *INSET* shows the modified device structure.

4. Conclusions

TCAD analysis of the CeO_x based ETL and the PCBM interface layer has been carried out in this simulation-based research work. Impact of oxygen composition of CeO_x is obtained, which shows a significant effect on the photovoltaic performance of the device due to change in mobility, showing the optimum performance (16.5% efficiency) at x = 1.96. To understand the carrier dynamics inside the device, recombination rate, energy band diagrams, and potential profile are obtained for $CeO_{1.96}$ and $CeO_{1.88}$ based devices. Moreover, PCBM thickness and its doping density are also optimized. The result shows a thin PCBM layer is desired for undoped or lightly doped PCBM based device, and thicker PCBM can also be employed by increasing the doping. Moreover, doping variation in $CeO_{1.96}$ has also been carried out to optimize the conductivity of the device. Result concludes 18.2% efficient optimized perovskite solar cell with $CeO_{1.96}$ based ETL. In addition, to further improve the moisture stability of the device a 30 nm thick CNT layer was added without showing any degradation in the

Table 3
Photovoltaic performance of simulated devices along with comparison from experimental work [18].

PSCs	V_{OC} (V)	J_{SC} (mA.cm^{-2})	FF (%)	PCE (%)
Ref. PSC [18] (CeO_x + PCBM) based ETL	1.057 ± 0.013	22.90 ± 0.50	69.62 ± 1.81	16.85 ± 0.33
Simulated PSC (CeO_x + PCBM) based ETL with CeO_x doping 1×10^{18} cm^{-3}	0.993	25.04	67.68	16.82
Optimized PSC (with 5 nm thick PCBM and CeO_x doping of 1×10^{19} cm^{-3})	0.993	25.06	73.18	18.21
Optimized PSC with stacked CNT between perovskite/HTL	0.995	25.22	74.13	18.60

performance of the device. With all the optimization and addition of CNT, this reported work achieved PCE of 18.60% from a device which is free from hysteresis and moisture degradation. The analysis reported in this work evidently shows that the proposed device may open new doors for new generation stable PSC solar cells.

Acknowledgment

The authors gratefully acknowledge the support from Microelectronics Research Laboratory, Delhi Technological University to carry out the work. The author, Dr. Rahul Pandey thanks to Dr. Rajnish Sharma from Chitkara University for his continuous support during the preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vacuum.2018.10.033.

References

[1] I.S.E.V.p, ITRPV, 2016, ITRPV, 2016.
[2] P. Loper, S.-J. Moon, S. Martin de Nicolas, B. Niesen, M. Ledinsky, S. Nicolay, J. Bailat, J.-H. Yum, S. De Wolf, C. Ballif, Organic-inorganic halide perovskite/crystalline silicon four-terminal tandem solar cells, Phys. Chem. Chem. Phys. 17 (3) (2015) 1619–1629.
[3] A. Richter, M. Hermle, S.W. Glunz, Reassessment of the limiting efficiency for crystalline silicon solar cells, IEEE J. Photovol. 3 (4) (2013) 1184–1191.
[4] K. Yoshikawa, H. Kawasaki, W. Yoshida, T. Irie, K. Konishi, K. Nakano, T. Uto, D. Adachi, M. Kanematsu, H. Uzu, K. Yamamoto, Silicon Heterojunction Solar Cell with Interdigitated Back Contacts for a Photoconversion Efficiency over 26% vol. 2, (2017), p. 17032.
[5] M.M. Lee, J. Teuscher, T. Miyasaka, T.N. Murakami, H.J. Snaith, Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites, Science 338 (6107) (2012) 643–647.
[6] W.E.I. Sha, X. Ren, L. Chen, W.C.H. Choy, The efficiency limit of $CH_3NH_3PbI_3$ perovskite solar cells, Appl. Phys. Lett. 106 (22) (2015) 221104.
[7] H. Zhou, Q. Chen, G. Li, S. Luo, T.-b. Song, H.-S. Duan, Z. Hong, J. You, Y. Liu, Y. Yang, Interface engineering of highly efficient perovskite solar cells, Science 345 (6196) (2014) 542–546.
[8] W.S. Yang, J.H. Noh, N.J. Jeon, Y.C. Kim, S. Ryu, J. Seo, S.I. Seok, High-performance photovoltaic perovskite layers fabricated through intramolecular exchange, Science 348 (6240) (2015) 1234–1237.
[9] W. Nie, H. Tsai, R. Asadpour, J.-C. Blancon, A.J. Neukirch, G. Gupta, J.J. Crochet, M. Chhowalla, S. Tretiak, M.A. Alam, H.-L. Wang, A.D. Mohite, High-efficiency solution-processed perovskite solar cells with millimeter-scale grains, Science 347 (6221) (2015) 522–525.
[10] G. Yang, H. Tao, P. Qin, W. Ke, G. Fang, Recent progress in electron transport layers for efficient perovskite solar cells, J. Mater. Chem. 4 (11) (2016) 3970–3990.
[11] H.-S. Kim, C.-R. Lee, J.-H. Im, K.-B. Lee, T. Moehl, A. Marchioro, S.-J. Moon, R. Humphry-Baker, J.-H. Yum, J.E. Moser, M. Grätzel, N.-G. Park, Lead iodide perovskite sensitized all-solid-state submicron thin film mesoscopic solar cell with efficiency exceeding 9%, Sci. Rep. 2 (2012) 591.
[12] G. Pellegrino, A.L. Magna, G.G. Condorelli, C. Bongiorno, C. Mocuta, E. Smecca, A. Alberti, A strategy to stabilise the local structure of Ti4+ and Zn2+ species against aging in TiO2/aluminium-doped ZnO bi-layers for applications in hybrid solar cells, J. Appl. Phys. 116 (5) (2014) 054907.
[13] K.-C. Wang, P.-S. Shen, M.-H. Li, S. Chen, M.-W. Lin, P. Chen, T.-F. Guo, Low-temperature sputtered nickel oxide compact thin film as effective electron blocking layer for mesoscopic NiO/CH3NH3PbI3 perovskite heterojunction solar cells, ACS Appl. Mater. Interfaces 6 (15) (2014) 11851–11858.
[14] A.S. Subbiah, A. Halder, S. Ghosh, N. Mahuli, G. Hodes, S.K. Sarkar, Inorganic hole conducting layers for perovskite-based solar cells, J. Phys. Chem. Lett. 5 (10) (2014) 1748–1753.
[15] K. Mahmood, S. Sarwar, M.T. Mehran, Current status of electron transport layers in perovskite solar cells: materials and properties, RSC Adv. 7 (28) (2017) 17044–17062.
[16] Z. Yu, F. Qi, P. Liu, S. You, K.K. Kondamareddy, C. Wang, N. Cheng, S. Bai, W. Liu, S. Guo, X.-z. Zhao, A composite nanostructured electron-transport layer for stable hole-conductor free perovskite solar cells: design and characterization, Nanoscale 8 (11) (2016) 5847–5851.
[17] J. Choi, S. Song, M.T. Hörantner, H.J. Snaith, T. Park, Well-defined nanostructured, single-crystalline TiO2 electron transport layer for efficient planar perovskite solar cells, ACS Nano 10 (6) (2016) 6029–6036.
[18] X. Wang, L.-L. Deng, L.-Y. Wang, S.-M. Dai, Z. Xing, X.-X. Zhan, X.-Z. Lu, S.-Y. Xie, R.-B. Huang, L.-S. Zheng, Cerium oxide standing out as an electron transport layer for efficient and stable perovskite solar cells processed at low temperature, J. Mater. Chem. 5 (4) (2017) 1706–1712.
[19] Ü. Özgür, Y.I. Alivov, C. Liu, A. Teke, M.A. Reshchikov, S. Doğan, V. Avrutin, S.J. Cho, H. Morkoç, A comprehensive review of ZnO materials and devices, J. Appl. Phys. 98 (4) (2005) 041301.
[20] S.-H. Wu, C.-H. Chan, C.-H. Chien, M.T. Yaseen, C.-T. Liang, Y.-C. Chang, Enhanced emission and photoconductivity due to photo-induced charge transfer from Au nanoislands to ZnO, Appl. Phys. Lett. 108 (4) (2016) 041104.
[21] J. Dong, Y. Zhao, J. Shi, H. Wei, J. Xiao, X. Xu, J. Luo, J. Xu, D. Li, Y. Luo, Q. Meng, Impressive enhancement in the cell performance of ZnO nanorod-based perovskite solar cells with Al-doped ZnO interfacial modification, Chem. Commun. 50 (87) (2014) 13381–13384.
[22] L. Liang, Z. Huang, L. Cai, W. Chen, B. Wang, K. Chen, H. Bai, Q. Tian, B. Fan, Magnetron sputtered Zinc oxide nanorods as thickness-insensitive cathode interlayer for perovskite planar-heterojunction solar cells, ACS Appl. Mater. Interfaces 6 (23) (2014) 20585–20589.
[23] S.-H. Wu, M.-Y. Lin, S.-H. Chang, W.-C. Tu, C.-W. Chu, Y.-C. Chang, A Design Based on Stair-case Band Alignment of Electron Transport Layer for Improving Performance and Stability in Planar Perovskite Solar Cells, (2017) arXiv preprint arXiv:1708.03153.
[24] X. Dong, H. Hu, B. Lin, J. Ding, N. Yuan, The effect of ALD-Zno layers on the formation of CH3NH3PbI3 with different perovskite precursors and sintering temperatures, Chem. Commun. 50 (92) (2014) 14405–14408.
[25] C. Xiao, C. Wang, W. Ke, B.P. Gorman, J. Ye, C.-S. Jiang, Y. Yan, M.M. Al-Jassim, Junction quality of SnO2-based perovskite solar cells investigated by nanometer-scale electrical potential profiling, ACS Appl. Mater. Interfaces 9 (44) (2017) 38373–38380.
[26] K. Eguchi, T. Setoguchi, T. Inoue, H. Arai, Electrical properties of ceria-based oxides and their application to solid oxide fuel cells, Solid State Ionics 52 (1) (1992) 165–172.
[27] S. Vangelista, R. Piagge, S. Ek, T. Sarnet, G. Ghidini, C. Martella, A. Lamperti, Structural, chemical and optical properties of cerium dioxide film prepared by atomic layer deposition on TiN and Si substrates, Thin Solid Films 636 (Supplement C) (2017) 78–84.
[28] H.L. Tuller, A.S. Nowick, Small polaron electron transport in reduced CeO2 single crystals, J. Phys. Chem. Solid. 38 (8) (1977) 859–867.
[29] M. Mogensen, N.M. Sammes, G.A. Tompsett, Physical, chemical and electrochemical properties of pure and doped ceria, Solid State Ionics 129 (1) (2000) 63–94.
[30] R. Pandey, R. Chaujar, Numerical simulations: toward the design of 27.6% efficient four-terminal semi-transparent perovskite/SiC passivated rear contact silicon tandem solar cell, Superlattice. Microst. 100 (2016) 656–666.
[31] C. Fu-Chien, L. Chih-Ming, Optical and electrical characterizations of cerium oxide thin films, J. Phys. Appl. Phys. 43 (7) (2010) 075104.
[32] C. Liu, S. Xiao, X. Shu, Y. Li, L. Xu, T. Liu, Y. Yu, L. Zhang, H. Liu, Y. Li, Synthesis and photovoltaic properties of novel monoadducts and bisadducts based on amide methanofullerene, ACS Appl. Mater. Interfaces 4 (2) (2012) 1065–1071.
[33] L.J. Phillips, A.M. Rashed, R.E. Treharne, J. Kay, P. Yates, I.Z. Mitrovic, A. Weerakkody, S. Hall, K. Durose, Dispersion Relation Data for Methylammonium Lead Triiodide Perovskite Deposited on a (100) Silicon Wafer Using a Two-step Vapour-Phase Reaction Process 5 (2015), pp. 926–928 Data in Brief.
[34] M. Filipič, P. Löper, B. Niesen, S. De Wolf, J. Krč, C. Ballif, M. Topič, CH(3)NH(3)PbI(3) perovskite/silicon tandem solar cells: characterization based optical simulations, Optic Express 23 (7) (2015) A263–A278.
[35] P.S. University, PCBM Band Structure Parameters Set, Analysis of Microelectronic and Photonic Structures, Material Data, 2018, http://www.ampsmodeling.org/materials/PCBM_densityofstates.htm.
[36] F. Liu, J. Zhu, J. Wei, Y. Li, M. Lv, S. Yang, B. Zhang, J. Yao, S. Dai, Numerical simulation: toward the design of high-efficiency planar perovskite solar cells, Appl. Phys. Lett. 104 (2014) (2014) 253508-253508.
[37] G. Xosrovashvili, N.E. Gorji, Numerical simulation of carbon nanotubes/GaAs hybrid PV devices with AMPS-1D, Int. J. Photoenergy 2014 (2014) 6.
[38] M. Samiee, S. Konduri, B. Ganapathy, R. Kottokkaran, H.A. Abbas, A. Kitahara, P. Joshi, L. Zhang, M. Noack, V. Dalal, Defect density and dielectric constant in perovskite solar cells, Appl. Phys. Lett. 105 (15) (2014) 153502.

[39] C. Wehrenfennig, G.E. Eperon, M.B. Johnston, H.J. Snaith, L.M. Herz, High charge Carrier mobilities and lifetimes in organolead trihalide perovskites, Adv. Mater. 26 (10) (2014) 1584–1589.
[40] I.K. Naik, T.Y. Tien, Small-polaron mobility in nonstoichiometric cerium dioxide, J. Phys. Chem. Solid. 39 (3) (1978) 311–315.
[41] S.M. Sze, Semiconductor Devices: Physics and Technology, John Wiley & Sons, 2008.
[42] F. Fu, T. Feurer, T. Jäger, E. Avancini, B. Bissig, S. Yoon, S. Buecheler, A.N. Tiwari, Low-temperature-processed efficient semi-transparent planar perovskite solar cells for bifacial and tandem applications, Nat. Commun. 6 (November) (2015) 8932-8932.
[43] W.-J. Huang, P.-H. Huang, S.-H. Yang, PCBM doped with fluorene-based polyelectrolytes as electron transporting layers for improving the performance of planar heterojunction perovskite solar cells, Chem. Commun. 52 (93) (2016) 13572–13575.
[44] C. Kuang, G. Tang, T. Jiu, H. Yang, H. Liu, B. Li, W. Luo, X. Li, W. Zhang, F. Lu, J. Fang, Y. Li, Highly efficient electron transport obtained by doping PCBM with Graphdiyne in planar-heterojunction perovskite solar cells, Nano Lett. 15 (4) (2015) 2756–2762.
[45] R. Fang, S. Wu, W. Chen, Z. Liu, S. Zhang, R. Chen, Y. Yue, L. Deng, Y.-B. Cheng, L. Han, W. Chen, [6,6]-Phenyl-C61-Butyric acid methyl ester/cerium oxide bilayer structure as efficient and stable electron transport layer for inverted perovskite solar cells, ACS Nano 12 (3) (2018) 2403–2414.
[46] A. Corma, P. Atienzar, H. García, J.-Y. Chane-Ching, Hierarchically mesostructured doped CeO2 with potential for solar-cell use, Nat. Mater. 3 (2004) 394.
[47] A. Tschöpe, E. Sommer, R. Birringer, Grain size-dependent electrical conductivity of polycrystalline cerium oxide: I. Experiments, Solid State Ionics 139 (3) (2001) 255–265.
[48] S.N. Habisreutinger, T. Leijtens, G.E. Eperon, S.D. Stranks, R.J. Nicholas, H.J. Snaith, Carbon nanotube/polymer composites as a highly stable hole collection layer in perovskite solar cells, Nano Lett. 14 (10) (2014) 5561–5568.
[49] K. Aitola, K. Domanski, J.P. Correa-Baena, K. Sveinbjörnsson, M. Saliba, A. Abate, M. Grätzel, E. Kauppinen, E.M. Johansson, W. Tress, High temperature-stable perovskite solar cell based on low-cost carbon nanotube hole contact, Adv. Mater. 29 (17) (2017).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor : 6.887
Volume 6 Issue II, February 2018- Available at www.ijraset.com

Performance Analysis of 4 Stroke Compression Ignition Engine by Using Diesel and Biodiesel Blend of Hingot Oil

Mr. Akash Singh Raikwar[1], Mr. Amit Agarwal[2], Dr. Amit Pal[3]
[1]*Department of Mechanical Engineering, SRCEM Bamore, Morena)*
[2]*Department of Mechanical Engineering, SRCEM Bamore, Morena)*
3Department of Mechanical Engineering, Delhi Technical University, Delhi India

Abstract : Looking to present scenario of rapidly increases the pollution in our country and worldwide too the need of some clean and non toxic fuels which helps the environment from the pollution, also the rise of petroleum is hike day by day and we all know petroleum is limited we can't have petroleum after next 50 year if the same rate of consumption is continue so for a future we needs some alternatives of petroleum, there are lot of alternative fuels is discovered in present days out of which biodiesel is most convenient fuel from our society. Biodiesel is clean, non toxic, biodegradable and environment friendly alternative fuels Since biodiesel is produce from the vegetable oil or animal fats hence its generally renewable fuels.

In this paper I have tested the performance of compressed ignition engine in varying load conditionby using biodiesel which is produced in a laboratory of Delhi Technological University, Delhi from a seed of Hingot (Balanitesaegyptiaca Del.) by Mechanical stirring method, testing is done with the diesel blend which have different properties at B0, B10, B20 and B30. Among the all blends, the the specific fuel consumption, brake specific fuel consumption and brake thermal efficiency is calculate and compare with the all samples of blends.

Keyword: Biodiesel, Performance test, Brake power, Blend, Brake thermal efficiency

I. TRODUCTION

Compressed Ignition Engine is a type of IC Engine which is generally used for a run heavy engine and generator and uses Diesel a fuel in this if we use the blend of biodiesel in place of pure biodiesel without any modification in engine then it run smoothly up to a certain percentage of biodiesel in diesel called blend. There aremany of past experiments says that CI engine run smoothly with biodiesel blends. Gopal el al.[1] performed experiments on a constant speed single cylinder, 4 stroke ,air cooled diesel engine for agriculture purpose used to test the fuel. Raw material used for biodiesel production was waste cooking oil which was collected from difference sources, specific energy consumption changes with increase in blends due to the lower value of biodiesel. Shahidet al.[2]observed B100 is also used an engine but B20 are increases the fuel consumption and 10% decreases the brake thermal efficiency. Zakira el al.[3] also used Waste cooking oil for producing biodiesel and perform the test at B20, B40, B60, B80 and B100 and observed the lowest BHP. Higher BHP obtained when engine was fueled by blends with low biodiesel contained also it was found that average brake specific fuel consumption of B100 is 9.45% higher than diesel fuel. Sharmaet al [4] uses WCO for produce biodiesel, B20 and B40 blend result the brake thermal efficiency is higher with respect to diesel at higher load, this is due to better combustion of biodiesel blends due to present of extra amount of oxygen, in case of brake specific fuel values are higher at the beginning due to higher viscosity. Srinivas et al. [5] observed at B20 blend brake thermal efficiency is 37.855% and it is 0.275% less than of standard diesel operation. Rao et al.[5] used cooking methyl ester UCME which has low calorif value than that of biodiesel hence specific fuel consumption is slightly higher than that of biodiesel for UCME and its blends. Brake thermal efficiency of UCME and its blend is lower than pure diesel. Algasim et al.[6] used turbocharged diesel engine for testing, standard diesel produced the higher brake torque and power at the engine for all engine speeds. This is due to lower calorific value of the biodiesel compare to standard value, at full load operation, the lowest specific fuel consumption was for the standard diesel compare to biodiesel.

II. PERFORMANCE TESTING

A. Engine test setup

The setup of engine is consist of single cylinder, four stroke, kirloskar diesel engine connected to an eddy current dynamometer for loading. The setup has stand-alone panel box consisting of air box, two fuel tanks for duel fuel testing, fuel measuring unit,

©IJRASET (UGC Approved Journal): All Rights are Reserved

manometer, transmitters for air and fuel flow measurements, process indicator and engine indicator. Rotameters are provided for cooling water and calorimeter water flow measurement. The setup enables study of engine of engine performance various parameters like brake power, indicated power, frictional power, break mean effective pressure (BMEP), indicated mean effective pressure (IMEP), indicating thermal efficiency, mechanical efficiency, volumetric efficiency, specific fuel consumption, air/fuel ratio and heat balance. The aim of these experiments is to study the effects of various parameters on performance of kirloskar engine for different blends of diesel and biodiesel.

B. Engine Specifications

Table:1Constructional Specifications of engine test setup.

S.No.	Component	Specifications
1.	Product	Kirloskar engine test setup 1 cylinder, 4stroke diesel engine.
2.	Product code	234
3.	Engine	Make kirloskar, type 1 cylinder, 4 stroke diesel, water cooled, power 3.5kW at 1500rpm, stroke 110mm, bore 87.5mm. 661cc CR17.5
4.	Dynamometer	Type eddy current, water cooled with loading unit
5.	Propeller shaft	With universal joint
6.	Air box	M S fabricated with orifice meter and manometer
7.	Fuel tank	Capacity 15L with glass fuel metering column
8.	Calorimeter	Type pipe in pipe
9.	piezo powering unit	Make cuadra, Model AX-409
10.	Digital milivoltmeter	Range 0-500mV, panel mounted
11.	Load indicator	Digital, range 0-50Kg, Supply 230V AC
12.	Load sensor	Load cell, type strain gauge, range 0-50Kg
13.	Rotameter	Engine cooling 40-400LPH: calorimeter 25-250LPH

Table: 2 Technical Specifications of engine test setup.

S.No.	Component	Specifications
1.	Number of cylinder	1
2.	Number of stroke	4
3.	Fuel	HS diesel
4.	Rated power	3.5KW @ 1500rpm
5.	Cylinder diameter	87.5mm
6.	Stroke length	110mm
7.	Connecting rod length	234mm
8.	Compression ratio varies	12 to 18:1
9.	Orifice diameter	20mm
10.	Dynamometer arm length	185mm

C. Preparation of Blends

Three different blends of biodiesel were prepared as explained below:

1) B10 (10% biodiesel + 90% diesel): To make a sample 500ml of B10, 450ml pure diesel was mixed in 50ml of biodiesel prepared from vegetable oil in a beaker.

2) B20 (20% biodiesel + 80% diesel): To make a sample 500ml of B20, 400ml pure diesel was mixed in 100ml of biodiesel prepared from vegetable oil in a beaker.

3) B30 (30% biodiesel + 70% diesel): To make a sample 500ml of B30, 350ml pure diesel was mixed in 150ml of biodiesel prepared from vegetable oil in a beaker.

©IJRASET (UGC Approved Journal): All Rights are Reserved

Density of pure diesel=811Kg/m^3
Density of biodiesel=859.7Kg/m^3

Table: 3 Description of diesel, biodiesel and different blends of biodiesel

Type of blend	Amount of biodiesel (ml)	Amount of pure diesel (ml)	Resultant density (Kg/m^3)
Diesel	0	500	811
Biodiesel	500	0	859.7
B10	50	450	816.7
B20	100	400	820.3
B30	150	350	826.9

Table:4Properties of diesel, biodiesel and different blends of biodiesel

Type of blend	Viscosity of biodiesel (mPa.s)	Kinematics viscosity of pure diesel (mm^2/s)	Resultant density (Kg/m^3)
Diesel B0	2.4811	3.0593	811
Biodiesel B100	4.6340	5.3905	859.7
B10	2.5500	3.1222	816.7
B20	2.6487	3.2288	820.3
B30	2.7689	3.3066	826.9

D. Performance Evaluations

All the experiments were performed on a single cylinder, 4 stroke kirloskar diesel engine. There were two separate fuel inlet arrangement is available in a setup, one for the diesel and another for biodiesel blend. Fuel was supplied to the engine from an outside tank. A 5-10min warm up period is provided from first run data collection. Initial engine is started with the diesel and then after 5-10min. period blend is use. The gap of 4-5min. was provided between the two consecutive runs for accurate data obtaining. Engine speed, specific fuel consumptions, brake power, heat balance etc. were measure during each run. During load test take rpm constant (1500rpm) and load on engine varied from 0 to 12kg by adjusting the load knob provided on the control panel of the test ring. The tests were performed with pure diesel (B0)and biodiesel blends (B10, B20, B30)

E. Pure diesel: B0, Biodiesel Blend: B10, Biodiesel Blend: B20 and Biodiesel Blend: B30

Engine performance parameters obtained from performance testing in single cylinder four stroke CI Engine against different loads for pure diesel ie B0 and Biodiesel Blend B10, B20 and B30

III. RESULTS AND DISCUSSION

A. Brake thermal efficiency vs Brake Power for Diesel and Biodiesel Blends

From the fig. it is observed that Brake thermal efficiency BTHE for both pure diesel and biodiesel blends increases with increase in brake power. Also the efficiency of biodiesel blends is more than pure diesel i.e. B0 and maximum is observed in the blend of B10. The reason is that is presence of more oxygen in biodiesel blends compare to pure diesel. With further increases the percentage of biodiesel in the blend a slight decreases in BTHE is seen in a fig. The reason for the decreases in BTHE in high % blend B20, B30 may be due to less volatility of biodiesel as compare to pure diesel B0.The maximum brake thermal efficiency is achieved 33.67 at a brake power of 2.79kW for B10. The value of BTHE for pure diesel B0 at a same load is obtained 32.46%. A condition of minimum load of BTHE is minimum 11.73% for B0 and 13.01% for B10. The variation of BTHE vs BP for different blends of diesel and pure diesel is shown in fig.1

©IJRASET (UGC Approved Journal): All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor : 6.887
Volume 6 Issue II, February 2018- Available at www.ijraset.com

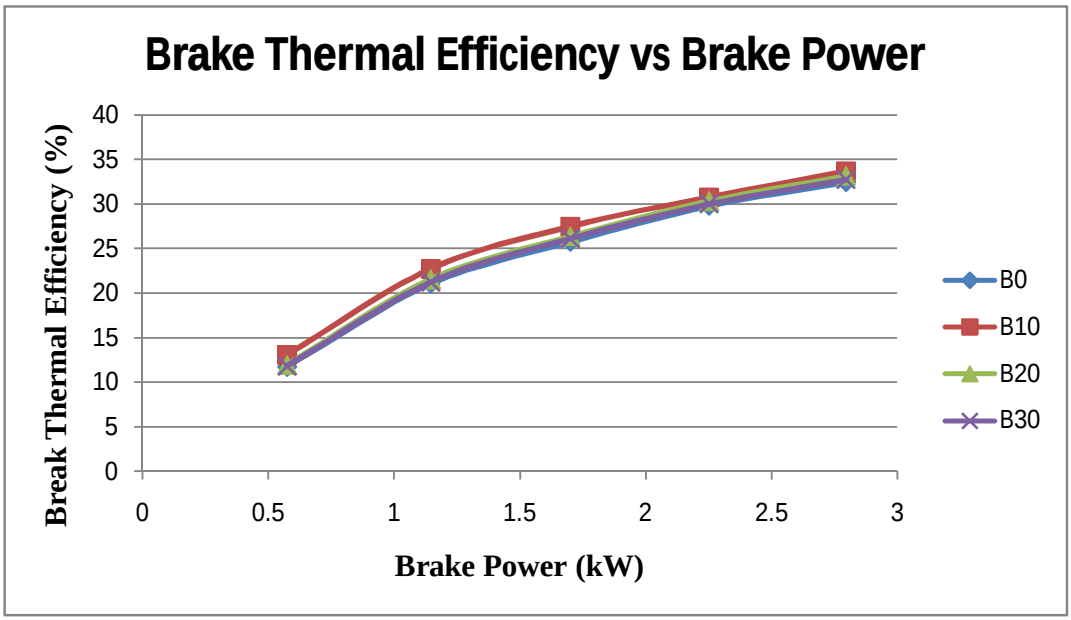


Fig.1 Brake thermal efficiency vs Brake Power for diesel and biodiesel blends

B. Brake Specific Fuel Consumption Vs Brake Power For Diesel And Biodiesel Blends

From the fig. it is observed that Brake specific fuel consumption BSFC for both pure diesel and biodiesel blends decreases with increase in brake power. As we know the total fuel consumption increases with increases the percentage of biodiesel increases in blends of biodiesel the reason is that biodiesel has low calorific value as compare to pure diesel so more fuel is induced in blends of biodiesel. Thus BSFC increases with increases in biodiesel percentage in blend.

It is Also observed from the fig that Brake specific fuel consumption BSFC for blends B10, B20, B30 at the maximum brake power of 2.79kW is 0.257, 0.261, 0.264 (in kg/kW-h) respectively. And for the minimum load (brake power 0.574kW) BSFC is observed for B10, B20, B30 is 0.661, 0.722, 0.731 (in kg/kW-h) respectively. The variation of BSFC vs BP for different blends of diesel and pure diesel is shown in fig.2

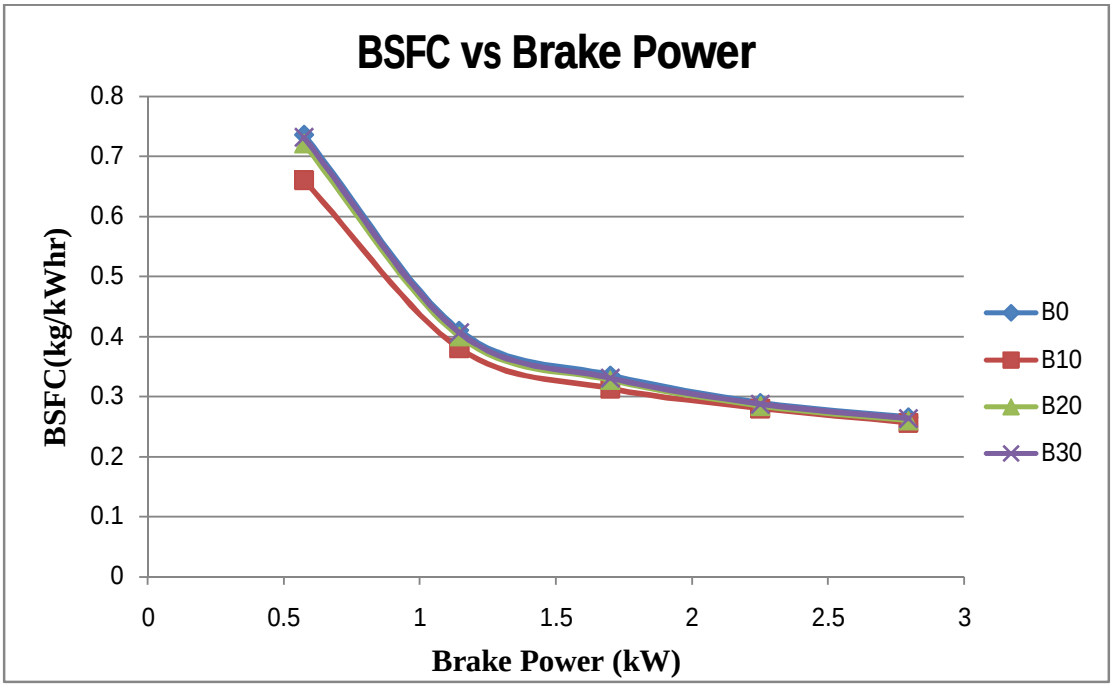


Fig.2 Brake specific fuel consumption vs Brake Power for diesel and biodiesel blends

©IJRASET (UGC Approved Journal): All Rights are Reserved

C. Brake Specific Energy Consumption Vs Brake Power For Diesel And Biodiesel Blends

The concept of brake specific energy consumption BSEC is same as brake specific fuel consumption BSFC because in mathematical expression of BSEC is the product of BSFC and calorific value of fuel. So the variation of BSFC and BSEC is same. BSEC is also increases with increases of percentage of biodiesel in blend.It is Also observed from the fig that Brake specific energy consumption BSEC for blends B10, B20, B30 at the maximum brake power of 2.79kW is 10.687, 10.855, 10.991 (in MJ/kW-h) respectively. And for the minimum load (brake power 0.574kW) BSFC is observed for B10, B20, B30 is 27.519, 30.078, 30.508 (in MJ/kW-h) respectively. The variation of BSFC vs BP for different blends of diesel and pure diesel is shown in fig.18

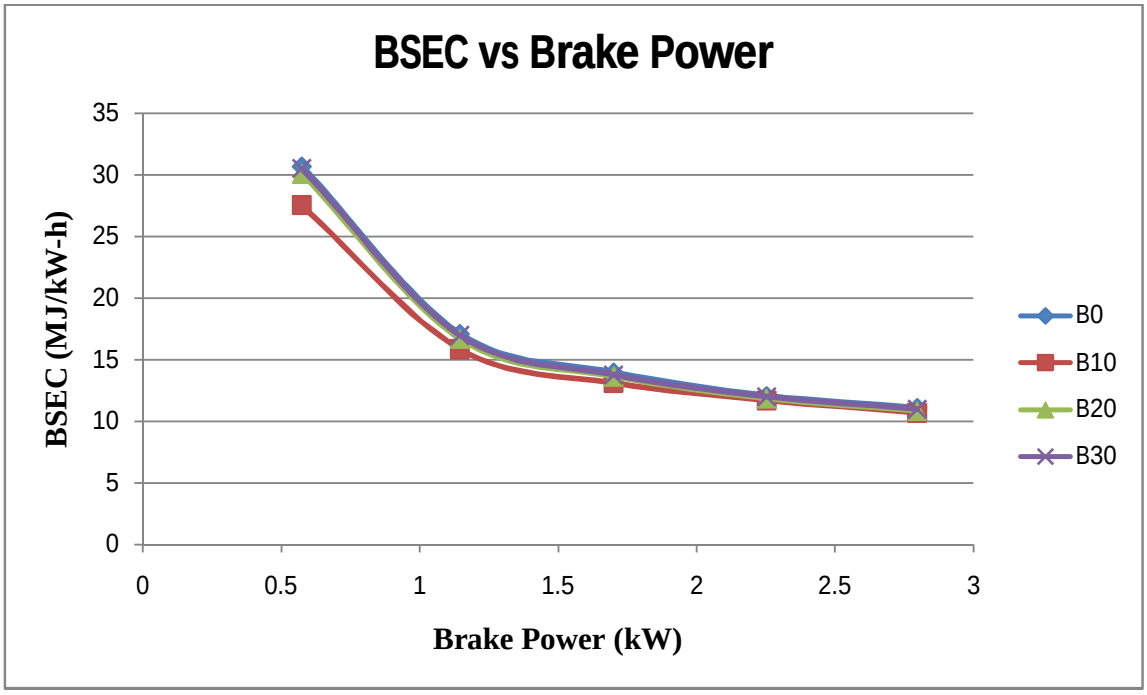


Fig.3 Brake specific energy consumption vs Brake Power for diesel and biodiesel blends

IV. CONCLUSION

A. Following Conclusion Has Been Made From The Experiments

1) For the engine performance testing it can be concluded that the performance parameter for blend of Hingot Biodiesel have gave better results than that of pure diesel without modification in engine setup.

2) The BTHE increases with increases brake power from low to high. This variation is shows because of better combustion biodiesel blends compare to pure diesel.

3) The BSFC decreases with increases in brake power. The BSFC is increases with increases in percentage of biodiesel in blend.

4) The BSEC also decreases with increases in brake power. The BSEC is increases with increases in percentage of biodiesel in blend just like BSFC.

5) It is also observed that BSEC and BSFC curves obtained by B20, B30 are very similar to pure diesel performed on the same engine at same load. Finally it can be concluded that the Hingot is also a one of the seed identified from which a biodiesel is also produced. Blend of Hingot Biodiesel is an industrial viable substitute for diesel engine.

REFRENCES

[1] K. NanthaGopal, Arindam Pal, Sumit Sharma, CharanSamanchi, K. sarthyananarayanan, T.Elango "Investigation of emissions and combustion characterstics of a CI Engine fueled with waste cooking oil methyl ester and diesel blends" , Alexanderia Engineering journal(2014) 53,pp281-287

[2] E.M Shahid, Y.Jamal, A.N. Shah, N. Rumzan, M. munsha. "Effect of used cooking oil methyl ester on Combustion Ignition Engine", Journal of Quantity and Technology Management, Volume VIII, Issue II, December 2012, pp.91-104

[3] Sheila M.Zakiya, Susun A. Roces, Nathaniel P.Dugos, Michael Angelo B. promentilla and Masahiro Shioj. "Emissions and Performance of Diesel Engine Fueled with Waste Cooking oil Methyl Ester Diesel Blends" , Iranica journal of Energy & Environment 5(1), (2014) pp.73-77

[4] R.B. Sharma, Dr. Amit Pal and JuhiSharaf. "Experimental Investigation of Bio-diesel Obtained from waste cooking oil and its Blends with diesel on Single Cylinder Engine", International Journal of Engineering Research and applications, ISSN: 2248-9622, Vol.4, Issue 1(Version 1), January 2014, pp.193-200.

©IJRASET (UGC Approved Journal): All Rights are Reserved

[5] K. Srinivas, N.Ramakrishna, B.Balunaik and K.Kalyaniradha. "Performance and emission analysis of waste vegetable oil and its blends with diesel and additives", Journal of Engineering Research and Application, ISSN: 2248-9622, vol.3, Issue 6, Nov-Dec 2013,pp.473-478

[6] Narayan Rao, G Lakshmi, S Sampath and K Rajagopal. " Experimental Studies on the combustipon and emission characterstics of a diesel engine fuelled with using cooking oil methyl ester and its diesel blends" World Academy of Science, Engineering and Technology Vol:2 2008, pp.01-25

[7] Hamad. M. Algasim and John Field house. "Performance and emissions of Compression Ignition Engines Using Waste Cooked Oil as fuel" Computing and Engineering Researcher's Conference, University of Huddersfield, Dec 2010, pp.65-70

[8] Amit Pal, Shashank Mohan and DhananjayTrivedi. " Production and Performance Testing of Waste Frying oil Biodiesel", International Journal of current Engineering and Technology, E-ISSN 2277-4104, Vol.4 No.3 (June-2014) pp. 347-516.

[9] R.B. Sharma, Dr. Amit Pal and JuhiSharaf. "Experiment Investigation of Bio-Diesel Obtained From Waste Cooking Oil and Its Blends with diesel on Single Cylinder Engine", International Journal of Engineering Research and Application, ISSN: 2248-9622, Vol.4, Issue 1(Version 1), January 2014, pp.193-200.

[10] NeeleshSoni, Om PrakashChaurasia "Experimental Investigation of 4 Stroke Compression Ignition Engine by using diesel and processed waste cooking blend" International Journal of Current Engineering and Scientific Research, ISSN: 2394-0697, Volume-3 Issue-1, 2016

©IJRASET (UGC Approved Journal): All Rights are Reserved

2018 3rd International Conference on Intelligent Transportation Engineering

Predictive Current Control of Feedback Linearized Induction Motor Drive

Ambrish Devanshu, Madhusudan Singh, Narendra Kumar
Electrical Engineering Department, Delhi Technological University, Delhi, India
e-mail: ambrish.devanshu@gmail.com, madhusudan@dce.ac.in, ndeshwalus@gmail.com

Abstract— **In this paper, predictive current controller (PCC) is implemented for controlling voltage source inverter (VSI) for a feedback linearization controlled (FLC) induction motor (IM) drive for better rotor speed and torque dynamics. Two new control inputs are developed using output of Proportional Integral (PI) controllers which regulate the rotor flux and rotor speed. Reference currents are the input to the PCC. The PCC regulates stator current of the IM according to the generated reference signal received. Error between predicted stator currents and reference current generated using control algorithm is minimized using the cost function defined in PCC. The best switching state which gives the minimum value of cost function is selected and applied to the switches of the three-leg VSI without any use of modulation stage and linear regulators. The proposed method is validated using MATLAB/Simulink for different operating points.**

Keywords-Predictive current control; Induction motor; Feedback Linearization, Voltage source inverter

I. INTRODUCTION

Induction motors are widely used in many industrial and commercial applications such as machine tools, pumps, fans, transportation, agricultural machinery, air compressors, etc. due to their low cost, long working life, low noise, reliable operation and easy maintenance. The vector control of induction motor (IM) evaluates the DC motor like behaviour in IM. The vector control of IM provides a very fast and accurate speed control of IM. The high performance IM drive with advancement of control techniques are a) when load is applied drop in speed and reinstate time should be minimum b) In Speed/torque, tracking and regulation error should be minimum c) dynamic performance of the drive should not be sensitive to variation in parameters d) Fast tracking of reference Speed/torque without overshoot. The theory of field oriented control (FOC) was established by Blaschke which partly linearizes the system in rotor rotating reference frame and this theory was applied to IM to achieve DC machine like control of speed and flux [1]-[10]. The main drawback of the FOC of IM is the existence of non-linearity in the mechanical equation during the variation of rotor flux linkage. However, positive results from FOC have been seen. Efforts to achieve new and more accurate control methods are made during last few decades by many researchers in the area of electric drives. One such development is the FLC or input-output (I/O) decoupling control. Non-linear control of IM was first presented in which author introduced a differential geometric approach to implement it for induction motor (IM) [11]-[15].

Inverter fed IM is widely used in industrial applications with high performance current controllers. For the advancement of such controllers, numerous investigations have been performed in literature [16]. Linear methods like PI controllers using PWM and nonlinear approach like hysteresis control are recognised in literature for FLC of IM drive. Simple implementation and fast transient response of hysteresis current control (HCC) makes it popular but with the change in operating conditions and variations in load parameters, it is seen that PWM frequency varies within a band which is a major disadvantage of HCC[17]-[19]. Because of the variable switching frequency and resonance ripple harmonics are generated. This causes phase lag in the fundamental current which increases at higher frequencies [18], [19]. This deviation in phases leads to deterioration in performance of IM drives. Switching frequency of the current control PWM is set constant by the carrier and the performance of the IM control method rest on the frequency of reference current and PI controller parameters. Tuning of the PI controllers are not an easy task especially when there are state and input constraints [20]. In the case of sinusoidal reference, PI controller can't assures the zero steady state error.

In this paper, PCC scheme is proposed for feedback linearized induction motor drive as shown in Fig.1. Fast dynamic performance of FLC IM drive system over a wide range of inverter operating frequencies is achieved using proposed PCC method, overcoming the limitations of the hysteresis current control and SVPWM current control techniques. Stator current is measured from the stator terminal of IM and the reference current is calculated from FLC technique. Error between predicted stator currents and reference current generated using control algorithm is minimized for each voltage vector for a cycle using the cost function defined in PCC. The best switching state which gives the minimum value of cost function is selected and applied to switches of the three-leg VSI. The operation of the FLC IM drive with PCC under different conditions and improvement in the drive dynamic performance is demonstrated.

II. MATHEMATICAL MODELLING OF IM

Non-linear coupled dynamic model of induction motor in αβ-stationary reference frame can be written as follow from (1) to (6)

$$\frac{di_{\alpha s}}{dt} = -\frac{1}{\sigma}[r_s + \frac{r_r l_m^2}{l_r^2}]i_{\alpha s} + \frac{l_m r_r}{\sigma l_r^2}\varphi_{\alpha r} + \frac{l_m}{\sigma l_r} p\omega_r \varphi_{\beta r} + \frac{v_{\alpha s}}{\sigma} \quad (1)$$

$$\frac{di_{\beta s}}{dt} = -\frac{1}{\sigma}[r_s + \frac{r_r l_m^2}{l_r^2}]i_{\beta s} + \frac{l_m r_r}{\sigma l_r^2}\varphi_{\beta r} - \frac{l_m}{\sigma l_r} p\omega_r \varphi_{\alpha r} + \frac{v_{\beta s}}{\sigma} \quad (2)$$

$$\frac{d\varphi_{\alpha r}}{dt} = -\frac{r_r}{l_r}\varphi_{\alpha r} - p\omega_r \varphi_{\beta r} + \frac{l_m r_r}{l_r} i_{\alpha s} \quad (3)$$

$$\frac{d\varphi_{\beta r}}{dt} = -\frac{r_r}{l_r}\varphi_{\beta r} + p\omega_r \varphi_{\alpha r} + \frac{l_m r_r}{l_r} i_{\beta s} \quad (4)$$

$$T_e = \frac{3}{2} P \frac{l_m}{l_r}(\varphi_{\alpha r} i_{\beta s} - \varphi_{\beta r} i_{\alpha s}) = K_T(\varphi_{\alpha r} i_{\beta s} - \varphi_{\beta r} i_{\alpha s}) \quad (5)$$

$$\frac{d\omega_r}{dt} = -\frac{B}{J}\omega_r + \frac{1}{J}(T_e - T_L) \quad (6)$$

Where r_s and r_r are per phase stator and rotor resistances l_s and l_r are the stator and rotor inductances per phase respectively and p is the differential operator l_m is the magentizing inductance of the motor. B is the coefficient of viscous friction, J is the rotor inertia, T_e is the torque developed and T_L is the load torque. ω_r is the rotor speed. Here $K_T = \frac{3}{2} P \frac{l_m}{l_r}$ and P is the number of the pole, Stator fluxes, voltages and currents are $(\varphi_{\alpha s}, \varphi_{\beta s})$ $(v_{\alpha s}, v_{\beta s})$, $(i_{\alpha s}, i_{\beta s})$, and the rotor fluxes in $\alpha\beta$ reference frame are $(\varphi_{\alpha r}, \varphi_{\beta r})$.

III. FEEDBACK LINEARIZATION CONTROL

The state equations of a nonlinear system is described as

$$\dot{x} = f(x) + g(x)u \quad (7)$$

$$y = h(x) \quad (8)$$

Where f, g and h are adequately smooth in a domain $D \subset R^n$, 'u' is the scalar control input, 'y'is the scalar control output. The mapping $f : D \to R^n$ and $g : D \to R^n$ are called vector fields on 'D' [1]-[10]. In I/O linearization, output 'y' of the nonlinear dynamic system is differentiated till the input 'u' come into view in the r[th] derivative of the output 'y'. Assume input 'u' developed in the r[th] derivative of 'y' which takes the form

$$y^r = \eta(x) + \xi(x)u \quad (9)$$

If $\xi(x)$ is zero, output 'y' is differentiated until input 'u' appears in the derivative of the output 'y' for $r \le n$. Here 'n' is the order of the system and $'r'$ is the relative degree and the smallest number where our control law is applied. When $\xi(x)$ is non-singular, then, in this case we can define the control law

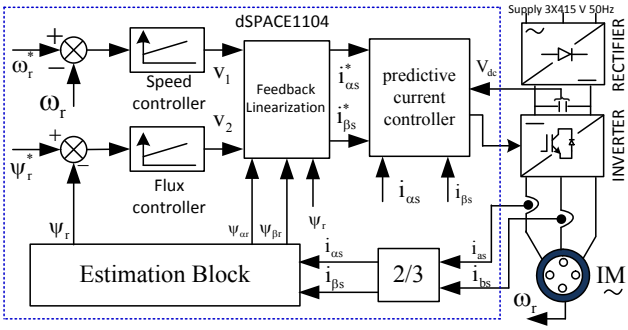


Figure 1. Block diagram of proposed method

$$u = \frac{1}{\xi(x)}[-\eta(x) + v] \quad (10)$$

Where $\xi(x) \neq 0$, v is the synthetic input or synthetic control that renders the linear differential equation $y^r = v$. Here, outputs are the rotor flux φ_r^2 and speed ω_r which is to be controlled.

$$\varphi_r^2 = \varphi_{\alpha r}^2 + \varphi_{\beta r}^2 \quad (11)$$

Differentiating the rotor flux φ_r from (11) with respect to time and substituting the value of $\frac{d\varphi_{\alpha r}}{dt}$ and $\frac{d\varphi_{\beta r}}{dt}$ from (3) and (4) yields the following equation

$$\frac{d\varphi_r}{dt} = -\frac{r_r}{l_r}\varphi_r + \frac{l_m r_r}{l_r \varphi_r}(i_{\alpha s}\varphi_{\alpha r} + i_{\beta s}\varphi_{\beta r}) \quad (12)$$

Using (5) and (6), speed state equation can be written as

$$\frac{d\omega_r}{dt} = -\frac{B}{J}\omega_r + \frac{1}{J}K_T(\varphi_{\alpha r} i_{\beta s} - \varphi_{\beta r} i_{\alpha s}) - \frac{T_L}{J} \quad (13)$$

State equations (12) and (13) are of rotor flux φ_r and speed ω_r respectively. Both are coupled by the two stator current control inputs $i_{\alpha s}$ and $i_{\beta s}$ which are decoupled by the feedback linearization I/O control technique with the help of two new control input v_1 and v_2 which are output of the PI controller. Let us define these new control inputs as

$$v_1 = \varphi_{\alpha r} i_{\beta s} - \varphi_{\beta r} i_{\alpha s} \quad (14)$$

$$v_2 = \frac{1}{\varphi_r}(i_{\alpha s}\varphi_{\alpha r} + i_{\beta s}\varphi_{\beta r}) \quad (15)$$

Now the state equaion (12) and (13) of the speed and flux as a function of new defined control input can be written as

$$\frac{d\omega_r}{dt} = -\frac{B}{J}\omega_r + \frac{1}{J}K_T v_1 - \frac{T_L}{J} \quad (16)$$

$$\frac{d\varphi_r}{dt} = -\frac{r_r}{l_r}\varphi_r + \frac{l_m r_r}{l_r} v_2 \quad (17)$$

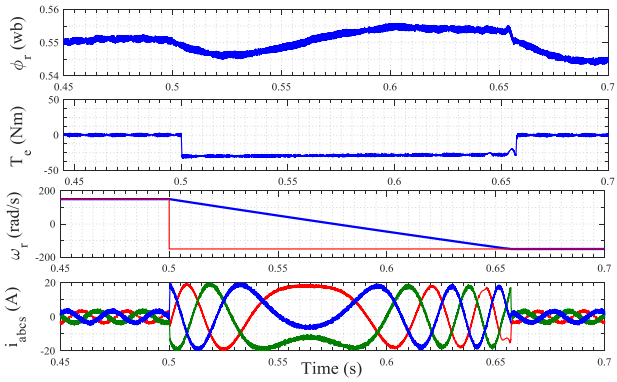


Figure 2. Speed reversal dynamics

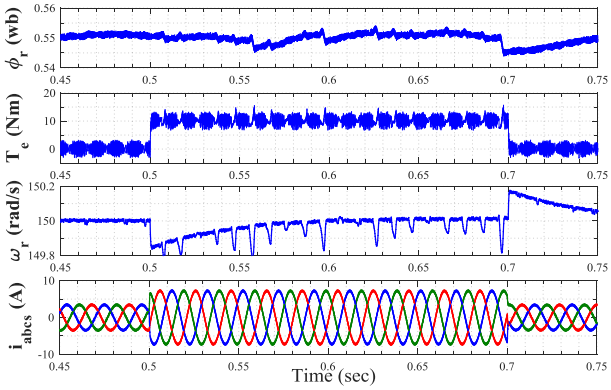


Figure 3. Sudden change in load dynamics

For non-zero value of rotor flux φ_r which is zero at the starting of the motor, exactly decoupled stator currents $i_{\alpha s}$ and $i_{\beta s}$ can be formulated with the help of feedback linearization method as

$$i_{\alpha s} = \frac{\varphi_{\alpha r}}{\varphi_r} v_2 - \frac{\varphi_{\beta r}}{\varphi_r^2} v_1 \qquad (18)$$

$$i_{\beta s} = \frac{\varphi_{\beta r}}{\varphi_r} v_2 + \frac{\varphi_{\alpha r}}{\varphi_r^2} v_1 \qquad (19)$$

IV. PREDECTIVE CURRENT CONTORLLER

A. Objective Function

Induction motor model is considered as a combination of resistive, inductive and back emf load for VSI which is used to predict the nature of the stator currents for every switching state generated by VSI. Objective function is defined for the selection of the best switching state which has to be applied to VSI and this objective function is calculated for the predicted values to be controlled. Error between predicted stator current and reference current generated by control algorithm is minimized using the cost function defined with PCC. The best switching state which gives the minimum value of cost function are selected and applied to the three-leg VSI without any use of modulation stage and linear regulators.

The error between reference currents produced using FLC and predicted stator currents are measured by cost function which is expressed in vector form as

$$\vec{g} = \vec{\imath^*}(k+1) - \vec{\imath^p}(k+1) \qquad (20)$$

Were $\vec{\imath^*}(k+1) = \vec{\imath_\alpha^*}(k+1) + j\vec{\imath_\beta^*}(k+1)$ and $\vec{\imath^p}(k+1) = \vec{\imath_\alpha^p}(k+1) + j\vec{\imath_\beta^p}(k+1)$. Here $\vec{\imath^*}(k+1)$ is the reference stator current vector produced using FLC algorithm, $\vec{\imath_\alpha^*}(k+1)$ and $\vec{\imath_\beta^*}(k+1)$ are the real and fictitious constituent of this current vector. $\vec{\imath^p}(k+1)$ is the predicted stator current vector. Currents $\vec{\imath_\alpha^p}(k+1)$and $\vec{\imath_\beta^p}(k+1)$ are the real and fictitious constituent of the current vector $\vec{\imath^p}(k+1)$ for a given voltage vector. Subscript 'p' is used for the predicted variables.

Now, the objective function in $\alpha\beta$ co-ordinate system can be defined for every voltage vector as

$$g = \left|i_\alpha^*(k+1) - i_\alpha^p(k+1)\right| + \left|i_\beta^*(k+1) - i_\beta^p(k+1)\right| \qquad (21)$$

B. VSI Modelling

Three phase voltage source inverter consists of six switches (T1, T2…T6) forming 3-leg of VSI and each leg of the inverter are controlled in complementary mode. Different switching states of the VSI are governed by the gating signals T_a, T_b and T_c as follows

$$T_a = \begin{cases} 1 \text{ if } T_1 \text{ on and } T_4 \text{ off} \\ 0 \text{ if } T_1 \text{ off and } T_4 \text{ on,} \end{cases} \quad T_b = \begin{cases} 1 \text{ if } T_2 \text{ on and } T_5 \text{ off} \\ 0 \text{ if } T_2 \text{ off and } T_5 \text{ on} \end{cases}$$

$$T_c = \begin{cases} 1 \text{ if } T_3 \text{ on and } T_6 \text{ off} \\ 0 \text{ if } T_3 \text{ off and } T_6 \text{ on} \end{cases}$$

The value of output voltage generated by inverter with different switching signals can be defined as $v = V_{dc}T$. Where, $T = \frac{2}{3}(T_a + aT_b + a^2T_c)$, is the inverter switching state vector and $v = \frac{2}{3}(v_{aN} + av_{bN} + a^2v_{cN})$ is the output voltage vector generated by the inverter. Voltage V_{dc} is the DC link voltage, $a = e^{j2\pi/3}$ is the unitary vector which gives 120° phase shift in different phases.

C. Discrete Time Model for Prediction

A model considering IM as load of the inverter is developed here. Equation is derived from equivalent circuit of the IM can be written as

$$v = ir_s + l_s\frac{di}{dt} + e \qquad (22)$$

Where v and I are the voltage and current vector respectively produced at the inverter terminal which is also stator voltage and stator current for the IM, e is the IM back emf vector. At the kth sampling instant, stator voltage and current are predicted using discrete time model of the stator current. Sampling time used for the entire discretization process is Ts.

Using the forward Euler approximation, stator current derivative $\frac{di}{dt}$ can be written as

$$\frac{di}{dt} \approx \frac{i(k+1) - i(k)}{T_s} \qquad (23)$$

For every voltage vector $v_x(k)$ which is produced by the inverter, future stator current can be predicted by substituting the value of $\frac{di}{dt}$ in (22) from (23) as

$$i^p(k+1) = \left(1 - \frac{r_s T_s}{l_s}\right) i(k) + \frac{T_s}{l_s}(v(k) - e(k)) \qquad (24)$$

Using equation (33) and (34), back-emf can be calculated as

$$e(k-1) = v(k-1) - \frac{l_s}{T_s} i(k) - \left(r_s - \frac{l_s}{T_s}\right) i(k-1) \quad (25)$$

Using an extrapolation of past back-emf $e(k-1)$, the present value $e(k)$of the back emf can be estimated. On the other-hand, the value of sampling frequency is much larger than the frequency of the back-emf and it is assumed that back emf will not change significantly in one sampling period. Therefore it is assumed that $e(k) = e(k-1)$.

D. Predicitve Current Control Algorithm

Fig.1 shows the control algorithm for predictive current controller used for FLC IM drive system. Two PI controllers are used to decouple the torque and flux. Input to the one PI controller is error of speed and input the second PI controller is the error of flux, output of these controllers are fed to the feedback linearized controller to produce torque and speed component of reference currents. Current i_{abc} is measured from the stator terminal of IM and the reference current i^*_{abc} is calculated from FLC technique. Back emf is calculated with the help of discrete stator current and voltage at the motor stator terminal. Predicted current $i^p(k+1)$ is determined for back emf $e(k)$ and each voltage vector $v(k)$. Error between predicted stator currents and reference current generated using control algorithm is minimized for each voltage vector for each sampling interval. The best switching state which gives the minimum value of cost function are selected and applied to the three-leg VSI.

V. SIMULATION RESULTS AND DISCUSSIONS

Simulation using MATLAB/Simulink is carried out for FLC IM drive with PCC. Parameters of IM drive used for simulation is shown in Table I.

A. Starting and Speed Reversal Dynamics of FLFOC IM drive

Fig. 2 shows the starting and speed reversal dynamics of FLC IM drive with PCC. Initially, induction motor is started without load with reference speed at 150 rad/s. Time taken by the motor to reach steady state speed 150 rad/s is 0.29s with overshoot in the speed is 0.428 rad/s. At t=0.5s, a sudden speed reversal command is given to the motor. As soon as speed reversal command is initiated, speed of the motor decreases and reaches to zero and then reverses its direction of rotation in opposite direction. Time taken by the motor to reverse its direction from 150 rad/s to -150 rad/s is 0.36s.

In reversal of speed, it is observed that torque is also developed in reverse direction. It results into increase in magnitude, decrease in frequency and change in phase sequence of the stator current. After reversal of the speed, torque of the motor comes down to zero and current drawn becomes equal to the no load current of the motor. From Fig. 2, it is seen that flux remains near to constant in the entire process.

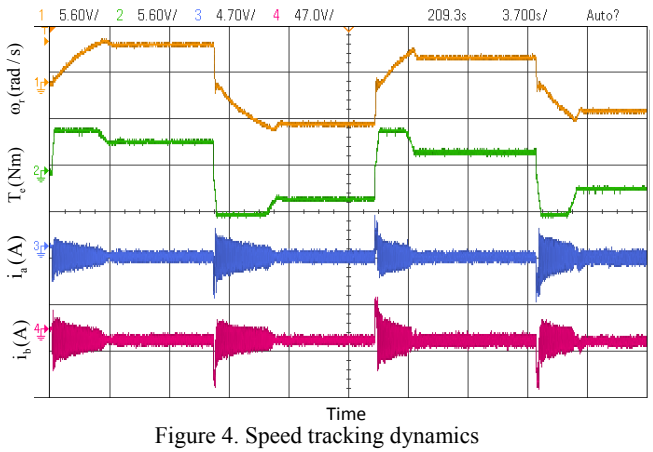


Figure 4. Speed tracking dynamics

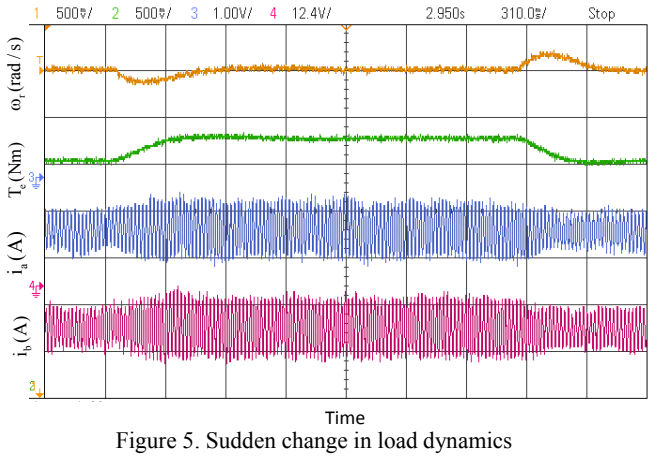


Figure 5. Sudden change in load dynamics

B. Dynamics Performance of FLC IM Drive with Sudden Change in Load

Fig.3 show the dynamics of the FLC IM drive with PCC when sudden load is applied to it. Initially, motor is run without any load with reference speed at 150 rad/s. Time required by the motor to reach at steady state speed 150 rad/s 0.29s with an overshoot of 0.428 rad/s. When motor reach steady state speed 150 rad/s at t=0.5s, sudden load Tl=10 Nm is applied to the motor. It is observed that drop in the speed of the motor is 0.162 rad/s with increase in the stator current to fulfil the demand of the load torque. Time taken by the rotor to regulate at its steady state speed is 0.098s. At t=0.7s, suddenly load torque is removed and overshoot of 0.162 rad/s in the speed is observed. Time taken by the motor to regulate at its steady state speed is 0.146s. When motor is running at reference speed 150 rad/s, torque developed is equal to zero. As soon as load torque Tl = 10 Nm is applied on the motor shaft, torque developed reaches to the load torque in 0.0003s to fulfil the demand of the load when FLC IM drive is operated. At t=0.7s, suddenly load is removed and therefore, torque developed comes down to zero in 0.0005s as shown in Fig. 3.

VI. Experimental Results and Discussions

Experimental setup is developed in laboratory using squirrel cage IM coupled to DC generator and with tachogenerator. DC generator is coupled with the IM and electrical load is applied at the armature terminal of the DC generator for loading of IM. Tachogenerator is also coupled on the same shaft to get the rotor speed of the motor. Switching pulses coming out from the dSPACE is first sent to a level shifter to ensure the sufficient voltage level of switching pulses applied to the IGBT switches of the inverter. Experiment is performed at sample period of 50 µs.

Figure 4-5 shows the experimental results of starting, speed tracking with reversal and step change in load dynamics. Results shows the waveform of the rotor speed, torque and stator currents. These results validated the effectiveness of the proposed controller.

TABLE I. Parameters of IM

Parameters	Value
Rated power	3 HP
Rated speed(ω_r)	1430 rpm
Number of poles(P)	4
Stator phase voltage(V)	346
Rotor resistance per phase(R_r)	0.877 Ω
Stator resistance per phase (R_S)	1.47 Ω
Rotor inductance per phase(L_r)	0.16514 H
Stator leakage inductance per phase(L_s)	0.16514H
Magnetizing inductance per phase (L_m)	0.1608H
Rotor inertia(J)	0.0015Kgm^2
Frequency(f)	50 Hz

VII. Conclusion

The FLC schemes of the IM gives exact decoupling of the motor speed and rotor flux amplitude control in both dynamic and steady states. Predictive current control method which replaces other classical current control techniques is implemented for FLC IM drive in MATLAB/Simulink to generate controlled switching of VSI fed IM. Although the PCC is more complex mathematically but because of the development of the power converter and micro-electronics, this scheme can be easily implemented with FLC IM drive system. Simulation results demonstrates that the speed and torque dynamics of FLC IM drive is very good when sudden change in load is applied to the motor shaft as well as on sudden change in speed. From the results obtained, it is concluded that proposed PCC has better dynamic response without any use of linear, nonlinear regulators and modulators as needed in classical control techniques. PCC functioning is independent of the rotor speed but dependent on the stator resistance and inductance.

References

[1] A. De Luca and G. Ulivi, "Full linearization of induction motors via nonlinear state-feedback," in *Proc. 26th conf. decision and control*, Los Angeles Dec. 1987.

[2] J. Dong-II, I. -J. Ha and M. Sam, "Control of induction motors for high dynamic perfromance and maximal power eficiency via feedback linearization with input output decoupling," pp. 324–329, 1989.

[3] A. De Luca and G. Ulivi, "Design of an exact nonlinear controller for induction motors," *IEEE Trans. Automat. Contr.*, vol. 34, no. 12, pp. 1304–1307, 1989.

[4] Ambrish Devanshu, Madhusudan Singh & Narendra Kumar (2018) Sliding Mode Control of Induction Motor Drive Based on Feedback Linearization, IETE Journal of Research, DOI: 10.1080/03772063.2018.1486743

[5] Y. Kuroe and Y. Yoneda "Design of a new controller for induction motors based on exact linearization," *in Proc. IECON* no. 1, pp. 621–626, 1991.

[6] C. Chong, W. Yan and W. Guanxiong, "Nonlinear decoupling control and exact linearization of induciuton motor drive using microprocessor," *in Proc. IECON,* 1991.

[7] I. Kanellakopoulos, P. T. Krein, and F. Disilvestro, "Nonlinear Flux-Observer-Based Control of Induction Motors," in *Proc. Am. Control Conf.* 1992, pp. 1700–1705.

[8] B. Chen, Y. Ruan. Y. Xu, and S. Gong, "Nonlinear decoupling control stratgy for variable frequency induction motor system fed by hysteresis band current controlled pwm inverter," in *Proc IEPEA* 1993 no. 3.

[9] R. Marino, S. Peresada and P. Valigi, "Adaptive input - output linearizing control of induction motors," *IEEE Trans. Auto. Control*, vol. 38, no. 2, february 1993

[10] T. Von Raumer, J. M. Dion and L. Dugard, "Applied nonlinear control of an induction motor using digital signal processing," *IEEE Trans. Control Syst. Technol.*, vol. 2, no. 4, pp. 327–335, 1994.

[11] A. Devanshu, M. Singh and N. Kumar, "DSP based feedback linearization control of vector controlled induction motor drive," *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, Delhi, 2016, pp. 1-6.

[12] M. P. Kazmierkowski and D. L. Sobczuk, "Sliding mode feedback linearized control of PWM inverter fed induction motor," in *Proc. IECON. 22nd Int. Conf. Ind. Electron. Control. Instrum.*, vol. 1, pp. 244–249, 1996.

[13] W. -J. Wang and C. -C. Wang,"A rotor flux observer based composite adaptive," *IEEE Trans. Energy Convers.*, vol. 12, no. 4, Dec. 1997.

[14] A. Rachid, "On induction motors control," *IEEE Trans. Control Syst. Technol.*, vol. 5, no. 3, pp. 380–382, May 1997.

[15] S. He, K. Relf and R. Unbehauen, "A neural approach for control of nonlinear systems with feedback linearization," *IEEE Trans. Neural Netw.*, vol. 9, no. 6, pp. 1409–1421, Nov. 1998.

[16] M. P. Kazmierkowski and L. Malesani "Current control techniques for three-phase voltage-source pwm converters: a survey," *IEEE Trans. Ind. Electron*, vol. 45, no. 5, Oct. 1998.

[17] H. Abu-Rub, J. Guziński, Z. Krzeminski, and H. A. Toliyat, "Predictive current control of voltage-source inverters," *IEEE Trans. Ind. Electron.*, vol. 51, no. 3, pp. 585–593, Jun. 2004.

[18] J. Rodriguez, J. Pontt, C. A. Silva, P. Correa, P. Lezana, P. Cortes and U. Ammann "Predictive current control of a voltage source inverter," *Ind. Electron. IEEE Trans.*, vol. 54, no. 1, pp. 495–503, Feb. 2007.

[19] J. M. Retif, L.-S. Xuefang and F. Morel, "Predictive current control for an Induction Motor," *IEEE Power Electron. Spec. Conf.* pp. 3463–3468, 2008.

[20] P. Cortes, M. P. Kazmierkowski, R. M. Kennel, D. E. Quevedo, J. Rodriguez,"Predictive control in power electronics and drives,"

cogent engineering

Probabilistic Approach Towards Darjeeling Himalayas Landslides- A Case Study

Sarkar Raju

Accepted Manuscript Version

This is the unedited version of the article as it appeared upon acceptance by the journal. A final edited version of the article in the journal format will be made available soon.

As a service to authors and researchers we publish this version of the accepted manuscript (AM) as soon as possible after acceptance. Copyediting, typesetting, and review of the resulting proof will be undertaken on this manuscript before final publication of the Version of Record (VoR). Please note that during production and pre-press, errors may be discovered which could affect the content.

© 2018 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

Publisher: Cogent OA

Journal: *Cogent Engineering*

DOI: http://dx.doi.org/10.1080/23311916.2018.1537539

Accepted Manuscript

Probabilistic Approach Towards Darjeeling Himalaya's Landslides- A Case Study

Raju Sarkar[1,2]

Center for Disaster Risk Reduction and Community Development Studies, Royal University of Bhutan, Rinchending, Bhutan
[2]Delhi Technological University, Bawana Road, Delhi, India

Abstract

The Himalayan region with its complex geological features, steep slopes, rugged topography and intense monsoonal rainfall create ideal conditions for landslides. The damage caused due to landslides is immense causing significant loss of life and property initiating a dire need to formulate strategies in minimizing its impact in areas affected by landslides. Several attempts have been presented by various researchers in order to establish rainfall intensity thresholds using various parameters of rainfall conditions. These methods provide deterministic thresholds i.e., landslide or no landslide, such thresholds are not always suited for landslides. In this paper, rainfall thresholds have been evaluated using a statistical method which results in the probability of landslide occurrence for single or multiple rainfall parameters leading to slide initiation. The results are expressed in terms of probabilities by analyzing two different variants of Bayes theorem i.e. 1D and 2D. Probabilistic thresholds were calculated for Kalimpong region of Darjeeling Himalayas using available rainfall and landslide data during the year 2010-2016. The probabilities calculated for landslide occurrence was found to be 0.37 for rainfall intensity greater than 10 mm/day. However, the probability for a combination of rainfall intensity of 30 mm/day with duration of 3 days was calculated to be 0.67. The results also depicted that landslides are related to rainfall event parameters especially with rainfall intensity.

Keywords: Landslides, Early Warning, Rainfall Thresholds, Darjeeling Himalayas, Probabilistic Method

1. Introduction

Landslides triggered by incessant rains, have been havoc in Indian Himalayas claiming several lives and displacing thousands of people besides damaging crops, properties and all means of communication. The number of deaths caused due to landslides stands third among the natural disasters. The Himalayan region accounts for 30% of the worldwide occurrence of landslides. Estimates indicate loss of land due to slides is up to 120 meter per km per year leading to yearly soil loss of about 2500 tons per sq km (NDMA report, 2011). Kalimpong is a hilly region situated in West Bengal state, India and is heavily affected by landslides. Most of the landslides are triggered during the monsoon season of June and October.

Rainfall-induced landslides originate due to the increase of pore water pressure in soil (Campbell, 1975; Wilson and Wieczorek, 1995). The reasons for an increase in pressure or changes in

Accepted Manuscript

moisture content in soil are responsible for landslides (Wilson and Wieczorek, 1995). Landslide prediction becomes difficult if the properties of soil parameters and slope conditions change greatly over short distances which make it impossible to forecast the timing, location and intensity of triggering events. Apart from heterogeneity of soil the other challenging factor is to analyse the mechanics of groundwater flow and the development of instability along with stability models (Wieczorek and Guzzetti, 2000). The actual failure of a slope is complicated involving numerous causes which influence hydrologic behavior and the shear stresses involved. Therefore a direct relation between rainfall and slope failure is difficult to formulate. Several studies have been done all over the world to develop relation between rainfall thresholds and landslide incidences using empirical, hydrological or physical (process-based) models (Capparelli and Versace, 2011; De Luca and Versace, 2017). The physical rainfall threshold models combine slope features, soil characteristics along with a steady or transient groundwater flow model (Baum et al., 2002; Chen and Zhang, 2014). Physical models integrate hydrological models for a simplified description of the dynamics of infiltration and saturation phenomena, together with geotechnical approaches for the stability analysis. Physical models usually attempt to account for infiltration and vertical movement of water into the ground surface, for modelling the longitudinal trend of groundwater.

Several rainfall thresholds have been devised using empirical models which correlate antecedent rainfall with landslide occurrence. The rainfall threshold relationship is fitted to an empirical equation $I = \square D^{-\square}$ (where I is the rainfall intensity (mm/h), D is the rainfall duration (h), α is a scaling parameter (the intercept) and β is the shape parameter that controls the slope of the power law threshold curve). Several rainfall thresholds for Indian Himalayan scenario have been proposed (Kanungo and Sharma, 2014; Dikshit and Satyam, 2018), using the threshold equation for local, regional, and global scales. However, since rainfall is not the only factor which causes slides, a certain extent of uncertainty is inevitable for calculation of rainfall thresholds (González, Caetano, 2017). The accuracy of empirical thresholds is restricted with the availability of rigorous rainfall and landslide data. Lack of data may increase uncertainty in determining thresholds using empirical methods. Moreover the results using empirical approach can be impacted due to several factors like the rainfall data resolution, rain gauge location including the landslide timing and location. In order to overcome the limitations, Bayesian approach is used to determine the probability of landslide occurrence with respect to rainfall event characteristics. The study area is Kalimpong town in Darjeeling Himalayas in the state of West Bengal which has a history of landslides caused mainly due to rainfall.

2. Methodology Adopted

Limit or threshold is characterized as the base or most extreme level of some amount required for a process to happen or a state to change (White et al., 1996). For precipitation instigated landslides, a threshold represents the minimum intensity, the minimum level of pore water pressure, soil moisture, slope angle, decrease in shear strength, or displacement required for a landslide initiation (Wilson and Wiezerock, 1995; Guzzetti et al., 2007).

Fig. 1a, describes deterministic precipitation thresholds where there is a distinct isolation amongst triggering and non triggering precipitation for slide initiation. Deterministic thresholds support a binary decision either 0 (no landslides) or 1 (landslides occurred) which in most cases the distinction isn't inconsequential (Fig. 1b). The reasons for landslide initiation are complex and are usually controlled by a combination of precipitation parameters and factors such as soil moisture,

Accepted Manuscript

pore pressure distribution, soil type, and field stress affect landslide occurrence (Berti et al., 2012). Likelihood based techniques are worthwhile as they coordinate different precipitation parameters and uncertainty giving a quantitative appraisal and better estimating of threshold limit unlike the definitive forecast of deterministic methods (Berti et al., 2012). Therefore if the output is not limited to either failure or no-failure condition for a given rainfall event, deterministic approach is not relevant and probabilistic model should be used.

While analyzing probability there may be several cases where a single rainfall event could cause multiple landslides in the same area. Multiple cases of landslide due to single rainfall event can also be considered using Bayesian analysis by adding another variable which counts the number of landslides triggered due to each rainfall event. However, multidimensional analysis is confined due to insufficient data and its pertinence in practical applications is limited. Generally, the landslide probability is directly proportional to the extent of the area i.e, with increase in the area leads to increase in landslide incidence.

2.1. One-dimensional Bayesian Probability

Conditional probabilities are computed using Bayes' theorem which would determine the probability of landslide occurrence caused due to rainfall expressed in terms of total rainfall, intensity or duration. Conditional probability is given as P(A|B) which means "probability of landslide occurrence (A) due to a particular rainfall parameter (B)." and is determined by the Bayes' theorem as:

$$P(B|A) = (P(B|A)P(A))/P(B) \quad \text{-(1)}$$

- P(B|A) = conditional probability of B given A i.e., probability of rainfall event of magnitude B when a landslide happens
- P(A) = probability of landslide occurrence despite the rainfall event of magnitude B occurs or not.
- P(B) = probability of rainfall of magnitude B, whether a landslide happens or not.
- P(A|B)= conditional probability of A given B, i.e., probability of a landslide occurrence when a rainfall event of magnitude B happens.

Let the number of rainfall events during a time period be N_R; number of landslide occurrences during the same time period be N_A , number of rainfall events of magnitude B be N_B and the number of rainfall events resulting in landslides be $N_{(B|A)}$, then the terms in eq. 1 can be expressed as:

$$P(A) \approx N_A/N_R \quad \text{-(2)}$$

$$P(B) \approx N_B/N_R \quad \text{-(3)}$$

$$P(B|A) \approx N_{(B|A)}/N_A \quad \text{-(4)}$$

In general, probabilities would help in determining the likelihood of landslides. The choice of variables in one-dimensional analysis depend on parameters which are key reasons for slide initiation, for our study variables considered are event rainfall, rainfall duration, and rainfall intensity. A full description of the approach briefly summarized here can be found in (Berti et al., 2012).

Accepted Manuscript

2.2 Two-Dimensional Bayesian Probability

Two-dimensional Bayesian probability estimates the conditional probability of event for joint occurrence of two parameters considered.

$$P(A|B,C) = (P(B|C,A)P(A))/P(B,C) \quad -(5)$$

where B, C denotes combined probability of having a certain or range of value of any two variables. If, B equals rainfall intensity and C is rainfall duration, the probability of landslide occurrence due to rainfall event of given duration and intensity is expressed using Eq. 5. Any pair of precipitation parameters for likelihood of landslide occurrence can be calculated using two-dimensional Bayesian probability (e.g., rainfall intensity, total event rainfall, rainfall duration), and their significance can be analyzed by comparing landslide probability with prior landslide probability P(A). Baye's method can likewise be utilized for multidimensional analysis with n-variables like joined impact of rainfall duration, rainfall intensity and groundwater conditions to evaluate probability of landslide occurrence.

3. About the Study Area

Darjeeling and Sikkim Himalayas cover more than 40% or 0.18 million sq. km of landslide prone areas in India (Geological Survey of India (GSI)). The focus of study was Kalimpong region in Darjeeling Himalayas situated in West Bengal state with an average elevation of 1247 m comprising of 1056 km^2 (Fig. 2). It is situated on a curved ridge surrounded by Teesta river in the west central and Relli river in the east (Dikshit and Satyam, 2018). About half of the region is covered by moderately steep to steep and very steep slopes (>25^o) and 32% of the study area is represented by high relief (>300 m). Landslides in this region are primarily caused due to poor lithological quality, erosion of river Teesta at the toe and its tributaries and high intensity of rainfall during monsoon (Chatterjee, 2010). The report by GSI, 2016 identified that 75% of the landslide occurrences in this region during 2006-2013 was triggered by rainfall. Geologically, the region comprises of moderately to highly weathered chlorite schist, phyllite, phyllitic quartzite belonging to Gorubathan Formation of Daling Group. The rocks are variably altered and are generally covered by a thin to thick heterogeneous debris material (Dikshit and Satyam, 2018). The soil in the region are generally reddish with intermittent presence of dark soils comprised mainly of phyllites and schists. The soil comprises a high percentage of gravel, sand and silt and varies from rock outcrops to coarse loamy to fine loamy. The eastern side of the Teesta river mostly comprises of sandy soils with soil characteristics varying from coarse to rocky with the increase in elevation (Dikshit and Satyam, 2018). The region comprises several cracks, joints which increases the probability of decomposing and disintegrating the rock to form unconsolidated matter.

3.1 Landslides in Kalimpong

Majority of landslides in this region is due to heavy rainfall and untrained water streams known as jhoras in the region (Dikshit and Satyam, 2018) which loosens up the soil by breaking of bonds and particle disintegration causing sloped surface area to be fully saturated causing landslides. Due to rainfall, soil absorbs water which increases the burden on soil by rise in pore water pressure thus

Accepted Manuscript

increasing the probability of landslides. Landslides are triggered by seepage of water in joints and cracks and facilitate the soil/debris to slide downward due to gravity. In addition to rainfall, a network of small rivers of various orders significantly drains the area. There are majorly five sub-basins present in the region which flow down as tributaries of Teesta river Fig. 3 represents the drainage map of the area. The complete basin comprises of first order which joins to form second and higher order streams (Dikshit and Satyam, 2018). Oftentimes the first order streams disappear from some location in the hill and then reappear down slope. The major streams are mostly perennial supported by the rainwater which percolates ground water. Heavy rainfall, especially during monsoon, enables these streams to flow with great vigour causing them to carry small pebbles to huge boulders. The increase in the use of concrete and asphalt for construction activities in upper parts of hill forbid rainwater percolation into soil draining out the excess rainwater as surface runoff. The absence of planned drainage system in the new built up areas allow the rainwater to find passage into any of the natural tributaries eventually feeding the major streams. This causes intense lateral and headward erosion resulting into successive bank failures mostly in the form of large landslides (Chakraborty et al., 2011).

3.2 Rainfall and Landslide Data

For the present study, daily rainfall values for 2010 to 2016 were used for the analysis (Fig. 4). The measurement of rainfall for landslide investigations should be site specific to each slope failure but since this area is largely ungauged the rain gauge at Teesta was taken as reference. Total of 214 rainfall events occurred between 2010-2016.

Landslide data was prepared from published literature along with local newspaper reports and reports from nongovernmental organizations. The landslide information included the location and rainfall triggering date. The landslide inventory map showing the locations of the landslides, based on which the model has been assessed is depicted in Fig. 5. The majority of the landslides are concentrated over $27^{O}4'$ 30" N and $88^{O}28'$ 30" E whereas the rest are scattered.

4. Application of Bayesian Approach in the Present Study

One-dimensional analysis is used to compute probability of landslide occurrence when a certain amount of rainfall event occurs. It calculates the relation of a variable B for a certain event A. For the present study, A represents landslide occurrences and B represents any parameters which describes rainfall event, such as rainfall duration or intensity. As discussed in section 2.1, the correlation between landslide probability P(A|B) and prior landslide probability P(A) indicates significance of B for occurrence of event A (Berti et al., 2012).

4.1 With One-Dimensional Bayesian Probability

The methodology was applied to Kalimpong region with the procedure explained in section 2.1. Rainfall probability P(B) is calculated for N_R = 214 rainfall events occurred between 2010-2016, and the conditional probability P(A|B) is calculated for N_A=36 rainfall events that triggered landslides. $P(A) = N_A/N_R = 36/214 = 0.17$. P(B|A) is calculated by considering various rainfall intensity values, for rainfall intensity greater than 50 mm per day, P(B|A) = P(I > 50|A). Similar analysis was conducted for various intensity range ($0 \leq I < 50$, $50 \leq I < 100$, $I > 150$ mm/day) and the landslide probability is represented using histogram.

Accepted Manuscript

Out of 214 rainfall events occurred in the study period, 58 rainfall events and 23 landslide events occurred for intensity between 0-50 mm/day, P(B)= 58/214=0.27 (Fig. 6e). Probability of landslides is given as : P(A|B)=P(A|I<50)= 0.29 . 0.14/0.27 = 0.145 (Fig. 6f). The increase in dissimilarity between P(B|A) and P(B) indicates increase in probability of landslides and high significance of variable used (Berti et al., 2012).

The probabilities calculated considering various rainfall parameters are depicted in Fig. 6. The graphs on left compare frequency distributions of triggering rainfall against total rainfall, that is P(B|A) against P(B). The ratio of probabilities multiplied with P(A) provides probability of landslides P(A|B) depicted on right side. The results in Fig. 6 indicate that rainfall parameters like cumulative rainfall, duration, intensity are very important for determining landslide probability. In each case the probabilities P(B|A) and P(B) are distinct and the computed landslide probability is greater than prior probability P(A). Rainfall intensity among other variables is the most used and significant variable for threshold calculation, having probability as high as 0.37 for intensity greater than 10 mm/day.

Landslide possibility increases with extremity of rainfall parameters like increase in rainfall amount, duration, or intensity even though the increase may be uneven due to irregular distribution of data. However, at the peak values of the parameters considered landslide probability decreases (Fig. 6(a)–(f)). This uncommon trend is generally due to two reasons. Firstly, calculated probabilities of such severe events are affected due to low sample sizes. Samples with little data are less informational and a little variation in the recorded number of landslides would result in very different probability. In order to understand the impact of such uncertainty the difference between P(A|B) and P(A) is depicted to understand the significance of each variable used.

4.2 With Two-dimensional Bayesian Probability

The landslide probability (two-dimensional probability) increases with both rainfall duration and intensity, but intensity affects the landslide probability more as shown in Fig. 7. The highest value of probability of 0.67 reaches for rainfall events for duration of 3 days with an intensity of 30 mm/day and also for rainfall of 4 days with intensity of 20 mm/day. The importance of such analysis can be understood from the fact that even a small probability cannot be neglected for vulnerable areas and has to be analyzed with the risk associated for each probability.

The two-dimensional approach is more reliable compared to one-dimensional as it involves a combination and also determines the most important parameter. As the results for one-dimensional signify that the rainfall intensity is the most important parameter, however when considering two-dimensional it shows that apart from rainfall intensity, rainfall duration also plays an important part. The analysis also suggests that the use of two-dimensional can be beneficial if the sample size is large as determining probabilities from small sample size can lead to contrasting results with a small change.

5. Results and Discussions

Probabilities were calculated for landslide occurrences using various rainfall parameters like rainfall intensity, event rainfall and rainfall duration. In case of event rainfall, the probability

Accepted Manuscript

variations are 0.35, 0.45, 0.60, 0.80, 0.70, 0.24 for 50, 100, 150, 200, 250, 300 mm of event rainfall, respectively (Fig. 6(a)-(b)). Similarly, in case of rainfall duration the probability of landslides are 0.02, 0.025, 0.24, 0.17, 0.18, 0.76, 0.245, 0.175, 0.98 for 1, 2, 3, 4, 5, 6, 7, 8, 9 days of rainfall, respectively (Fig. 6(c)-(d)). Fig. 6 (e)-(f) depict that in case of rainfall intensity probabilities are 0.38, 0.19, 0.12, 0.08, 0.07, 0.16, 0.02 for 10, 20, 30, 40, 50, 60, 70 mm rainfall intensity, respectively. In case of two-dimensional probabilities, the probabilities achieved highest value of 0.67 for rainfall of duration 3 days with an intensity more than 30 mm/day whereas lowest value of 0.10 for rainfall duration of 1 day with intensity of 10 mm/day (Fig. 7).

6. Conclusions

Though various methods have been proposed in literature of the present paper to relate rainfall threshold with landslide occurrence, such approach may not always be helpful to understand the effect and to forecast landslides as it only considers the rainfall which resulted in landslides. The important question which arises using such methods is the effect of rainfall which did not result in landslides and its interdependency with various rainfall parameters. In order to overcome this limitation, Bayesian methodology was used which considers all the rainfall aspects and provides more clarity on landslide occurrence. The analysis was carried out for a period of 7 years (2010-2016) using various rainfall parameters including duration, intensity and event rainfall. The probabilities for landslide occurrences were determined using two approaches (one-dimensional and two-dimensional). The former approach uses a single rainfall parameter whereas the latter uses a combination of two rainfall parameters. The conclusions from the study are as follows:

i) The results signify that probability of landslide occurrence is 0.37 for rainfall intensity greater than 10 mm/day for one dimensional case and 0.67 for rainfall of duration 3 days with intensity more than 30 mm/day for two dimensional case.
ii) The use of two-dimensional approach provides a better understanding of the landslide incidents compared to one-dimensional. However, it would depend on the richness of the data. Samples with little data are less informational and a little variation in the landslide information would depict contradictory probability.
iii) The use of probabilistic approach over deterministic methods is a better alternative to set up an early warning system for landslide affected areas and can be used as a first line of action.

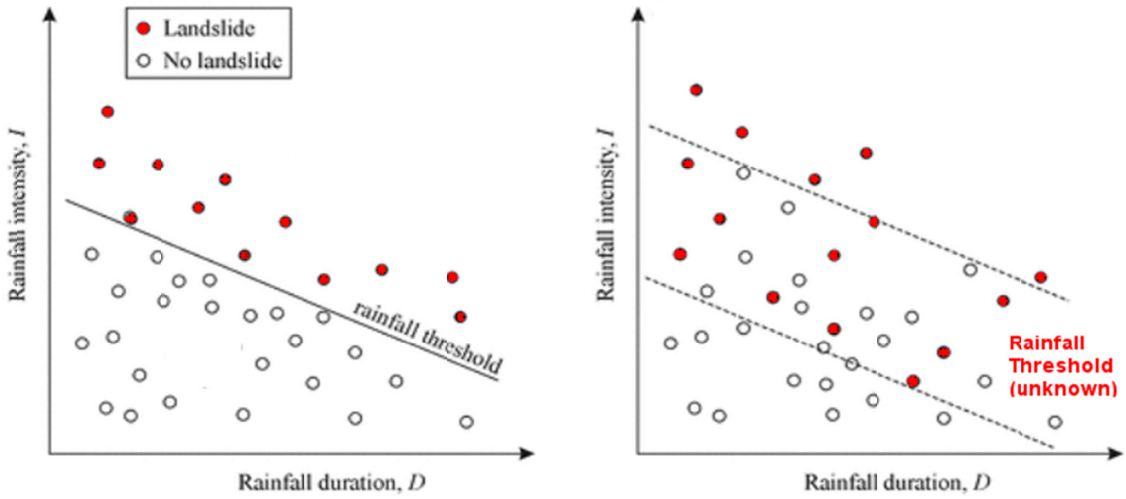
References:

1. Baum, R. L., Savage, W. Z., and Godt, J. W. (2002) TRIGRS – a FORTRAN program for transient rainfall infiltration and grid-based regional slope stability analysis, US Geological Survey Open-File Report 2002-424, 38 pp..
2. Berti, M., M. L. V. Martina, S. Franceschini, S. Pignone, A. Simoni, and M. Pizziolo (2012), Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach, J. Geophys. Res., 117, F04006, doi:10.1029/2012JF002367..
3. Campbell, R. H. (1975), Soil slips, debris flows, and rainstorms in the Santa Monica Mountains and vicinity, southern California, U.S. Geol. Surv. Prof. Pap., 851.
4. Chen HX, Zhang LM (2014) A physically-based distributed cell model for predicting regional rainfall-induced shallow slope failures. Engineering Geology doi. doi:10.1016/j.enggeo.2014.04.011

Accepted Manuscript

5. Capparelli, G. and Versace, P. (2011) FLaIR and SUSHI: two mathematical models for early warning of landslides induced by rainfall, Landslides, 8, 67–79
6. Chakraborty, I., Ghosh, S., Bhattacharya, D., & Bora, A. (2011). Earthquake induced landslides in the Sikkim-Darjeeling Himalayas—An aftermath of the 18th September 2011 Sikkim earthquake (8 p.). Kolkata: Report of Geological Survey of India.
7. De Luca, D. L. and Versace, P. (2017) A comprehensive framework for empirical modeling of landslides induced by rainfall: the Generalized FLaIR Model (GFM), Landslides, 14, 1009–1030, https://doi.org/10.1007/s10346-016-0768-5, 2017
8. Dikshit A., Satyam N. (2017) Rainfall thresholds for the prediction of landslides using empirical methods in Kalimpong, Darjeeling, India JTC1 Workshop on Advances in Landslide Understanding, Barcelona, Spain
9. Dikshit A, Satyam DN (2018) Estimation of rainfall thresholds for landslide occurrences in Kalimpong, India. Innov Infrastruct Solut 3:24–34
10. Ghoshal T.B., Sarkar N.K., Ghosh S. and Surendranath M. (2008) GIS Based Landslide Susceptibility Mapping- A Study from Darjeeling-Kalimpong Area, Eastern Himalaya, Kolkata Journal Geological Society of India Vol.72, December 2008, pp.763-773,
11. González, A. and Caetano, E. (2017) Probabilistic Rainfall Thresholds for Landslide Episodes in the Sierra Norte De Puebla, Mexico. Natural Resources, 8, 254-267. https://doi.org/10.4236/nr.2017.83014
12. Guzzetti, F., S. Peruccacci, M. Rossi, and C. P. Stark (2007), Rainfall thresholds for the initiation of landslides in central and southern Europe, Meteorol. Atmos. Phys., 98, 239–267, doi:10.1007/s00703-007-0262-7.
13. Guzzetti, F., M. Cardinali, and P. Reichenbach (1994), The AVI project: A bibliographical and archive inventory of landslides and floods in Italy, Environ. Manage., 18(4), 623–633, doi:10.1007/BF02400865.
14. Kanungo DP, Sharma S (2014) Rainfall thresholds for prediction of shallow landslides around Chamoli-Joshimath region, Garhwal Himalayas, India. Landslides 11(4):629–638.
15. Mukherjee, A. and Mitra, A.K.(2001) Geotechnical study of mass movements along the Kalimpong approach road in the Eastern Himalayas. Indian Journal of Geology Vol. 73, No. 4, p. 271-279,.
16. Ranit Chatterjee (2010) Landslide Hazard Zonation Mapping of Kalimpong, Publisher : VDM Verlag Dr. Muller GmbH & Co. KG
17. Sumantra SB, Raghunath P (2016) Causes of Landslides in Darjeeling Himalayas during June-July, 2015. J Geogr Nat Disast 6: 173. Doi: 10.4172/2167-0587.1000173
18. White ID, Mottershead DN, Harrison JJ (1996) Environmental systems, 2nd edn. Chapman & Hall, London, p 616
19. Wieczorek, G.W. and Guzzetti, F. (2000) A Review of Rainfall Thresholds for Triggering Landslides. In: Claps, P. and Siccardi, F., Eds., Mediterranean Storms, Proceedings Plinius Conference ’99, Maratea, Editoriale Bios, Cosenza, 14-16 October 1999, GNDCI pub. n. 2012, 407-414.
20. Wilson, R. C., and G. F. Wieczorek (1995), Rainfall thresholds for the initiation of debris flows at La Honda, California, Environ. Eng. Geosci., 1, 11–2

Accepted Manuscript



(a) (b)

Fig. 1. Rainfall intensity-duration thresholds cases (a) surface runoff and (b) deep-seated landslides (Berti et al., 2012).

Accepted Manuscript

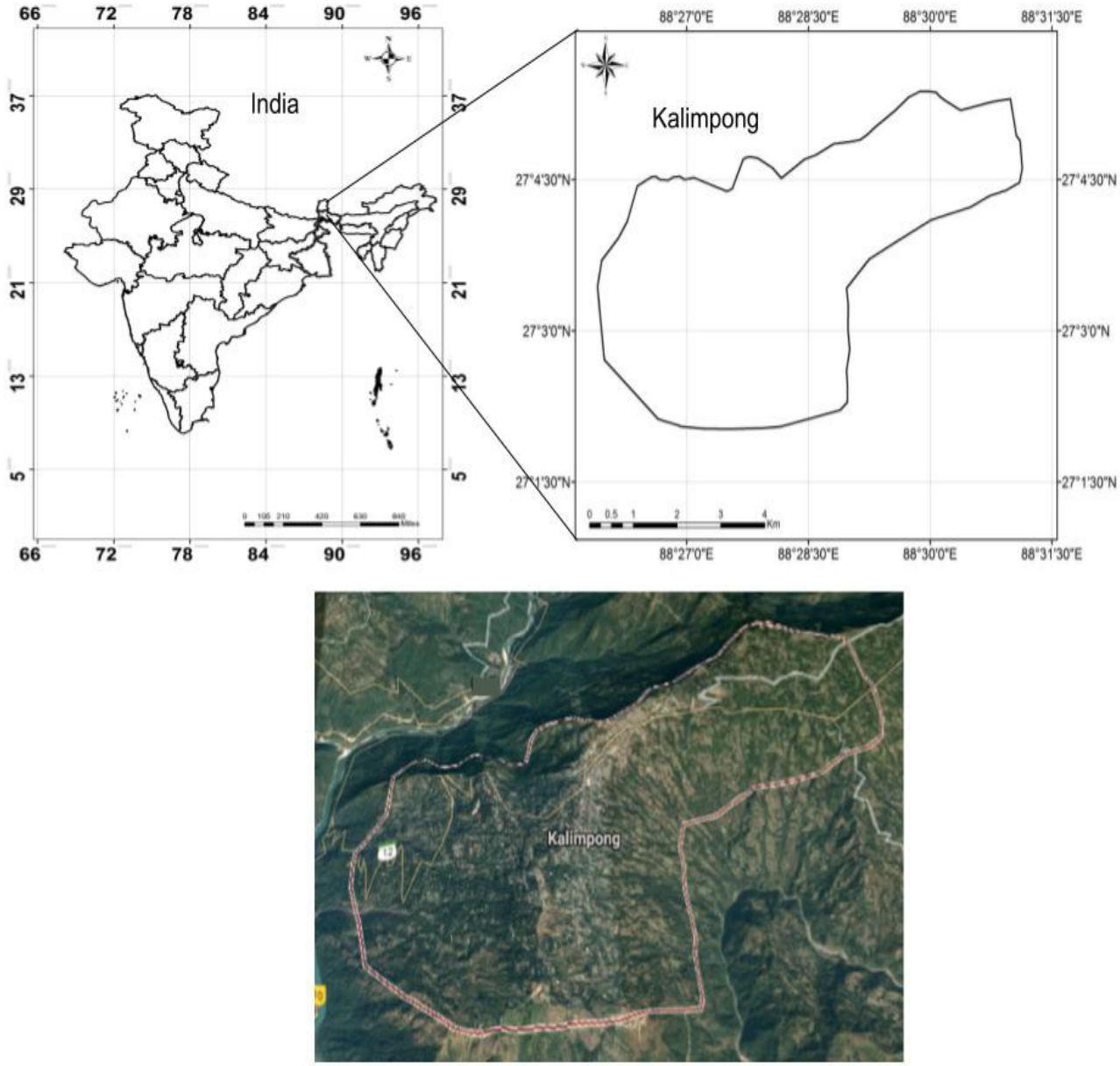


Fig. 2 Location map of study region

Acce

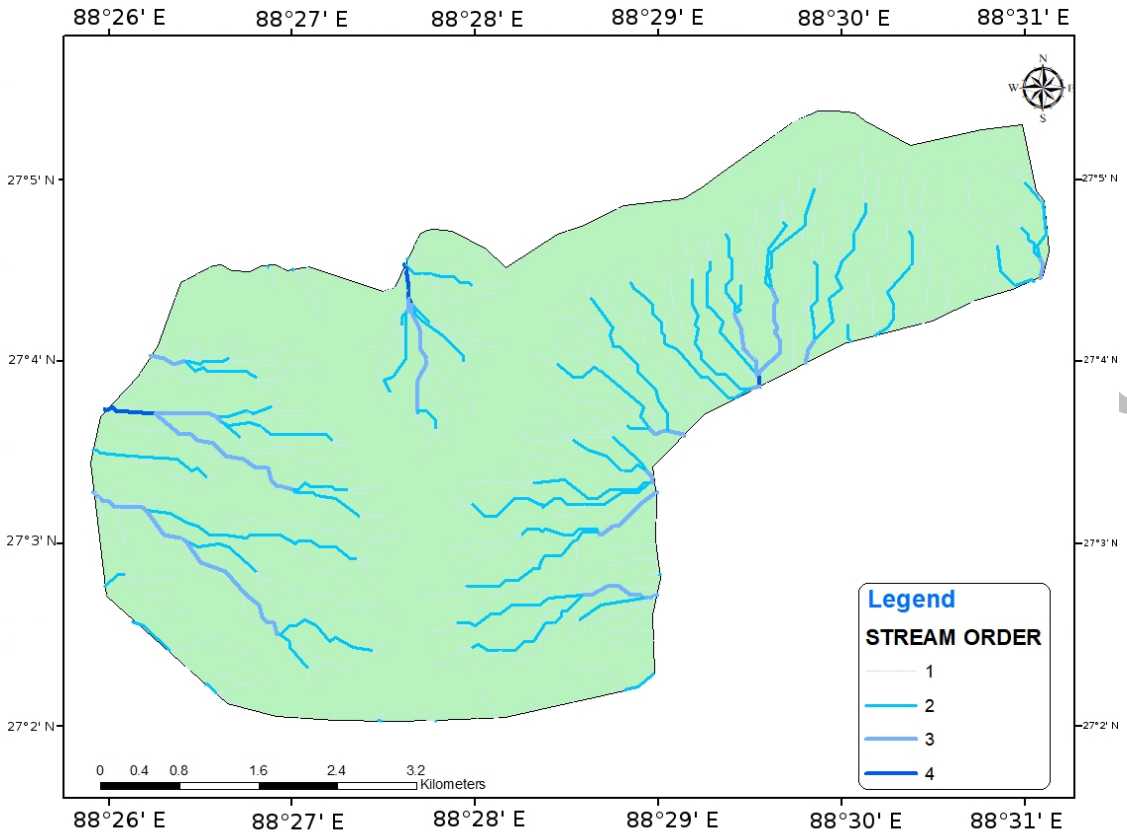


Fig. 3 Drainage Map of Kalimpong town

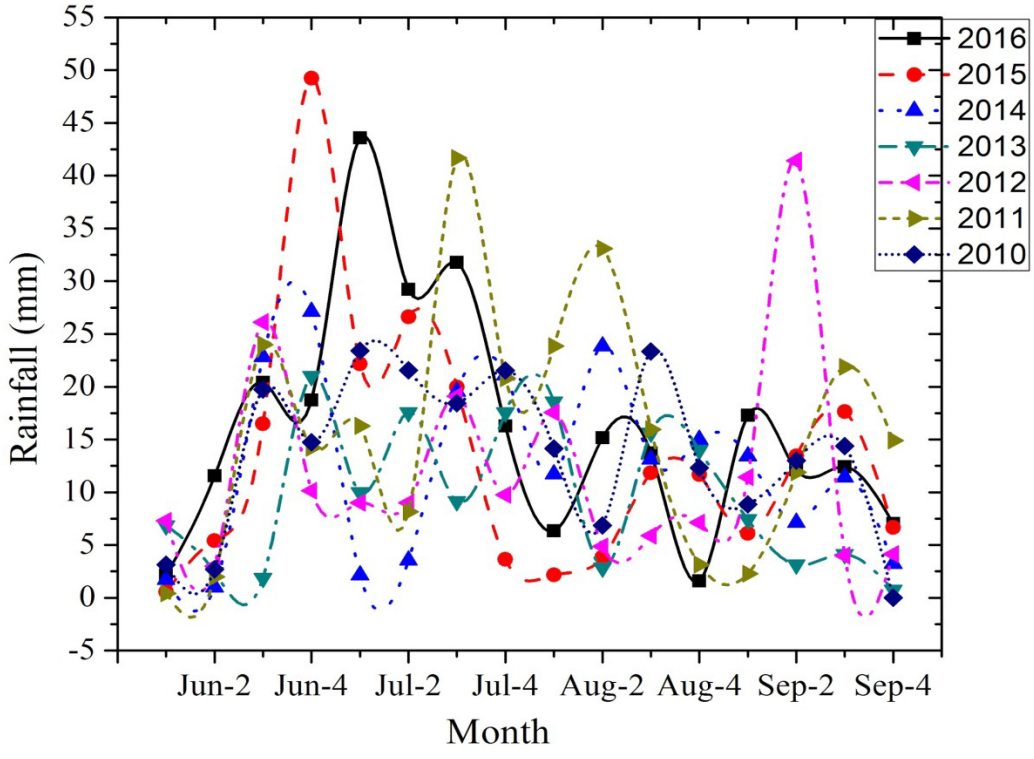


Fig. 4: Rainfall data for monsoon period (June-September, 2010-2016)

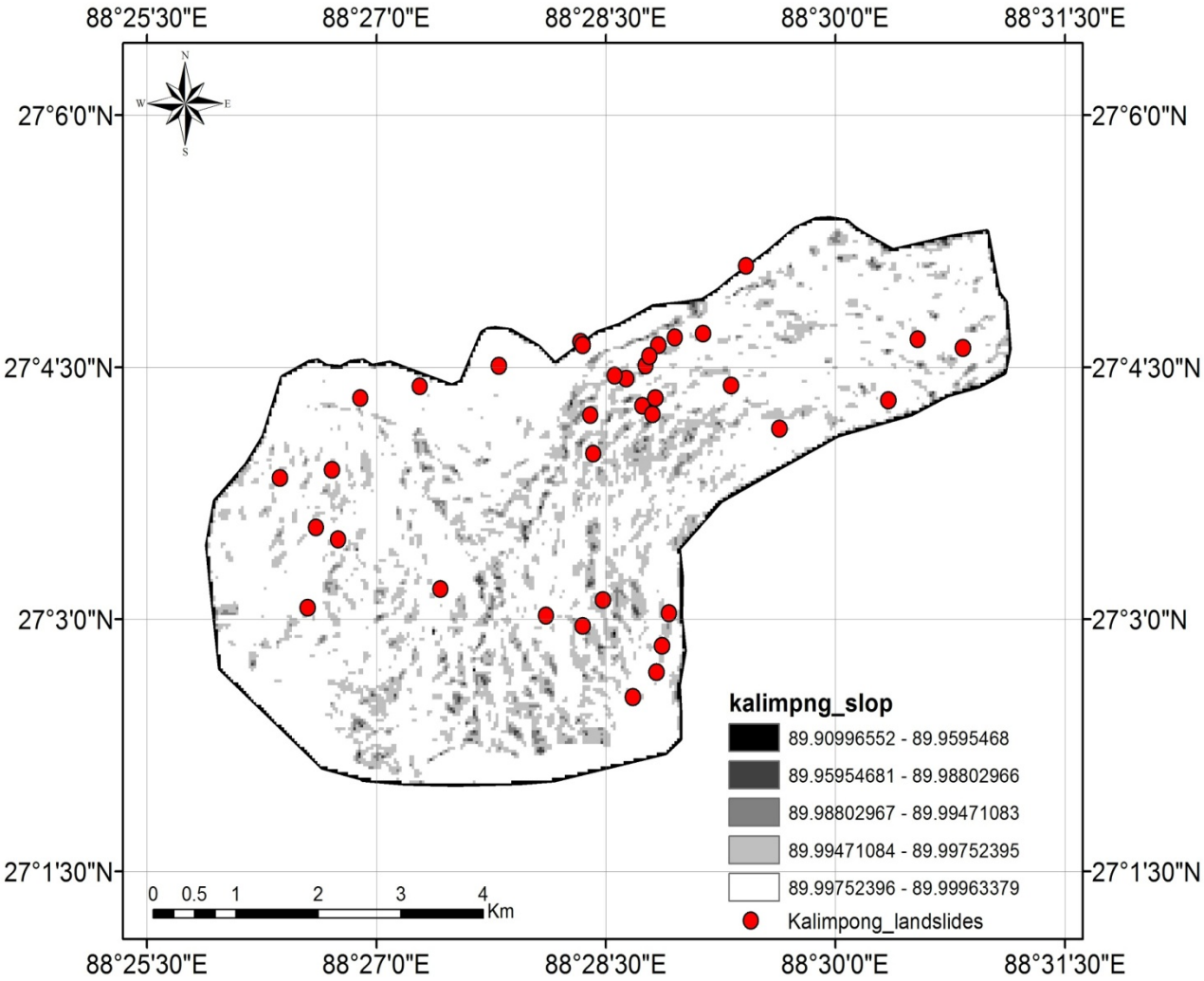
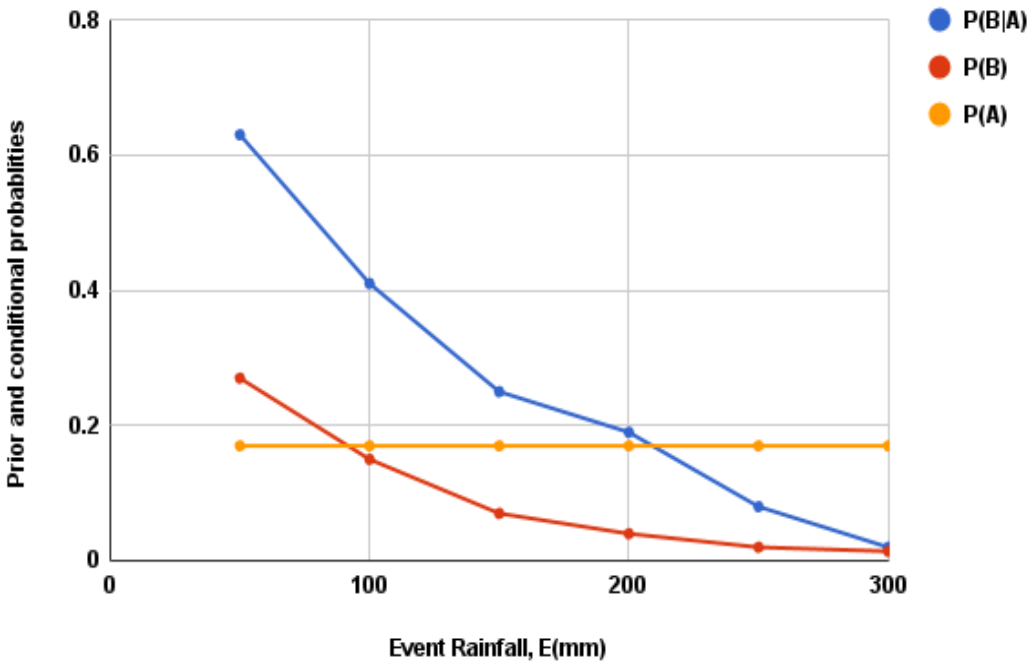
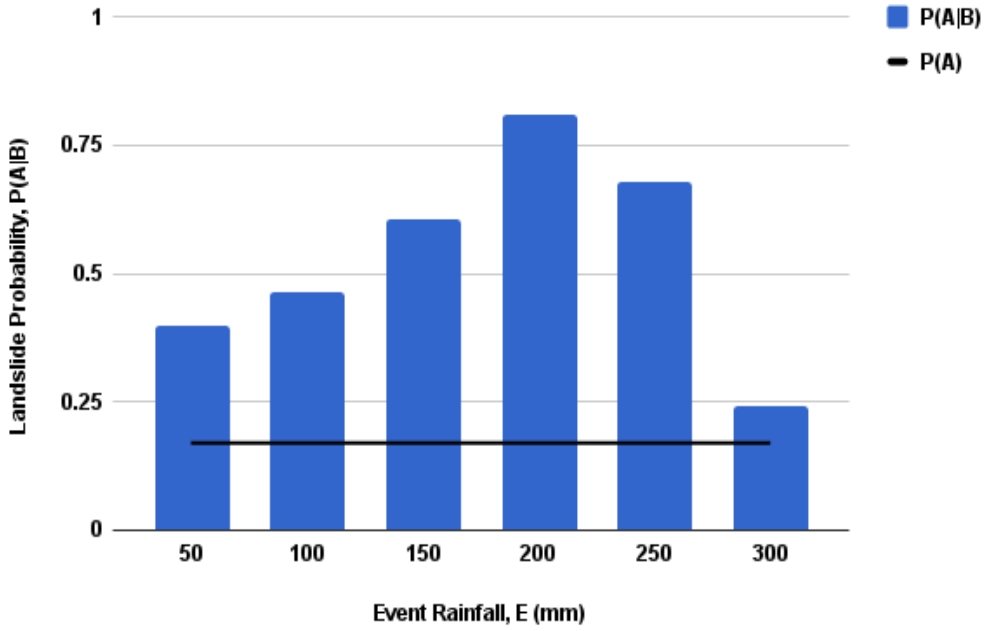


Fig. 5 : Landslide Inventory map

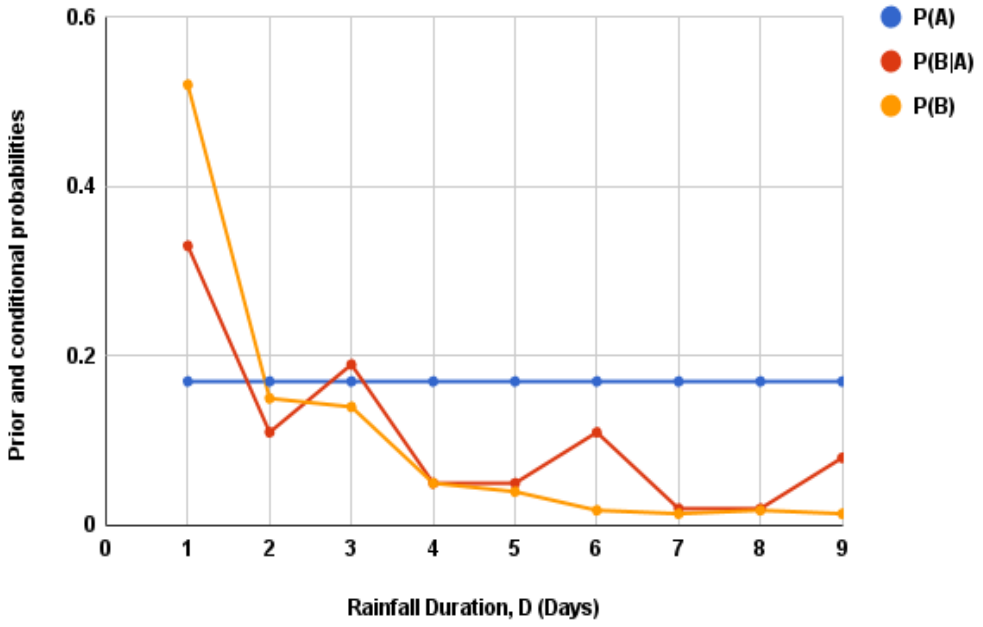
Accep


P(B|A)
P(B)
P(A)
Prior and conditional probabilities
0.8
0.6
0.4
0.2
0
0
100
200
300
Event Rainfall, E(mm)

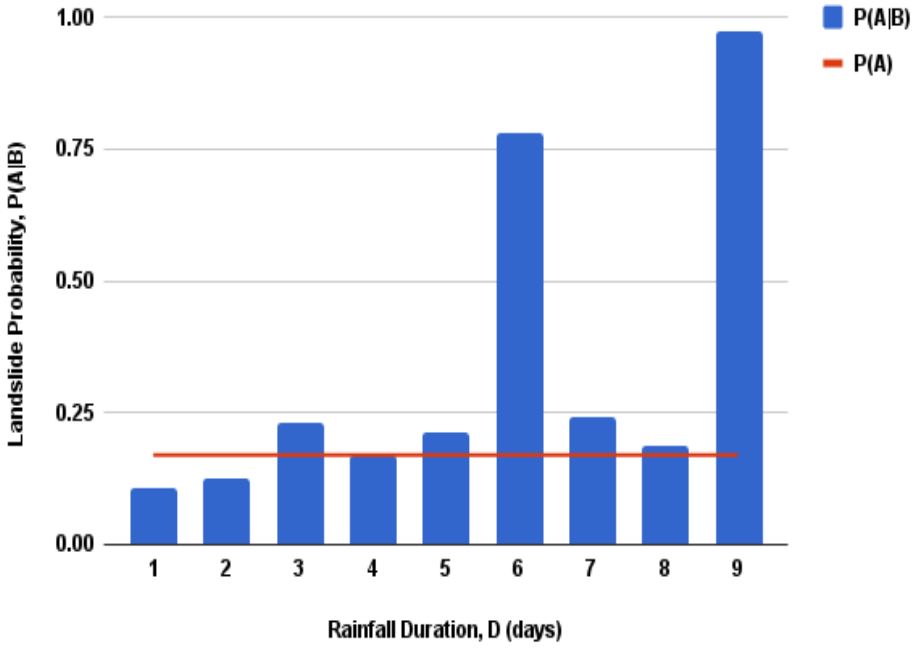
(a)


P(A|B)
P(A)
Landslide Probability, P(A|B)
1
0.75
0.5
0.25
0
50
100
150
200
250
300
Event Rainfall, E (mm)

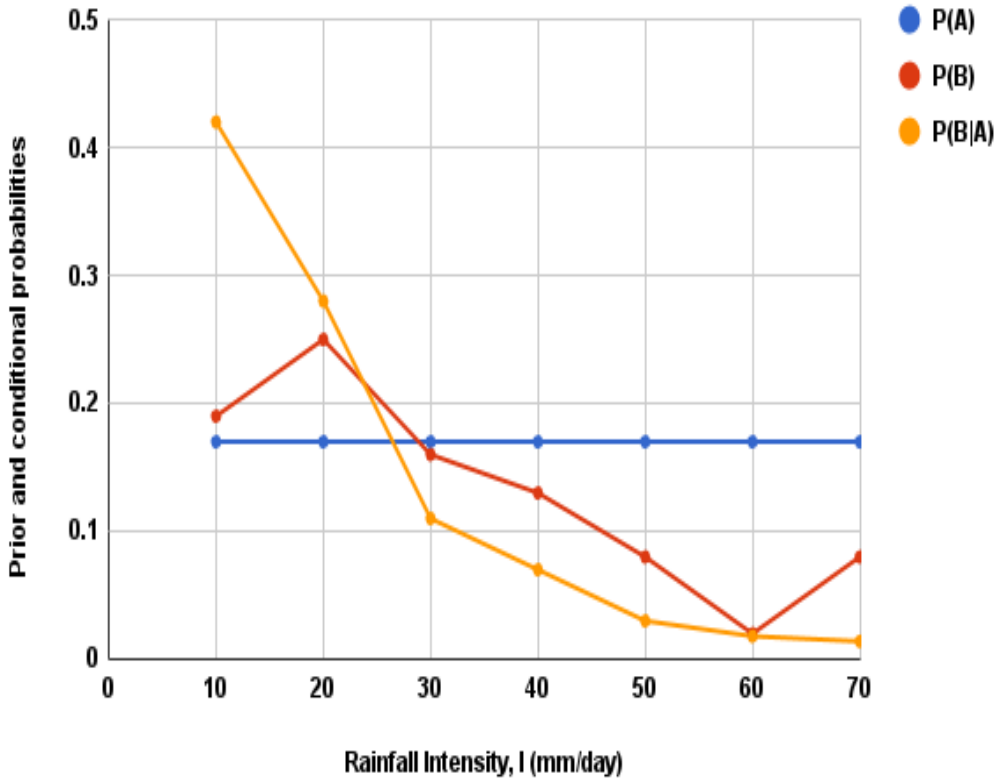
(b)


P(A)
P(B|A)
P(B)
Prior and conditional probabilities
0.6
0.4
0.2
0
0
1
2
3
4
5
6
7
8
9
Rainfall Duration, D (Days)

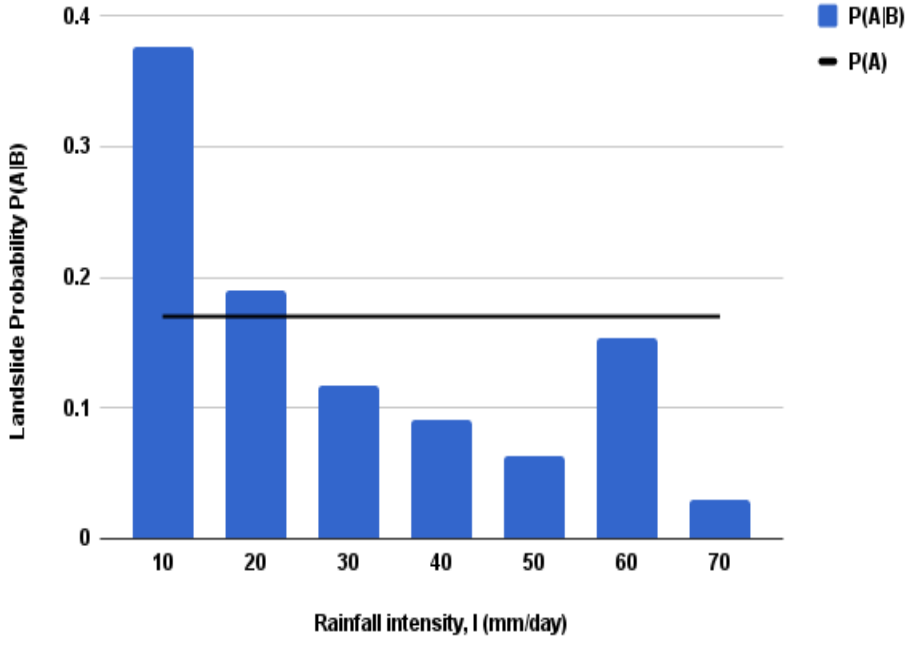
(c)


P(A|B)
P(A)
Landslide Probability, P(A|B)
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
Rainfall Duration, D (days)

(d)



(e)



(f)

Fig. 6: Computed probabilities for one-dimensional Bayesian analysis of Kalimpong region. Rainfall parameters considered are: (a–b) event rainfall, (c–d) rainfall duration, and (e–f) rainfall intensity. Graphs on the left depicts prior landslide probability P(A), prior rainfall probability P(B), and conditional probability P(B|A) for corresponding rainfall

parameter. Graphs on the right compare the calculated landslide probability P(A|B) with prior landslide probability P(A) to evaluate the significance of the considered variable.

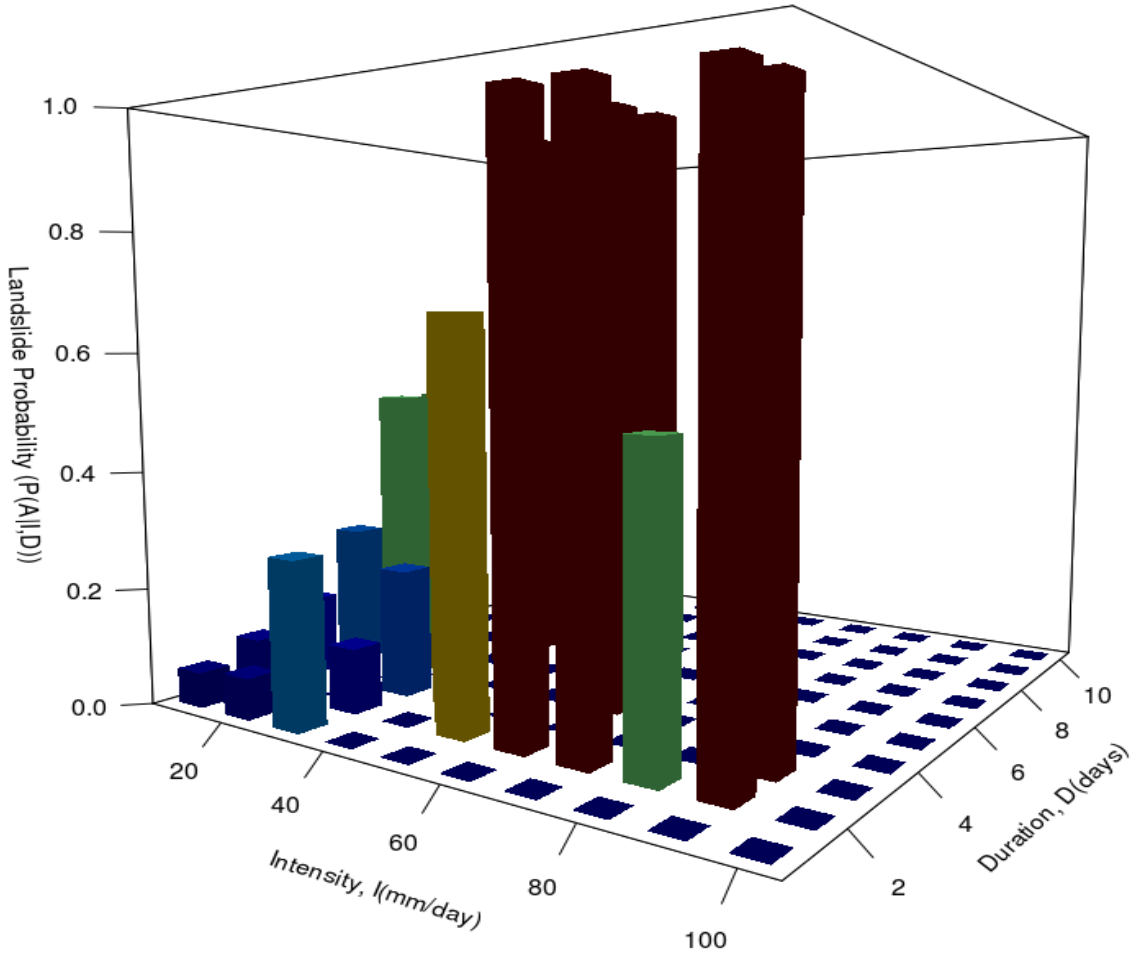


Fig. 7: Histogram of landslide probability as a function of rainfall duration and intensity

Accepted

Accepted Manuscript

Public Interest statement:

The increase of landslide incidences in the Himalayan region has led to the development of various kinds of study with an aim to set up an early warning system. The aim of the present work is to understand the effect of rainfall parameter to landslide triggering in terms of probabilities. This would help in determining the most important rainfall parameter and thereby devise strategies with an aim to save human lives. Such a study has rarely been conducted for Indian Himalayan region and hope it would encourage further research on various other regions.

Author statement:

Dr Raju Sarkar is a professor at the College of Science and Technology, Royal University of Bhutan with interests in Geotechnical aspects of Landslide and Earthquake Hazards, Rainfall Induced Landslides, Stabilization Methods for Landslide Mitigation, Vulnerability and Risk Assessment, Disaster Management Education. He heads the University's Disaster Risk Reduction and Community Development Studies Centre. In Bhutan, currently he works on rainfall induced landslides along the Phuentsholing-Thimbu Highway.

Accepted Manuscript

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/328265391

qualitative-and-quantitative-phytochemical-analysis-of-emcentella-asiaticaem-2329-6836-1000323

Article · May 2018

CITATIONS
0

READS
15

1 author:

Arpita Roy
Delhi Technological University
46 PUBLICATIONS **39** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Micropropagation of Centella asiatica View project

Medicinal Plants Research View project

All content following this page was uploaded by Arpita Roy on 13 October 2018.

The user has requested enhancement of the downloaded file.

Natural Products Chemistry & Research

ISSN: 2329-6836

Arpita Roy et al., Nat Prod Chem Res 2018, 6:4
DOI: 10.4172/2329-6836.1000323

Research Article Open Access

Qualitative and Quantitative Phytochemical Analysis of *Centella asiatica*

Arpita Roy, M Laxmi Krishnan and Navneeta Bharadvaja*

Department of Biotechnology, Delhi Technological University, New Delhi-42, India

***Corresponding author:** Navneeta Bharadvaja, Department of Biotechnology, Delhi Technological University, New Delhi-42, India, Tel: 0112787 1018; E-mail: navneetab@dce.ac.in

Received: May 01, 2018; **Accepted:** May 07, 2018; **Published:** May 10, 2018

Copyright: © 2018 Arpita Roy M, et al. 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

Centella asiatica commonly known by names like gotu kola, Mandukaparni or Indian pennywort is a small herbaceous perennial medicinal plant belonging to *Apiaceae* family. Its active metabolites are triterpenoids which includes asiaticoside, asiatic acid, madecassoside, madecassic acid which has wide range of pharmaceutical activities like wound healing, memory enhancer, treatment of skin diseases etc. The present work is aimed to perform phytochemical analysis and estimation of antioxidant activity by DPPH assay. The antioxidant activity by DPPH assay showed maximum IC_{50} value in CSC extract than shoot and callus extracts. The quantitative estimation of total phenolic, flavanoids and tannins indicates highest phytochemicals in SC when compared to CC and CSC.

Keywords: *Centella asiatica*; Shoot culture; DPPH; Total phenolic content; Total flavanoid contents; Total tannins contents

Introduction

Centella asiatica, a small edible, herbaceous medicinal plant belonging to *Apiaceae* family native to India, Sri Lanka, Iran, New Guinea, Australia, Indonesia southern and central Africa [1]. This plant is listed as an important drug in the Indian Herbal Pharmacopoeia, European Pharmacopoeia, Pharmacopoeia of the People's Republic of China and German Homoeopathic Pharmacopoeia [2]. In 2006, it was ranked third position in a priority list of most essential Indian medicinal plants based on their pharmaceutical and economic importance and also the demands for its raw material is constantly rising throughout the world [3]. *Centella* contains several valuable secondary compounds or terpenoids, known as centelloids, which includes asiaticoside, madecassoside, centelloside, centellose, brahminoside, thankunizide, sceffoleoside, brahmoside, and asiatic, centellic, brahmic, and madecassic acids [4-6]. The most biologically active compounds include triterpene saponin like asiatic acid, asiaticoside, madecassoside and madecassic acid, which are responsible for a wide range of therapeutic activity [7-9].

Antioxidant is any substance which prevents or retards deterioration richly available in fruits and green leafy vegetables which is known to reduce the risk of cancer, cardiovascular diseases. Some antioxidant components are ascorbic acid, carotenoids, vitamin E, polyphenols and other phytochemicals. A free radical is substance with unpaired electrons causes damage to lipids, protein, enzyme thus antioxidants prevents these damages by neutralizing them. Phytochemical imparts health benefits beyond basic nutrition [10]. Several plant extracts and different class of phytochemicals are known to have antioxidant activity [11]. Flavonoids are potent antioxidant and its capacity depends on their molecular structure and Quercetin is one the most abundant dietary flavonol usually used as standards. It was also reported that plant contains tannins, sugars, inorganic acids [12], amino acids like phenylalanine, aspartic acid, α alanine, glutamic acid and glycine [13]. The antioxidant activity of *Centella* plays a key role in acting against the reactive oxygen species in our body [14]. Due to health benefits like antioxidant activity, the usage of *Centella* in food and beverages has been increased over these years. In the present study qualitative phytochemical analysis of *Centella asiatica* was done in three different culture system i.e., shoot, callus and cell suspension culture. Further antioxidant activity, total phenolic, total flavonoid and total tannin content were estimated.

Materials and Methods

Shoot culture

Air dried plant material (0.5 g) was weighed and crushed to fine powder and then add 5 ml of 80% methanol was added and mixed properly. Sonicate for 15 minutes and left at room temperature for 24 hrs. The extract was centrifuged for 10 minutes at 1500 g and the supernatant was collected and stored at 4°C until further use.

Callus culture

For callus extracts, approximate 0.5 g of air dried callus was soaked in 5 ml of 80% methanol and mixed properly, then sonicate for 15 minutes and centrifuge for 10 minutes at 1500 g and the supernatant was collected and stored at 4°C until further use.

Cell suspension culture

For Cell suspension cells extract, the cell suspension was transferred into sterile falcon tubes and centrifuged at 1500 g for 15 minutes and then the supernatant was removed so that cells remains at the bottom as pellet. Approximate 0.5 g of cells was weighed and adds 5 ml of 80% methanol mix it properly. Then sonicate the samples for 10 minutes. Again, centrifuged the samples at 1500 g for 10 minutes and the supernatant was collected and stored at 4°C until further use.

Preliminary screening of phytochemicals

a) Test for Terpenoids (Salkowski test): About 5 ml of sample was mixed with 2 ml of chloroform in a test tube to which 3 ml of concentrated sulphuric acid was carefully added through the sides to

form a layer. If reddish brown colour appears at the interface, indicates presence of terpenoids.

b) Test for Saponin: In a test tube, few amount of sample was taken and mixed properly with 5 ml of distilled water and was vigorously shaken, if stable foam appears, then it indicates presence of saponin.

c) Test for Tannins: In a test tube, about 1 ml of the sample was taken and then add 1 ml of Potassium ferricyanide (0.008 M) then 1 ml of Ferric chloride (0.02 M) containing 0.1 N HCl was added. Blue-black coloration indicates the presence of tannin.

d) Test for Alkaloids: In a test tube few quantity of sample is mixed with 2 ml of Wagner's reagent (2 g of iodine and 6 g of potassium iodide in 100 ml distilled water). If reddish brown colored precipitate is formed indicates the presence of alkaloids.

e) Test for Flavonoids: In a test tube, about 5 ml of dilute ammonia solution was added to small amount of the sample followed by addition of concentrated sulphuric acid. A yellow colour indicates the presence of flavonoids which disappears after some time.

f) Test for Steroids: In a test tube, 1 ml of the sample was taken and dissolved in 5 ml chloroform, then equal volume (5 ml) of concentrated sulphuric acid was carefully added through the sides of test tube. If upper layer turns into red colour and the sulphuric acid layer turns yellow colour with slight green fluorescence which indicates the presence of steroid.

Antioxidant activity by DPPH free radical assay

Sample stock solutions of 1 mg/ml concentration prepared which were diluted with methanol to final concentrations of 5, 10, 25, 50, 125, 250, µg/mL to make a total volume of 2.5 ml. 1 mL of a 0.3 mM DPPH prepared methanol solution was added to different concentrations of sample solutions and allowed to react at room temperature. Blank was methanol (1.0 mL) and plant extract solution (2.5 mL). The negative control was DPPH solution (1.0 mL) and methanol (2.5 mL). The positive controls are the standard solutions which is ascorbic acid for which all the above procedure was repeated. After 30 min the absorbance values of standard and samples were measured at 518 nm and converted into the antioxidant activity in percentage using the following formula:

Antioxidant activity (%)=100-{[(Absorbance of sample-Absorbance of blank})*100]/Absorbance of control}

The I_0 value was calculated from the plots where the X axis represents the concentration of sample/standard and the Y axis is the percentage of antioxidant activity.

Experiment was performed in triplicates

Flavanoid estimation: The Aluminium chloride colorimetric assay was done to estimate flavanoids. Take 1 ml of sample and mix it with 4 ml of distilled water in a test tube. Then 0.3 ml of prepared sodium nitrite (5%) was added. 5 minutes later, 0.3 ml of prepared aluminium chloride (10%) was mixed properly. After 5 minutes, 2 ml of prepared sodium hydroxide (1 M) was added and made upto to 10 ml using distilled water. A standard solution of quercetin of concentration 20, 40, 60, 80 and 100 µg/ml was prepared using the above procedure. The absorbance of the sample and qurecetin standards was determined using an UV or Visible spectrophotometer at 510 nm. The total flavonoid content obtained was expressed in terms of mg of Qurecetin Extract per gram of extract.

Phenol estimation: The Total phenol content was estimated using Folin-Ciocalteu (F-C) reagent method. Firstly F-C reagent (10%) was prepared by 10 ml of F-C reagent added to 90 ml in water. Then sodium carbonate (5%) was prepared by dissolving 3 g of sodium carbonate in 50 ml of distilled water. Then in a test tube take about 200 µL sample to which 1.5 ml of prepared F-C reagent was added and kept at the dark condition for 5 minutes. After which, 1.5 ml of prepared sodium carbonate was added and then mix it properly. All the test tubes were again kept in the dark at dark conditions for about 2 hours.

The standard used was Gallic acid whose calibration curve was plotted. About 3 mg of Gallic acid was taken and dissolve it in 10 ml of methanol to get 300 mg/L concentration. Similarly 200, 100, 50 and 25 mg/L concentrations were prepared. Then the above mentioned procedure was followed for all the standard concentrations. The absorbance of all the samples and standards was determined using UV spectrophotometer at a wavelength of 750 nm.

Tannin estimation: The tannin content was estimated by Folin-Ciocalteu method. About 0.1 ml of the sample extract was treated with 7.5 ml of distilled water, with 0.5 ml of prepared F-C reagent (10%) reagent and adds 1 ml of the prepared sodium carbonate solution (35%) and then dilutes it to 10 ml using distilled water. The entire reaction mixture was mixed well and kept at room temperature for 30 min. The gallic acid was used as the standard and prepared to various concentrations like 20, 40, 60, 80 and 100 µg/ml with methanol. Finally, the absorbance of all the samples and gallic acid standards were determined using an UV or Visible spectrophotometer at 725 nm. The tannin content obtained was expressed as mg of gallic acid equivalent (GAE) per gram of extract

Results and Discussion

Preliminary phytochemical screening

When preliminary phytochemical screening was done in shoot, callus and cell suspension extracts, the positive results were obtained in tannins, flavanoids, terpenoid, saponin and steroids in all the three cultures and absence of alkaloids. Similar report has been obtained by Roy and Bharadvaja [15] in case of *Centella asiatica* shoot culture (Table 1).

Phytochemicals	Plant extract	Callus extract	Cell suspension cells extract
Alkaloids	-	-	-
Tannin	+	+	+
Flavanoids	+	+	+
Terpenoids	+	+	+
Saponins	+	+	+
Steroids	+	+	+
+ for positive result - for negative result			

Table 1: Preliminary Phytochemical screening of methanolic extracts of *Centella asiatica*.

Antioxidant activity

IC_{50} value was calculated from the %inhibition vs. concentration graph which indicates the concentration of sample that is required for scavenging 50% of DPPH free radical. The IC_{50} value was highest of 21.25 µg/ml for cell suspension culture followed by shoot culture 19 µg/ml and then callus extract with 10 µg/ml. The ascorbic acid was taken as the standard with the IC_{50} value of 26.25 µg/ml. It was reported that ultrasonic assisted extraction showed highest recovery and highest antioxidant activity with 79% scavenging activity [16]. The IC_{50} values reported in *C. asiatica* with 100% ethanol extracts is 35.6 ± 1.3 µg/ml, 50% ethanol extract is 7.1 ± 1.5 µg/ml of and water extract is 10.3 ± 1.2 µg/ml [17] (Table 2 and Figure 1).

Samples	IC_{50} value (µg/ml)
Plant extract	19 ± 0.83
Callus extract	10 ± 0.79
Cell suspension sample	21.25 ± 1.2
Ascorbic acid(standard)	26.25 ± 0.64

Table 2: The IC_{50} values calculated from the graphs.

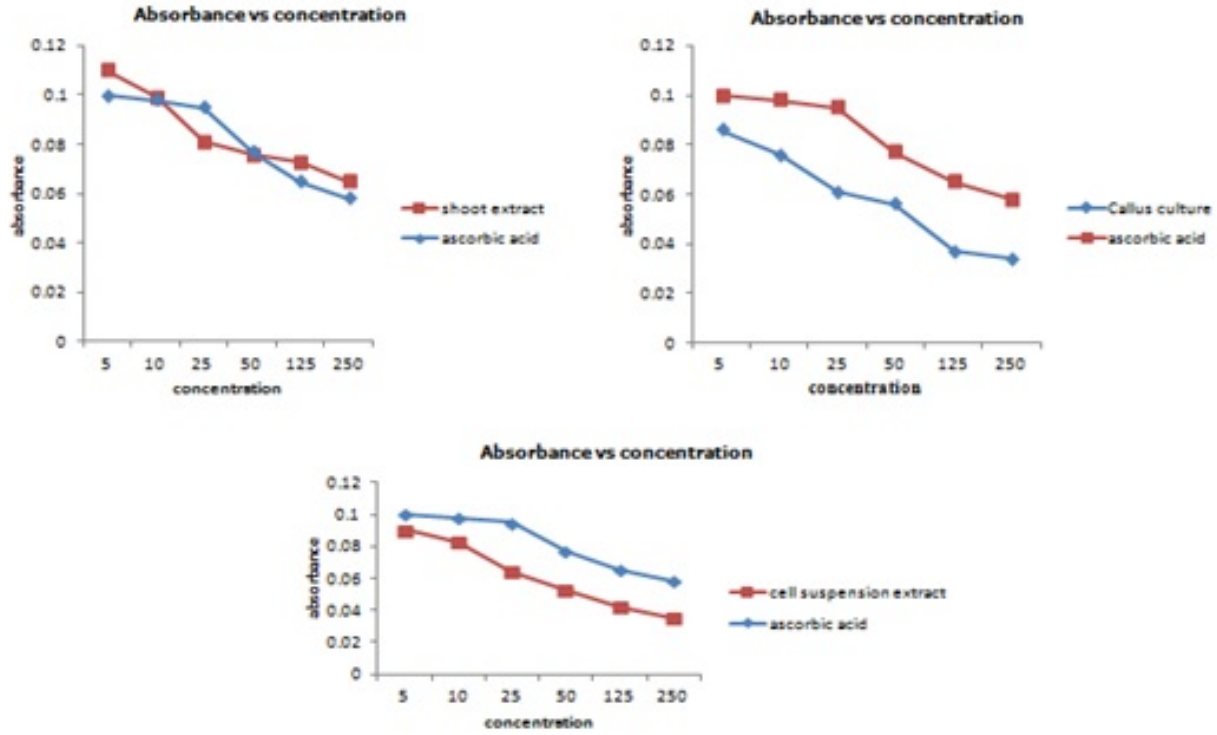


Figure 1: Absorbance vs concentration graph of ascorbic acid and a) shoot extract, b) Callus extract, c) Cell suspension extract.

Total phenolic, flavonoid and tannin estimation

The polyphenol content expressed in terms of mg of gallic acid equivalent per gram of extract was found to be highest in plant extract of 372.5 mg of GAE/g of extract followed by callus extracts having 182.5 mg of GAE/g of extract and least in cell suspension extract, 70 mg of GAE/g of extract. The total flavanoid estimation indicated in terms of qurecetin equivalent per gram of extract with maximum value of 275 ± 0.8 of mg QE/g of extract, following callus 145 ± 0.4 mg QE/g of extract and 100 ± 0.7 QE/g of extract in cell suspension cells. The Total tannin content estimate using gallic acid as standard found to maximum in shoot extract, 260 ± 0.55 mg of GAE/g of extract, followed by callus, 205 ± 0.7155 mg of GAE/g of extract and 55 ± 0.8 mg of GAE/g of extract in cell suspension extracts. According to reports, *Centella* extracts by ultrasonic assisted extraction showed Total Phenolic Content of 1350 mg GAE/100 g Dry Weight and Total Flavonoid Content, 599 mg CE/100 g Dry weight [16]. The polyphenols in 100% ethanol extract was 21.1 ± 0.1 Pyrogallol Equivalent and flavanoid is 9.3 ± 0.3 Quercetin Equivalent [17]. The polyphenols in *Centella* was found to be 150 mg tannic acid/100 g for *C. asiatica* [18]. The total antioxidant capacity was found to be very less in *C. asiatica* (623.78 µmol of ascorbic acid/g of sample). The IC_{50} value was estimated to be 19.89 mg/ml for *Centella asiatica* (Table 3) [19].

View publication stats

S No.	Samples	Total phenolic content (mg/g)*	Total flavanoid estimation (mg/g)	Total tannin content(mg/g)*
1	Shoot extracts	372.5	275 ± 0.8	260 ± 0.55
2	Callus extracts	182.5	145 ± 0.4	205 ± 0.71
3	Cell suspension cells extracts	70	100 ± 0.7	55 ± 0.8
*Expressed in terms of mg of gallic acid equivalent per gram of extract.				

Table 3: Total phenolic, flavanoid and tannin content in Shoot, Callus and Cell suspension extracts.

Conclusion

Qualitative phytochemical analysis showed the presence of flavanoids, tannins, terpenoid, saponin and steroids in all the three culture systems whereas absence of alkaloids. Antioxidant activity was maximum in cell suspension culture i.e. 21.25 µg/ml. Highest amount of phenolic, flavanoid and tannin content was present in the shoot culture.

References

1. Hoang N, Nguyen LT (2010) Asiaticoside production from Centella (Centella Asiatica L. Urban) Cell Culture. Biotechnol Bio Eng 15: 1065-1070.
2. Schaneberg BT, Mikell JR, Bedire E, Khan IA (2002) An improved HPLC method for quantitative determination of six triterpenes in Centella asiatica extracts and commercial products. Pharmazie 58: 381-384.
3. Sivakumar G, Alagumanian S, Rao MV (2006) High Frequency in vitro Multiplication of Centella asiatica: An Important Industrial Medicinal Herb. Eng Life Sci 6: 597-601.
4. Roy A, Kundu K, Saxena G, Kumar L, Bharadvaja N (2016) Effect of different media and growth hormones on shoot multiplication of in vitro grown Centella asiatica accessions. Adv Tech Biol Med 4: 172.
5. Roy A, Bharadvaja N (2017) Effect of different culture medias on shoot multiplication and stigmasterol content in accessions of Centella asiatica. Intern J Ayurv Herb Med 7: 2643-2650.
6. Roy A, Attre T, Bharadvaja N (2017) Anticancer agent from medicinal plants: A review. New aspects in medicinal plants and pharmacognosy. JB Books.
7. Kundu K, Roy A, Saxena G, Kumar L, Bharadvaja N (2016) Effect of Different Carbon Sources and Elicitors on Shoot Multiplication in Accessions of Centella asiatica. Med Aromat Plants 5: 251.
8. Roy A, Kundu K, Saxena G, Bharadvaja N (2017) Estimation of Asiaticoside by Using RP-HPLC and FAME Analysis of Medicinally Important Plant Centella asiatica. J Plant Biochem Physiol 5: 198.
9. Roy A, Bharadvaja N (2017) Medicinal Plants in the Management of Cancer: A Review. Int J Complement Alt Med 9: 291.
10. Oomah BD, Mazza G (2000) Functional foods. In: Francis FJ (ed) The Wiley encyclopedia of science and Technology, 2nd edn. Wiley, New York, pp: 1176-1182.
11. Vani T, Rajani M, Sarkar S, Shishoo CJ (1997) Antioxidant properties of the Ayurvedic formulation Triphala and its constituents. Int J Pharmacognosy 35: 313-317.
12. Kapoor LD (2005) CRC Handbook of Ayurvedic Medicinal Plants, (CRC Press LLC, Florida), pp: 208-209.
13. Malhotra CL, Das PK, Sastry MS, Dhalla NS (1961) Chemical and pharmacological studies on Hydrocotyle asiatica Linn. Indian J Pharm, p: 23.
14. Mustafa RA, Hamid AA, Mohamed S, Bakar FA (2010) Total phenolic compounds, flavonoids, and radical scavenging activity of 21 selected tropical plants. J Food Sci 75: 28-35.
15. Roy A, Bharadvaj N (2017) Qualitative analysis of phytocompounds and synthesis of silver nanoparticles from Centella asiatica. Innov Techn Agri 1: 88-95.
16. Nithyanandam R, Shapheri MR, Nassir M (2014) Antioxidant potential of Malaysian herb Centella asiatica, 3rd international conference on environment. Chem Bio 78: 1-7.
17. Rahman M, Hossain S, Rahaman A, Fatima N, Nahar T, et al. (2013) Antioxidant activity of Centella asiatica (Linn.) Urban: Impact of extraction solvent polarity. J Pharmaco Phytochem 1: 27-32.
18. Gupta S, Gowri BS, Lakshmi AJ, Prakash J (2013) Retention of nutrients in green leafy vegetables on dehydration. J Food Sci Technol 50: 918-925.
19. Shyamala BN, Gupta S, Lakshmi JA, Prakash J (2005) Leafy vegetable extracts-antioxidant activity and effect on storage stability of heated oils. Innov Food Sci Emerg Technol 6: 239-245.

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/TII.2018.2875035, IEEE Transactions on Industrial Informatics

Manuscript No. TII-18-1737.R1

Robust Repetitive Control Design for 3P4W Shunt Active Power Filter

Gitanjali Pandove[1], *Member, IEEE* and Mukhtiar Singh[2], *Member, IEEE*

***Abstract*—This paper presents a discrete repetitive control technique for three phase four wire (3P4W) shunt active power filter (SAPF). Generally, the control design for power electronics devices involves two control loops: slow acting outer voltage loop and fast acting inner current control loop. The reference for inner current loop is periodic in nature and cannot be easily tracked by PI regulator. The repetitive controllers (RC) are well known for their tracking ability of periodic signals and offers high gain at all the frequencies. The high gain in higher frequency range may leads towards instability. Therefore, in proposed work, the regular RC is modified by squaring its sensitivity function. This approach results in low amplitude of sensitivity function while offering deep notches at low to mid frequencies range and smaller notches at higher frequencies. This control approach has been simulated and implemented on 3P4W SAPF.**

***Index Terms*—Repetitive Control, Active Power filter, harmonic compensation, Power Quality, 3P4W, Load Compensation, Non-linear load, Unbalanced Load**

I. Introduction

THE power system network at distribution level faces many types of challenges, primarily related to load unbalance, reactive power and current harmonics etc. From power system operator's point of view, these types of issues must be addressed at distribution level itself so that generation and transmission could perform smoothly without any disturbance. Conventionally, the electromechanical devices like tap-changing transformer, passive filter, synchronous condenser have been used to sort out the issues related to power quality. However, with the development of sophisticated power electronics technology, lot of custom power devices like dynamic voltage restorer (DVR), active power filters (APF) may be series or shunt type, static compensator (STATCOM), unified power quality controller (UPQC) are widely being used to mitigate power quality problem. Normally, voltage or current regulated voltage source inverter is utilized as series or shunt connected device depending on situation.

The SAPF is more preferable to solve load current related problems, where it can be regulated to supply the load current harmonics, unbalance and reactive current demand so that grid side current always remain balanced set of sinusoidal current at unity power factor (UPF). Since, SAPF is current controlled device, where the inner current control loops are much faster than outer slow voltage loops, it becomes a typical design problem to develop a fast-responding controller. Moreover, the inner current signals to be handled are periodic in nature and they need to transformed into equivalent DC signals in order to regulate them using simple PI controller [1]-[3]. Thus, the performance of controller depends on the accuracy of transformation which is highly sensitive to variation in frequency and phase angle. A large number of control techniques for inner current loop of SAPF have already been proposed. Majority of them are mainly based on PI, proportional resonant (PR), hysteresis, passivity-based control (*PBC*) and deadbeat (DB) controllers. Here, PI is the simplest one but cannot be used for oscillating signals [4]-[5], hysteresis controller may cause resonance issue due to variable switching frequency [6]-[7], the PR can be tuned only for selective harmonics [8]-[9], PBC [10]-[11]and DB [12] require accurate model of system and control performance may suffer with parameter variations. Some artificial intelligence-based control techniques have also been reported in [13]-[14]. The major issue with such kind of techniques is lack of suitable training data for off line tuning and complex computational burden for online adaptive tuning purpose.

Since, most of power electronics converters deals with periodic signals where the prior information about input signal is available to controller, it is advantageous to use repetitive controllers (RC) [15]-[18]. Internal model principle theory (IMPT) [19] forms the basis for the design of RC where it can be used to achieve zero steady-state error while tracking multiple frequencies harmonic components. Here, high gains are induced at those harmonic frequencies which are needed to be tracked. Therefore, RC proves to be very effective in mitigating the harmonics with the proper control of SAPF. The traditional digital RC may be designed by introducing n no. of delays. Here n must be taken as the integer no. of samples in fundamental cycle of input signal. Designing a robust RC is vary tedious task as any numerical error or inaccurate modelling may leads the controller to unstable region [20]. Moreover, the control performance of simple RC may suffer with variation in signal frequency (In this case, it is grid voltage frequency). Since, the RC is tuned to reject the unwanted harmonics which are integer multiple of fundamental frequency, it may lead to false operation of SAPF in case of frequency variations. To sort out this issue, the variable sampling rate approach has been proposed for RC design for proper harmonics rejection [21]-[22]. A similar method where fractional delay approach is proposed using Lagrange-interpolation- based FIR filter to model the fractional part of n [23]-[24]. However, the variable-sampling-rate and fractional

The manuscript received on July 04, 2018, revised on Sept. 10, 2018 and finally got accepted for publication on Oct. 01, 2018.

[1] Department of ECE, D. C. R. University of Science and Technology, Murthal 131039, Haryana, India (pandove_geet@yahoo.co.in).

[2] Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India (smukhtiar_79@yahoo.co.in).

1551-3203 (c) 2018 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.

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/TII.2018.2875035, IEEE Transactions on Industrial Informatics

delay methods implies lot of changes in system dynamics and involves lot of mathematical computations, respectively [25]. A similar approach based on cycling sampling strategy has been reported in [26], where the sampling points are shifted in every cycle to have fractional order phase lead compensation with enhanced stability. However, shifting in sampling points may compromise with its steady state error tracking capability. Some of the authors also presented parallel structure repetitive control (PSRC) which employs multiple order harmonic internal models suitable for PWM inverter/converter control [27]-[28]. The major issue with PSRC is to set the gains for multiple parallel loops. Further, the individual harmonic frequency components are need to be tuned for faster convergence which is very difficult to achieve in any practical application.

Therefore, the main objective of proposed work is to simplify the RC design approach with enhanced control efforts. For this purpose, a simple approach as reported in [29] has been utilized, where the sensitivity function has been modified and taken as square of the actual sensitivity function. This approach results in increased deep notches with wider control spectrum for multiple fundamental frequencies without any need of extensive computational burden. The wider notch spectrum implies that the small variation in grid frequency has no significant impact on control performance. Moreover, the low amplitude of sensitivity function while offering deep notches at all the frequencies integer multiple of fundamental frequency ensures the knockout of unwanted multiple frequency components without any phase lag. This approach is found to be very much suitable for the control of APF whose main task is also to eliminate the unwanted frequencies in order to have reduced THD level. Therefore, the main contribution implies in adapting such robust controller for APF, where it is designed, simulated and successfully validated through experimental results.

The proposed control approach is designed and developed in MATLAB Simulink and tested on four-leg SAPF using *simpowersystem* blockset. The simulation results have been provided with highly unbalanced non-linear load, where the SAPF is able to supply the unbalance non-linear load demand while compensating reactive load demand. This makes the grid side currents as balanced set of sinusoidal currents. The extensive experimental analysis has been performed to validate the claims of proposed control method.

This paper is structured in five main sections, where overall system description and its control description has been presented in Section-II. The section-III presents the modelling of proposed RC while the simulation results with experimental validation have been presented in section-IV. The conclusion of whole paper is summarized in Section-V.

II. System Modelling and Control

Fig. 1 shows the complete schematic diagram of proposed system. Here, IGBT based 4-leg power electronics converter is controlled as shunt APF, where the switching period of individual IGBT is manipulated in a power frequency cycle such that the combined effect of unbalanced nonlinear load and SAPF injected current appears to be a balanced resistive load while seen from grid side. This results into perfect set of balanced resistive load currents to be drawn from grid and thus, automatically enables the unity power factor grid operation.

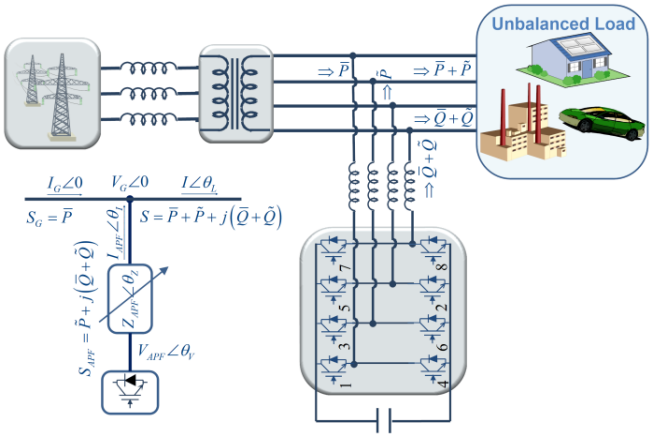


Fig. 1. Schematic of 3P4W Shunt Active Power Filter

Here, it is worthwhile to mention that the 4th leg of SAPF is utilized to supply the neutral load current demand (unbalance current component) so that the grid side neutral current could be limited to zero. This also indicates the balanced operation of grid. Let us consider a non-linear unbalance load drawing a complex power:

$$S = P + jQ \;= \bar{P} + \tilde{P} + j\left(\bar{Q} + \tilde{Q}\right) \tag{1}$$

Now, to have UPF grid operation, the grid voltage and grid current must be in same phase with sinusoidal wave shape (i.e. $V_G\angle 0, I_G\angle 0$). This can be achieved only if the fundamental active power ($\bar{P}$) is to be supplied from grid while the SAPF supplies the harmonic active power demand ($\tilde{P}$), fundamental reactive power demand ($\bar{Q}$) as well as harmonic reactive power demand ($\tilde{Q}$) of load. The complex power to be supplied from SAPF side can be expressed as:

$$S_{APF} = V_G I_{APF}^{*} \tag{2}$$

Here, I_{APF} is the SAPF injected current expressed as

$$I_{APF}^{*} = \frac{V_{APF}\angle\theta_V - V_G}{Z_{APF}\angle\theta_Z} \tag{3}$$

$$V_{APF} = V_G + I_{APF} Z_{APF}\angle\theta_I + \theta_Z$$

Where V_{APF} is SAPF terminal voltage with phase angle θ_V and Z_{APF} is series impedance with phase angle θ_Z. On substituting equation (3) in (2), we have:

$$S_{APF} = V_G\left(\frac{V_{APF}\angle-\theta_V - V_G}{Z_{APF}\angle-\theta_Z}\right) = \frac{V_G V_{APF}\angle\theta_Z - \theta_V}{Z_{APF}} - \frac{V_G^2\angle\theta_Z}{Z_{APF}}$$

$$= \left(\frac{V_G V_{APF}}{Z_{APF}}\cos\left(\theta_Z - \theta_V\right) - \frac{V_G^2}{Z_{APF}}\cos\left(\theta_Z\right)\right) \approx \tilde{P} \tag{4}$$

$$+ j\left(\frac{V_G V_{APF}}{Z_{APF}}\sin\left(\theta_Z - \theta_V\right) - \frac{V_G^2}{Z_{APF}}\sin\left(\theta_Z\right)\right) \approx Q$$

1551-3203 (c) 2018 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.

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/TII.2018.2875035, IEEE Transactions on Industrial Informatics

From equation 4, it is clear that the SAPF injected currents can be manipulated by controlling its terminal voltage, phase angle and the impedance. This particular objective can be achieved by switching the power electronics switches in such a way that the SAPF's impedance and load impedance becomes purely resistive load to be supplied from grid. Therefore, the switching sequence of various switches in SAPF forms the operating principle of load compensation.

The operation of SAPF can be controlled using outer voltage and inner current control loops. The current control loops have very fast dynamics in comparison to slow acting voltage control loop. The main aim of voltage control loop is to regulate the dc-link capacitor voltage which indirectly sets the reference for active power required by APF in order to meet out its switching losses. The *PI* regulator is used to track the error between actual (V_{dc}) and reference (V_{dc}^*) dc-link voltages. The output of *PI* regulator sets the reference of active current component (i_d^*). Since, the APF is controlled in indirect control mode, the reactive current component (i_q^*) is set to zero. The grid synchronizing angle (θ) obtained from phase lock loop (PLL) is utilized to determine the grid side reference currents (i_a^*, i_b^*, and i_c^*). In order to have grid side balanced operation, the neutral reference current (i_n^*) is set to zero. The actual grid side currents are compared with the predetermined reference currents, where the proposed repetitive controller tracks the error currents to zero value. The complete schematic with control description is shown in Fig. 2.

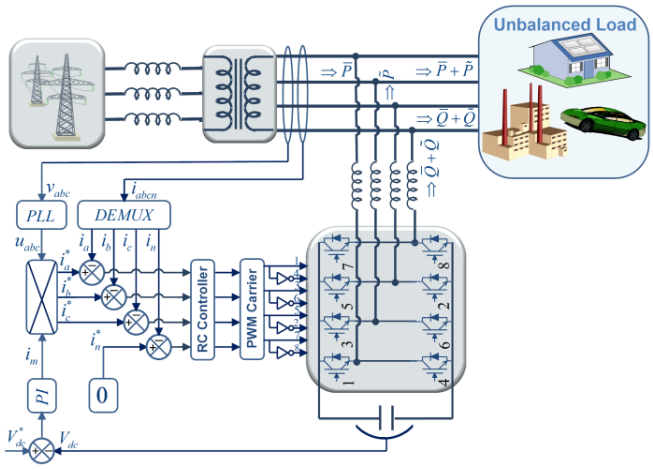


Fig. 2. Control Description of 3P4W Shunt Active Power Filter.

Control Description

The control of grid connected power electronic converters involves *abc-dq* transformation commonly known as Park's transformation which is based on rotating-reference frame theory. The main advantage of this particular transformation is that it results in quasi-DC signals which are easy to track. The development of control system for a simple resistive-inductive load is very simple and can be easily controlled using a low bandwidth proportional-integral (PI) controller. The conversion of signals from rotating to stationary reference frame or vice-versa is performed using Clarke's and Park's transformation or their inverse. However, the issue arises in case of system is single-phase or there is unbalance loading without an isolated neutral in case of three-phase. In such cases, the synchronous coordinate transformation cannot be applied directly or in other words the advantage of having quasi-DC current components to be regulated is completely lost. Therefore, there may be certain conditions where the controller design with wider control bandwidth is required to track sinusoidal reference currents.

Let us consider a open loop system having a power converter connected to load characterized by transfer function $G_p(s)$ and a control unit characterized by a transfer function $G_c(s)$. The combined open-loop transfer function, $G_o(s)$ may be written as:

$$G_o(s) = G_c(s).G_p(s) = \frac{A(s)}{B(s)} \tag{5}$$

Let a sinusoidal reference signal applied to this combination:

$$f(t) = I_m \sin \omega t \Rightarrow F(s) = \frac{I_m s}{s^2 - \omega^2} \tag{6}$$

Then error feedback signal can be expressed as:

$$E(s) = \frac{F(s)}{1 + G_o(s)} = \frac{B(s)}{B(s) + A(s)} \frac{I_m s}{s^2 + \omega^2} \tag{7}$$

On applying partial fraction expansion, we have

$$E(s) = \frac{a_1}{s_1 + \omega_1} + \cdots + \frac{a_i}{s_i + \omega_i} + \frac{b_1}{s_o - j\omega_o} + \frac{b_2}{s_o + j\omega_o} \tag{8}$$

The poles of aforementioned function must fall on left hand side of the s-plane to have stable bounded output. However, if any of the poles lies on the right-hand side of the s-plane, the steady state error may increase to an unbounded limit and ultimately lead towards instability. The other possibility of poles being complex conjugate at known frequencies, then the poles corresponding to these frequencies will disappear from the error-transfer function which further results in the reduction of the steady-state error to zero. Therefore, the transfer function should include the terms corresponding to these sinusoidal frequencies while designing the controller. However, designing a control for signal having other multiple frequency components shall be very cumbersome as additional terms very specific to individual frequency components need to be included while designing a controller. Fig. 3. shows a simple example for designing a closed-loop current control having multiple frequencies.

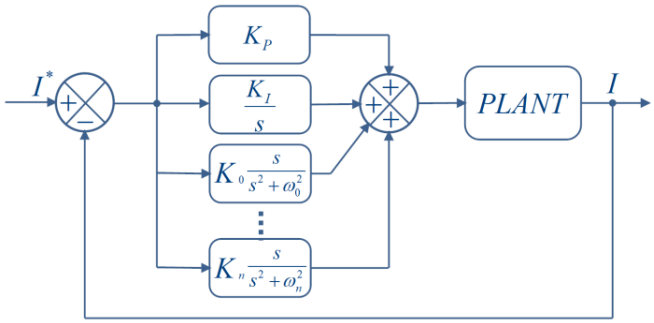


Fig. 3. Closed loop control of Typical Power Electronics System.

For active power filter (APF) application, the main

1551-3203 (c) 2018 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.

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/TII.2018.2875035, IEEE Transactions on Industrial Informatics

objectives are to have sinusoidal balanced grid currents in phase with their respective phase voltages while maintaining constant DC-link voltage. Keeping the DC-link voltage constant is a simple regulation issue where the reference is well defined both in shape and amplitude. However, in case of inner current control loops, it becomes a typical tracking issue as only the shape of signal and not the actual signal is defined. Therefore, the inner current controller is required to track the current signal of fundamental frequency where the shape but not the amplitude is predefined. Since, the signals in inner current control loops are periodic signals, it is advantageous to use repetitive controller instead of control structure with PI having multiple loops corresponding to different multiple frequency components as shown in Fig.3.

III. Repetitive Controller Design

The repetitive controllers are mainly used to solve those control problems where the disturbances or signals to be controlled are periodic with defined time intervals. According to IMPT, a generator with periodic signal must be included in controller feedback loop to track the periodic reference signal. The periodic signal generator uses large time delay terms corresponding to periods of reference signal. In actual practice, a zero-phase FIR low pass filter $H(z)$ is used to reduce the gain of those frequencies where the behavior of system is not well defined. Here, it is pertinent to mention that this FIR low pass filter decreases the gain of controller to finite values for all the frequencies and thus takes the shape of typical lowpass filter $q(z, z^{-1})$. A typical repetitive controller with zero-phase error low pass filter is shown in Fig. 4.

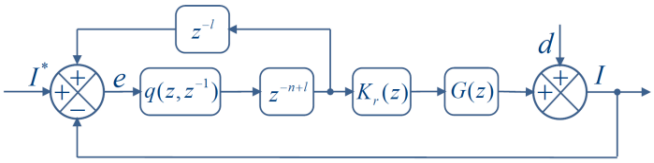


Fig. 4. Structure of Typical Repetitive Controller.

Here, $K_r(z)$ is internally stabilizing controller which is normally taken as inverse of plant model $G(z)$ and may also be calculated as proposed in [15], [17]. The generator of periodic signal comprises lowpass filter (LPF) $q(z, z^{-1})$, delays z^{-l} and z^{-n+l}. Here n represents the time interval of periodic signal and l is the combined delay of both plant and controller. The sensitivity function of the such RC may be written as:

$$\begin{aligned} S_o &= \frac{1-q(z,z^{-1})z^{-n}}{1-q(z,z^{-1})z^{-n}+q(z,z^{-1})z^{-n+l}K_r(z)G(z)} \\ &\approx 1-q(z,z^{-1})z^{-n} \\ &if\ z^{+l}K_r(z)G(z)\approx 1, n \gg l \end{aligned} \tag{9}$$

On substituting $z=e^{j\omega T}$ where T is sampling time and keeping the LPF $q(z, z^{-1})$ amplitude to unity, the sensitivity function S_0 reduces to zero at integer multiple of $(1/nT)$ Hz and doubles at fundamental frequencies. In typical repetitive control design, the disturbance d and reference signal I^* are supposed to be periodic. Therefore, as per the sampling theorem the response of these signal appears at fundamental frequencies. The signal error e may be written as $e=S_0I^*$, where the function S_0 as mentioned in equation 9 knocks out the various frequency components of reference signal I^* by offering deep notches at the fundamental frequencies. This results into decrease in error e for a given periodic reference signal I^*. However, to have better performance, it is always desirable to have sensitivity function S_0 of smaller amplitude at the fundamental frequencies [29].

Here, the main objective of this particular controller is to achieve smaller magnitude of sensitivity function at the fundamental frequencies. For this purpose, the function S_0 as in equation 9 is squared by modifying the structure of filter $q(z,z^{-1})$ as shown in Fig. 5 and the corresponding new sensitivity function S_{new} may be written as

$$\begin{aligned} S_{new} &= S_0^2 \approx [1-q(z,z^{-1})z^{-n}]^2 \\ &= 1+q(z,z^{-1})z^{-2n}-2q(z,z^{-1})z^{-n} \\ &= 1-\underbrace{q(z,z^{-1})[2-q(z,z^{-1})z^{-n}]}_{Modified\ q(z,z^{-1})z^{-n}\ filter\ structure}z^{-n} \end{aligned} \tag{10}$$

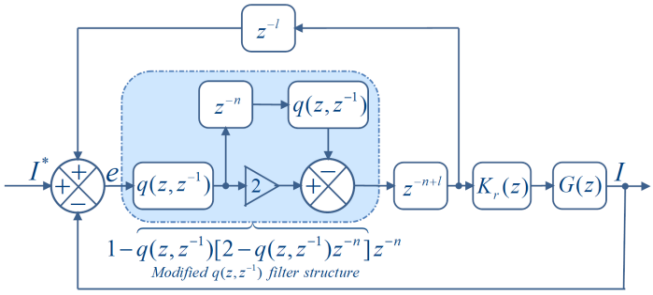
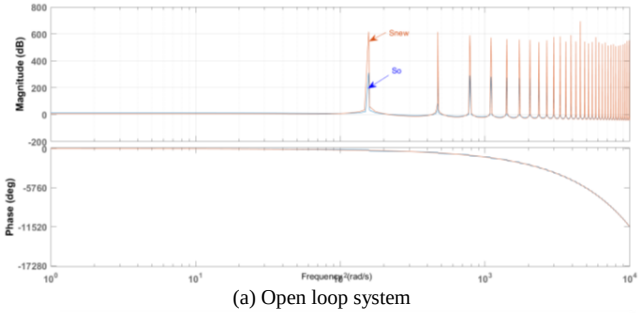
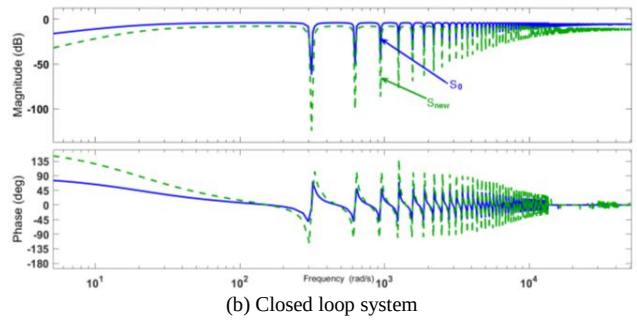


Fig. 5. Structure of Modified Repetitive Controller.

On comparing equation 9 and 10, it can be observed that on squaring the function S_0 as in equation 9, the resultant modified filter structure's amplitude changes quite significantly. This fact can also be verified by analyzing their corresponding bode plot as shown in Fig. 6, where Fig. 6(a) shows the Bode diagram for open loop system and Fig. 6(b) shows the Bode diagram for closed loop system.



(a) Open loop system



(b) Closed loop system

Fig. 6. Bode Plot of Simple and Modified Repetitive Controller.

1551-3203 (c) 2018 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.

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/TII.2018.2875035, IEEE Transactions on Industrial Informatics

From Bode diagram as in Fig. 6, it can be easily seen that the modified RC has better control bandwidth in comparison to simple RC. The modified RC offers deep notches with wide spectrum at all the multiple frequencies without any phase lag and hence makes it suitable for small variations in operating power frequency. Moreover, the gains of closed loop system start reducing at higher frequencies and almost becomes zero. This is possible due to the presence of low pass filter $q(z, z^{-1})$, which attenuates the controller response at higher frequencies and enhances the system stability. Here, the structure of LPF must be carefully designed to ensure the stability of the system. For the proposed work, LPF of unity amplitude with zero phase has been opted to keep the poles and zeroes within unity circle, which guarantees the stability of system [30].

IV. Simulation Results & Discussion

The proposed system is simulated using the various modules of MATLAB/Simulink *SimPowerSystem* tool box. A 4-leg inverter is not readily available in Simulink library. Therefore, the 8 no. of IGBT's have been taken to form a current controlled voltage source converter, where it is connected in parallel with load to act as SAPF. The driver circuit for individual IGBT is also developed in Simulink itself. The individual switches of SAPF are controlled to obtain the 3-phase balanced set of fundamental grid currents at unity power factor with their respective grid phase voltages despite of highly unbalanced non-linear load connected at point of common coupling (PCC) under varying load conditions. An unbalance 3-phase 4-wire non-linear variable load is simulated using both single phase as well as 3-phase load with rectifier, resistive and inductive load profile. The simulation results comprising the waveforms of grid side voltage (V_G), grid side current (I_G), unbalanced load side current (I_L) and injected SAPF currents (I_{APF}) are shown in Fig.7 and Fig. 8.

Here, it is worthwhile to mention that the simulation study has been carried out with two different controllers on same system with same parameters for the comparison purpose. The Fig. 7 shows the simulation results with the RC controller as used in our previous work [15], [17]. The Fig. 8 shows the simulation results with proposed RC controller. As per color coding, the red color indicates phase A, green-phase B, blue-phase C and Violet-Neutral N. On carefully analyzing the load current waveforms, it can be observed that initially, a three-phase balanced nonlinear rectifier load is connected in the system. At t = 0.6 sec, the non-linear load is suddenly increased by 75% of its initial loading and again at t = 0.8 sec., the load is reduced to its initial value. At t = 1.0 sec, a RL load is connected to the phase A of supply. This introduction of RL load, suddenly increases the phase A current by almost 100% of its initial value while the load on phase B and phase C remains same. This highly unbalanced non-linear current results in flow of lot of neutral current.

On analyzing the waveforms of Fig. 7 (d), it can be seen that initially the SAPF injected currents are kept zero. This indicates the non-compensation mode and thus grid current

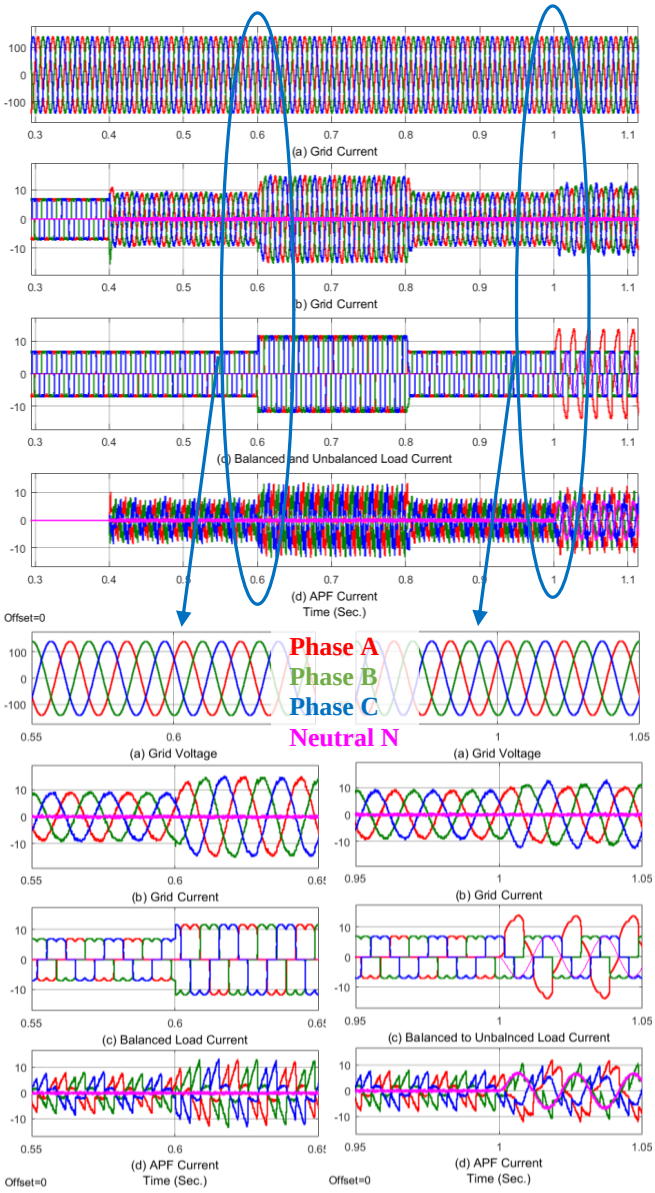


Fig. 7. Simulation results with RC: (a) Grid side voltages, (b) Grid side currents, (c) Unbalanced load side currents, (d) APF currents.

profile and load current profile remain same. At t = 0.4 sec, just after the introduction of SAPF, the grid currents almost become sinusoidal as the SAPF starts injecting the non-linear harmonic load component. However, the harmonic compensation capability of SAPF slightly got affected with sudden increase in load current at t = 0.6 sec. This fact can also be verified on closely observing the grid current profile in the zoomed section of Fig. 7 (b). Moreover, the SAPF performance further deteriorate with introduction of unbalanced load at t = 1.0 sec., where it tries its best to make the grid currents a balanced set of sinusoidal current but in vain. Therefore, the RC as discussed in [15], [17] is able to handle only balanced non-linear load currents only, but unable to handle the wide variations in load current, especially the unbalanced load current, where its effort is promising but not up to the mark.

The Fig. 8 shows the simulation results of proposed RC

1551-3203 (c) 2018 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.

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/TII.2018.2875035, IEEE Transactions on Industrial Informatics

controller. Here, it can be observed that just after the introduction of SAPF at t = 0.4 sec, it starts supplying the reactive and nonlinear load demand so that the grid currents become balanced set of purely sinusoidal currents. This claim can be easily verified from the zoomed part of results in Fig. 8 (b), where it not only handles the steady state current easily but also able to deal with any change in dynamic load conditions smoothly. Moreover, the proposed RC also exhibit the same performance while handling the unbalanced non-linear currents. Just after the introduction of unbalance load at t = 1.0 sec, the proposed RC is able to control the SAPF such that it starts injecting the non-linear as well as unbalance load current demand so that the grid currents always remain balanced with sinusoidal profile. This fact can also be verified by the zero value of grid side neutral current while the SAPF neutral current is true replica of load neutral current. This indicates that the SAPF supplies the neutral current demand of load and this automatically makes the grid side currents perfectly balanced.

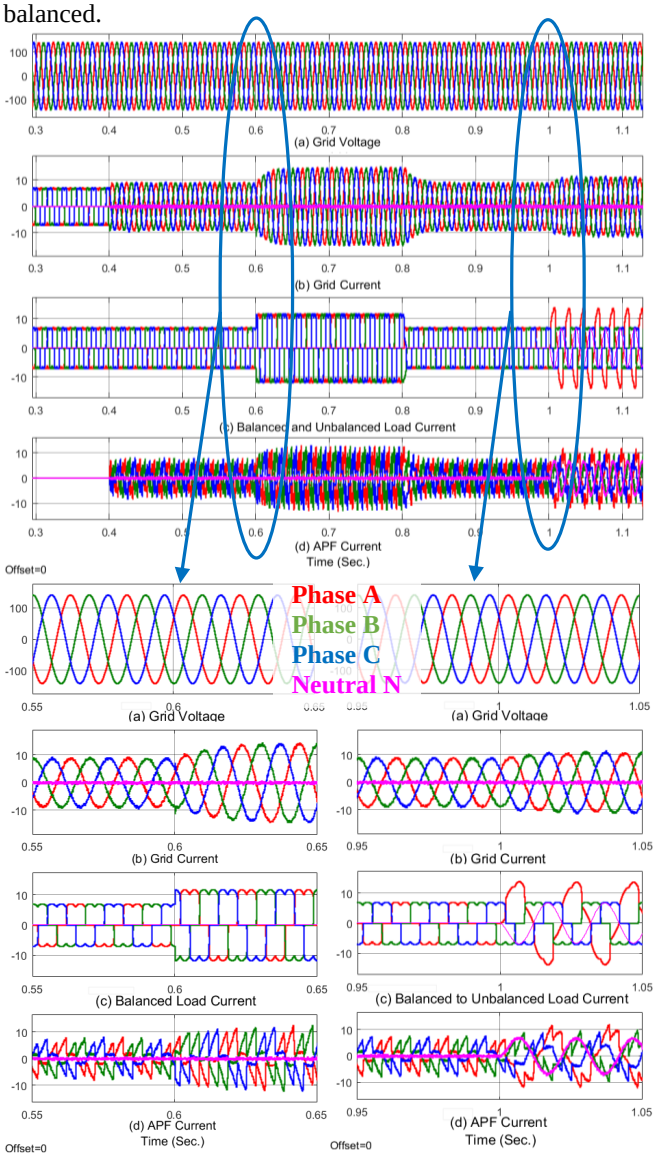


Fig. 8. Simulation results with Modified RC : (a) Grid side voltages, (b) Grid side currents, (c) Unbalanced load side currents, (d) APF currents.

V. Experimental Results & Discussion

A hardware prototype of SAPF is also designed to validate this particular control approach. For this purpose, IGBT based 4-leg converter is used, where the fourth leg is used to compensate the load neutral current. Three phase diode bridge rectifier with unbalanced RL load is utilized to have non-linear unbalanced loading. The experimental study has been carried out in real time using *dSPACE 1104*. Here, both the RC as discussed in [15] and proposed robust RC have been implemented and results are shown in Fig. 9.

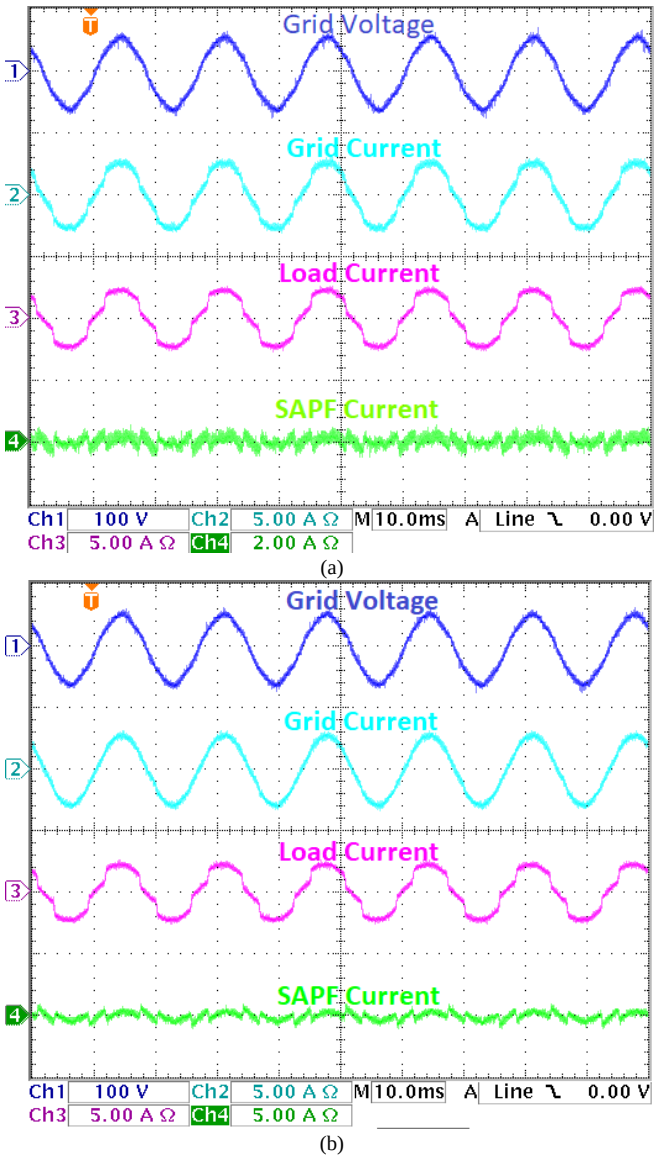


Fig. 9. Experimental Results: (a) with Repetitive Controller (b) with proposed Robust Repetitive Controller

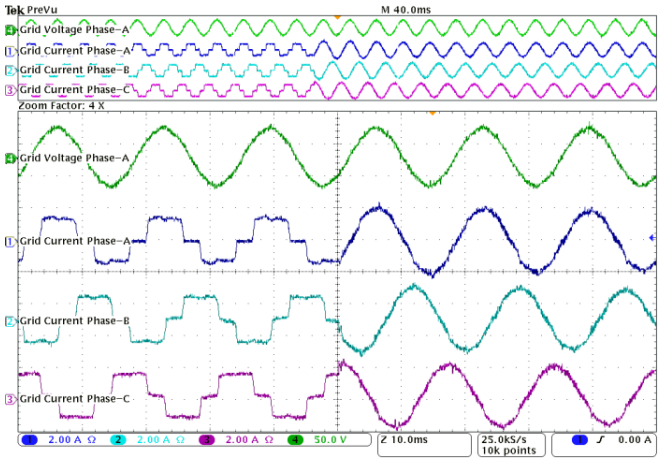
On analyzing the experimental results as shown in Fig. 9 (a), it may be observed that the RC controller is able to compensate the load harmonics to some extent. However, due to higher computational time (140 µsec.), there is some time delay in SAPF injected current, due to which the grid currents are not purely sinusoidal. On the other hand, with the proposed robust RC, the computational time is reduced drastically to

1551-3203 (c) 2018 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.

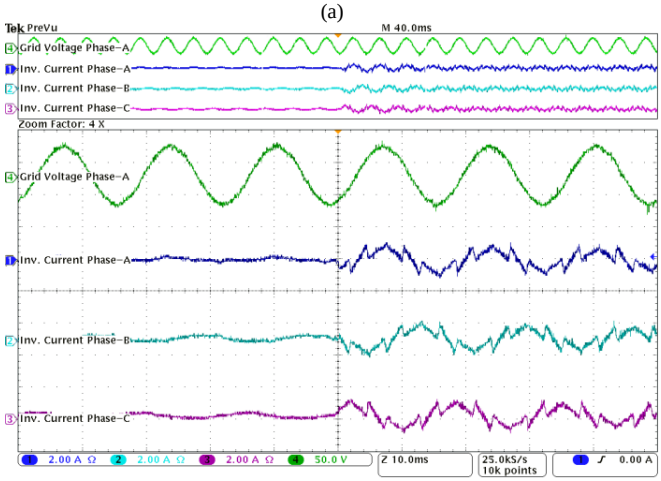
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2875035, IEEE Transactions on Industrial Informatics

100 µsec. This results in accurate SAPF current injection, which makes the grid currents perfectly sinusoidal as evident from Fig. 9(b).

Finally, the experimental results with robust RC are provided under two different operating conditions: 1. Balanced non-linear loading condition, 2. Unbalanced 3P4W non-linear loading conditions.



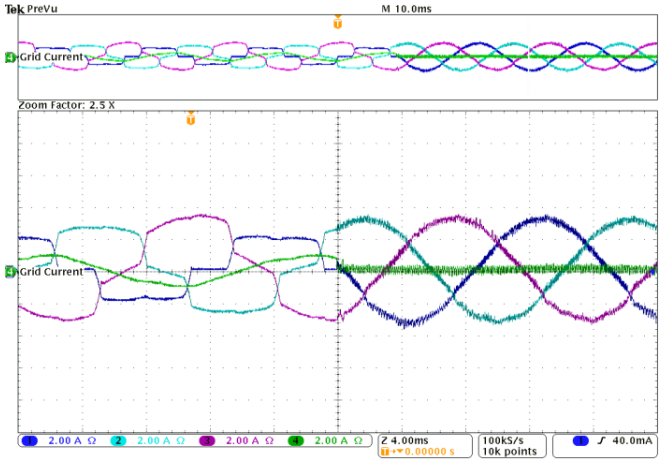
(a)



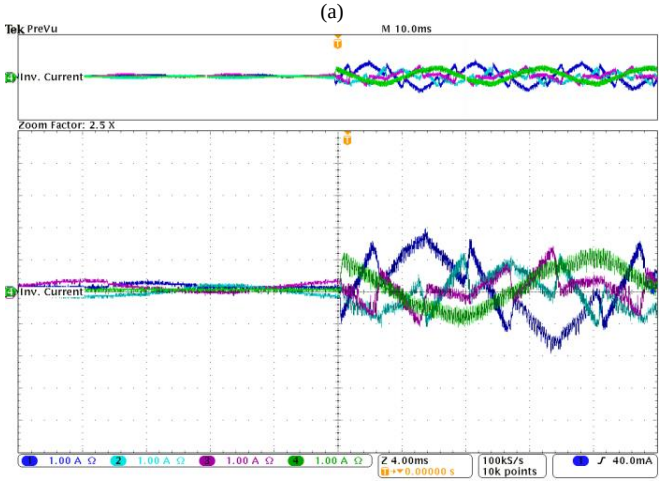
(b)

Fig. 10. Experimental Results: Under Balanced Non-linear load conditions, (a) Grid voltage and Grid side Current without and with compensation (b) SAPF Currents without and with compensation

The Fig. 10, shows the experimental results under balanced non-linear load conditions, where initially the SAPF is not in operation and therefore, whatever load demand is there that is to be supplied from grid side only. In Fig. 10. (a), the waveforms of grid side voltage and grid side currents are shown with and without compensation. Here, it can be observed that initially the grid is supplying the non-linear load demand. But as soon as SAPF is switched on, the grid currents become perfectly sinusoidal and in phase with grid voltages. The corresponding SAPF currents without and with compensation are shown in Fig. 10. (b), where it can be observed that initially, the SAPF currents are zero. But once, SAPF is switched on, it starts supplying only the non-linear load demand.



(a)



(b)

Fig. 11. Experimental Results: With Unbalanced Non-linear load, (a) Grid Current without and with compensation (b) SAPF Currents without and with compensation

The Fig. 11 shows the experimental results with unbalanced non-linear load, where initially the SAPF is not in operation and therefore, whatever load demand is there that is being supplied from grid side on. In Fig. 11. (a), the traces of 3P4W grid side currents are shown with and without compensation. Here, it may be observed that initially, the grid is supplying the 3P4W non-linear load demand. But as soon as SAPF is switched on, the grid side currents become purely sinusoidal and in phase with their respective grid side phase voltages. The corresponding SAPF currents without and with compensation are shown in Fig. 11. (b), where it can be observed that initially, the SAPF currents are zero. But once, SAPF is switched on, it starts supplying only the non-linear and unbalance load demand. This fact may also be verified from neutral current, where under without compensation, the grid is supplying the neutral current. But once, SAPF is switched on, the grid side neutral current becomes zero, while the SAPF starts supplying the same.

The Fig. 12 shows the waveforms of grid side voltage, grid side current, load current and SAPF current waveforms. Here, it can be observed that load current comprises of both non-linear as well as reactive current component. Initially, the SAPF is switched off and thus, the grid is supplying the total

1551-3203 (c) 2018 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.

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/TII.2018.2875035, IEEE Transactions on Industrial Informatics

load demand as also evident from grid and load current profile. But as soon as SAPF is switched on, the grid current becomes purely sinusoidal and in phase with their respective grid side phase voltage while the SAPF keep on supplying the non-linear part of load current demand.

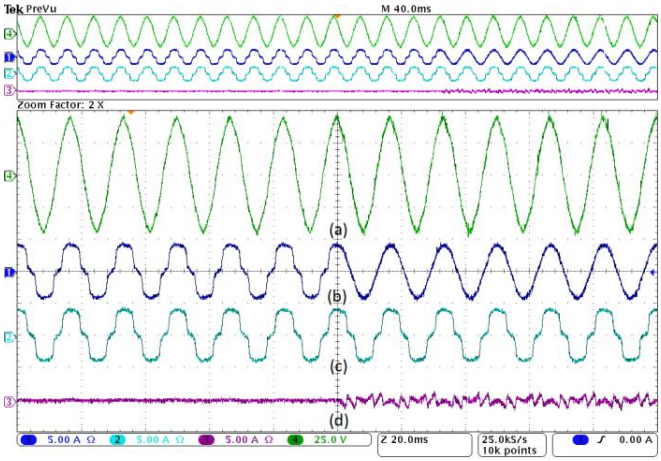


Fig. 12. Experimental Results: Without and With compensation, Traces of (a) Grid side voltage, (b) Grid side current, (c) Load current, (d) SAPF current.

The Fig. 13 shows the experimental results under dynamic load conditions. Here, it can be observed that the load current profile remains perfectly sinusoidal and in phase with grid voltage despite of huge variation in load current (almost 100% of its initial value). Thus, the proposed RC exhibit excellent performance under different operating conditions: variable and unbalanced loading conditions.

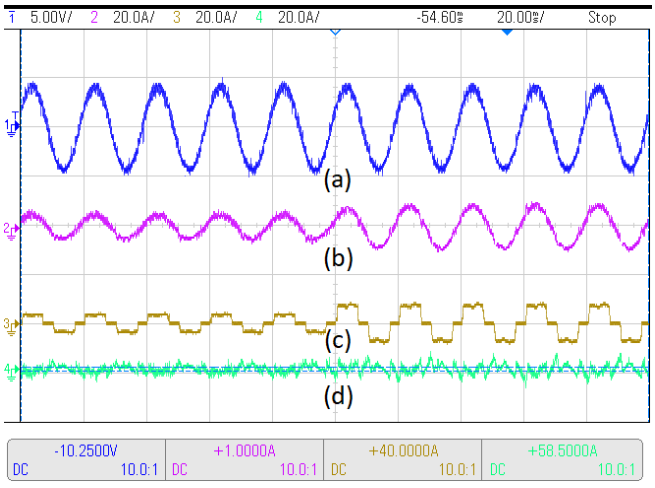


Fig. 13. Experimental Results: With variable load, Traces of (a) Grid side voltage, (b) Grid side current, (c) Load current, (d) SAPF current.

VI. Conclusion

A modified RC controller modelling, design and simulation for the control of SAPF have been presented in this paper. From the design and modelling prospective, it has been demonstrated that the modified RC controller has better control bandwidth than conventional RC. The proposed controller provides satisfactory results under the dynamic operating conditions. It has also been demonstrated that the SAPF is able to handle the non-linear load demand both under balanced and unbalanced conditions. To validate these facts, simulations results are provided. Finally, the extensive experimental results have also been presented to support the claims. Here, it can be noticed that the current harmonics, current unbalance and reactive current demand of load are compensated in such a way that the grid side currents are always maintained as a 3-phase balanced set of fundamental current at UPF. Moreover, the load neutral current demand is supplied locally from the fourth leg of SAPF so that the grid side neutral current could be limited to zero value.

Appendix

Table-I

Grid Currents	Before Compensation			After Compensation		
	Current (r.m.s.)	% THD	(%) Unbalance Factor	Current (r.m.s.)	% THD	(%) Unbalance Factor
Phase-A	3.05	15.2	23.98	2.72	3.1	0.25
Phase-B	1.92	24.1	21.95	2.72	3.1	0.25
Phase-C	2.41	18.4	02.03	2.70	3.2	0.50
Neutral	0.9			0		
Power Factor	0.87			0.99		

System Parameters

3-phase Supply (r.m.s.)	:	V_g=45 *V*, 50 *Hz*
3-phase Non-linear Load	:	R=40Ω, L=15 *mH*
1-phase Linear Load (A-N)	:	R=40Ω, L=15 *mH*
1-phase Non-Linear Load (C-N)	:	R=74Ω, L=15 *mH*
DC-Link Capacitance & Voltage:		C_{dc}=3400 μF, V_{dc}=150 *V*
Coupling Inductance	:	L_{sh}=2.0 *mH*
Switching Frequency	:	10 kHz
Sampling Frequency	:	10 kHz
No. of Period, n	:	200
No. of Delays, l	:	10

References

[1] Pablo Acu˜na, Luis Mor´an, Marco Rivera, Juan Dixon, and Jos´e Rodriguez, "Improved Active Power Filter Performance for Renewable Power Generation Systems," IEEE Trans. Power Electronics., vol. 29, no. 2, pp. 687–694, Feb. 2014.

[2] M. Aredes, J. Hafner, and K. Heumann, "Three-phase four-wire shunt active filter control strategies," IEEE Trans. Power Electron., vol. 12, no. 2, pp. 311–318, Mar. 1997.

[3] H. Akagi, "New trends in active filters for power conditioning," Ind. Appl. IEEE Trans. On, vol. 32, no. 6, pp. 1312–1322, 1996.

[4] C. Lascu, L. Asiminoaei, I. Boldea, I. Boldea and F. Blaabjerg, "High performance current controller for selective harmonic compensation in active power filters," IEEE Trans. Power Electron., vol. 22, no. 5, 1826-1835, Sep. 2007.

1551-3203 (c) 2018 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.

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/TII.2018.2875035, IEEE Transactions on Industrial Informatics

[5] Q. N. Trinh and H. H. Lee, "An advanced current control strategy for three-phase shunt active power filters," IEEE Trans. Ind. Electron., vol. 60, no. 12, pp. 5400-5410, Dec. 2014.

[6] Y. Suresh, A. K. Panda and M. Suresh, "Real-time implementation of adaptive fuzzy hysteresis-band current control technique for shunt active power filter," IET Power Electron. vol. 5, no. 7, pp. 1188-1195, 2012.

[7] L. Wang, C. S. Lam, M. C. Wong, N. Y. Dai, K. W. Lao and C. K. Wong, "Non-linear adaptive hysteresis band pulse-width modulation control for hybrid active power filters to reduce switching loss," IET Power Electron., vol. 8, no. 11, pp. 2156-2167, Nov. 2015.

[8] F. Briz, P. Garcaí, M. W. Degner, D. Daíz-Reigosa and J. M. Guerrero, "Dynamic behavior of current controllers for selective harmonic compensation in three-phase active power filters," IEEE Trans. Ind. Appl., vol. 49, no. 3, pp. 1411-1420, May-Jun. 2013.

[9] C. Lascu, L. Asiminoaei, I. Boldea and F. Blaabjerg, "Frequency response analysis of current controllers for selective harmonic compensation in active power filters," IEEE Trans. Ind. Electron., vol. 56, no. 2, pp. 337-347, Feb. 2009.

[10] D. D. Puerto-Flores, J. M. A. Scherpen, M. Liserre, M. M. D. Vries, M. J. Kransse and V. G. Monopoli, "Passivity-based control by series/parallel damping of single-phase PWM voltage source converter," IEEE Trans. Control Syst. Technol., vol.22, no. 4, pp. 1310-1322, Jul. 2014.

[11] Y. Gui, W. Kim and C. C. Chung. "Passivity-based control with nonlinear damping for type 2 STATCOM systems," IEEE Trans. Power System., vol. 31, no. 4, pp. 2824-2833, Jul. 2016.

[12] Y. Han, L. Xu, M. M. Khan, C. Chen, G. Yao and L. D. Zhou, "Robust deadbeat control scheme for a hybrid APF with resetting filter and ADALINE-based harmonic estimation algorithm," IEEE Trans. Ind. Electron., vol. 58, no. 9, pp. 3893-3904, Sep. 2011.

[13] M. Singh, and A. Chandra, "Real Time Implementation of ANFIS Control for Renewable Interfacing Inverter in 3P4W Distribution Network," IEEE Trans. Industrial Electron, vol. 60, no.1, pp. 121-128, Jan. 2013.

[14] M. Qasin, P. Kanjiya, and V . Khadkikar, "Artificial-neural-networkbased phase-locking scheme for active power filters," IEEE Trans. Power Electron., vol. 61, pp. 3857 – 3866, Aug. 2014.

[15] A. Trivedi, and M. Singh, "Repetitive Controller for VSIs in Droop Based AC-Microgrid," IEEE Trans. Power Electronics, Vol. 32, No. 8, pp. 6595-6604, Aug.-2017.

[16] Qingzeng Yan, Xiaojie Wu, X. Yuan, and Y. Geng, "An Improved Grid-Voltage Feedforward Strategy for High-Power Three-Phase Grid-Connected Inverters Based on the Simplified Repetitive Predictor," IEEE Trans. Power Electronics, Vol. 31, No. 5, pp. 3880-3897, May. 2016.

[17] G. Pandove, A. Trivedi, and M. Singh, "Repetitive control-based single-phase bidirectional rectifier with enhanced performance," IET Trans. Power Electronics., vol. 9, no.5, pp. 1029-1036, Apr. 2016.

[18] B. G. Sun, Y. X. Xie, H. Ma and L. Cheng, "Analysis and application of repetitive control scheme for three-phase active power filter with frequency adaptive capability," Journal of Electrical Engineering and Technology, vol. 11, no. 3, pp. 618-628, May. 2016.

[19] B. Francis and W. Wonham, "The internal model principle for linear multivariable regulators," Applied Mathematics and Optimization, vol. 2, no. 2, pp. 170–194, 1975.

[20] P. Schuelting, C. H. van der Broeck and R. Doncker, "Analysis and Design of Repetitive Controllers for Applications in Distorted Distribution Grids," in IEEE Transactions on Power Electronics, 2018.

[21] E. Kurniawan, Z. Cao, and Z. Man, "Design of Robust Repetitive Control With Time-Varying Sampling Periods," IEEE Trans. Ind. Electron., vol. 61, no. 6, pp. 2834–2841, Jun. 2014.

[22] J. C. Vasquez, J. M. Guerrero, M. Savaghebi, J. Eloy-Garcia, and R. Teodorescu, "Modeling, Analysis, and Design of Stationary-Reference-Frame Droop-Controlled Parallel Three-Phase Voltage Source Inverters," IEEE Trans. Ind. Electron., vol. 60, no. 4, pp. 1271–1280, Apr. 2013.

[23] Z.-X. Zou, K. Zhou, Z. Wang, and M. Cheng, "Frequency-Adaptive Fractional-Order Repetitive Control of Shunt Active Power Filters," IEEE Trans. Ind. Electron., vol. 62, no. 3, pp. 1659–1668, Mar. 2015.

[24] Y. Yang, K. Zhou, H. Wang, F. Blaabjerg, D. Wang, and B. Zhang, "Frequency Adaptive Selective Harmonic Control for Grid-Connected Inverters," IEEE Trans. Power Electron., vol. 30, no. 7, pp. 3912–3924, Jul. 2015.

[25] C. Xie, Xin Zhao, M. Savaghebi, L. Meng, Josep M. Guerrero, J. C. Vasquez, "Multi-Rate Fractional-Order Repetitive Control of Shunt Active Power Filter Suitable for Microgrid Applications," IEEE Journal of Emerging and Selected Topics in Power Electronics, 2016.

[26] Y. Ye, Y. Wu, G. Xu and B. Zhang, "Cyclic Repetitive Control of CVCF PWM DC–AC Converters," in IEEE Transactions on Industrial Electronics, vol. 64, no. 12, pp. 9399-9409, Dec. 2017.

[27] W. Lu, K. Zhou, M. Cheng, J. Shen and L. Wu, "Parallel structure general repetitive controller for general grid-connected PWM converters," in IET Power Electronics, vol. 10, no. 3, pp. 338-347, 3 10 2017.

[28] Lu, W., Zhou, K., Wang, D., et al.: 'A generic digital nk ± m order harmonic repetitive control scheme for PWM converters', IEEE Trans. Ind. Electron., vol. 61, no. 3 pp. 1516–1527, 2014.

[29] Byung-Sub Kim, and Tsu-Chin Tsao, "Robust repetitive controller design with improved performance," in the proceedings of American control conference, Arlington, USA, June 25-27, 2001.

[30] Q. Zhao and Y. Ye, "A PIMR-Type Repetitive Control for a Grid-Tied Inverter: Structure, Analysis, and Design," in IEEE Transactions on Power Electronics, vol. 33, no. 3, pp. 2730-2739, March 2018.

BIOGRAPHY

Gitanjali Pandove has done her B. Tech from Nagpur University, India M. Tech from AMU Aligarh, India and Ph.D from Dr. K. N. Modi University, India. Presently, she is working as Professor in Electronics and Communication Engineering deptt. at Deenbandhu Chhotu Ram University of Science and Technology (D.C.R.U.S.T.) Murthal, Sonipat, Govt. of Haryana, India. Her research interest includes, Digital Control, Power Electronics, Renewable Energy, Optical communication, Image Processing and Digital signal processing.

Mukhtiar Singh received the B.Tech. and M.Tech. degrees in electrical engineering from National Institute of Technology (Erstwhile REC Kurukshetra), Kurukshetra, India, in 1999 and 2001, respectively, and the Ph.D. degree from Ecole de Technologie Superieure, University of Quebec, Montreal, QC, Canada. He is presently working as full Professor in the Department of Electrical Engineering, Delhi Technological University (DTU), New Delhi, India.

Prof. Singh is recipient of the IEEE student scholarship for one of the best paper in 34th IEEE Conference on Industrial Electronics, IECON-2008, held at Orlando, FL, USA. He is also the winner of Researchers Merit Scholarship of the University of Quebec, Montreal, Canada for the three years consecutively, 2008, 2009, and 2010.

1551-3203 (c) 2018 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.

Applied Physics A (2018) 124:739
https://doi.org/10.1007/s00339-018-2158-4

Applied Physics A
Materials Science & Processing

Source/Gate Material-Engineered Double Gate TFET for improved RF and linearity performance: a numerical simulation

Skanda Shekhar[1] · Jaya Madan[1,2] · Rishu Chaujar[1]

Received: 14 June 2018 / Accepted: 4 October 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
This work seeks to present a comparative analysis of linear and analog/RF performances of a silicon (Si)-source double gate tunnel field effect transistors (DG-TFET), germanium (Ge)-source DG-TFET and Gate Material Engineered germanium (GME-Ge)-source DG-TFET. The objective of this analysis is to probe the viability of the Ge-source-DG-TFET and GME-Ge-source-DG-TFET as possible components of RF/microwave systems vis-à-vis conventional Si-source-DG-TFETs. The Ge-source-DG-TFET has been shown to have superior performance parameters as compared to the Si-source-DG-TFET. It is realized that the GME-Ge-source-DG-TFET preserves the superior performance of the Ge-source-DG-TFET and displays improved device reliability in contrast to the latter. With these ideas in the backdrop, a study of parasitic capacitances, transconductance (g_{m1}) and its higher order coefficients (g_{m2} and g_{m3}), and signal performance metrics such as VIP2, VIP3, third-order Input Intercept Point (IIP3) and third-order Intermodulation Distortion (IMD3) sheds light on the linear performances of the three devices. Further, the analog/RF performances of the three devices are investigated through analog/RF Figures of Merit (FOM) including current gain, unilateral power gain and maximum available power gain (Gma). This analysis reveals that the GME-Ge-source-DG-TFET is suitable for high-frequency, low-power operation required in modern communication systems.

1 Introduction

Since its inception, Metal Oxide Semiconductor Field Effect Transistors (MOSFETs) have become the mainstay of modern electronics with them being used everywhere from computer chips to communication systems. The primary factors that made them attractive transistor candidates in the past were their small size, high efficiency and optimal current switching ratio (I_{ON}/I_{OFF}). However, limitations arising due to physical laws governing the operation of MOSFETs act as impediments to mankind's increasing need to make devices smaller and more energy efficient [1, 2]. These limitations manifest in the form of drain-induced barrier lowering, velocity saturation and hot carrier degradation leading to a reduction in the threshold voltage (V_{th}) of the transistor at high drain voltages (V_{ds}), an adverse impact on the transfer characteristics of the device due to ballistic conduction and an increase in the leakage current at the drain of the transistor respectively. These are collectively called Short Channel Effects (SCEs) as they arise due to reduction of the channel length of the transistor which is inevitable in the scale-down of MOSFETs [3, 4]. Further, the subthreshold swing (SS) of the MOSFET is limited to 60 mV/decade. This limitation is also fundamental in nature as it follows directly from Fermi–Dirac statistics that govern the energy distribution of electrons in the transistor, leaving little scope for improvement of SS of MOSFETs. These limitations have given impetus to the development of devices that are driven by mechanisms of majority carrier injection other than thermal carrier injection that is prevalent in MOSFETs. One such device is the Tunnel Field Effect Transistor (TFET), i.e., a gated p–i–n device that is usually reverse biased. The drive current in TFETs is a result of quantum mechanical Band-to-Band Tunneling (BTBT) allowing the device to

✉ Rishu Chaujar
chaujar.rishu@dtu.ac.in

Skanda Shekhar
skanda.shekhar15@gmail.com

Jaya Madan
jaya.madan@chitkara.edu.in

[1] Microelectronics Research Lab, Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110042, India

[2] Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

Published online: 08 October 2018

have SS lower than 60 mV/decade [5–8]. The main strength of TFETs lies in the ease with which they can be integrated into the existing CMOS architecture allowing us to use similar techniques to fabricate devices with TFETs [9]. Thus, TFETs have become increasingly promising transistor candidates over time.

Work done earlier in the field of Source Material Engineering (SME) has revealed the superiority of Ge, as opposed to conventional Si, as the source material of Double Gate TFET (DG-TFET) due to improved transfer characteristics and performance parameters like V_{th}, SS and I_{ON}/I_{OFF} [10]. A Ge-source is able to induce steeper band bending at the source–channel interface on the application of gate voltage as compared to a Si-source owing to the lower bandgap of Ge (0.67 eV) than Si (1.14 eV) [11–13]. This reduces the length of the tunneling barrier (L_{BW}) at the source–channel junction allowing for the observed performance enhancements. Further, a Ge–Si heterojunction does not produce a kink at the source–channel junction as is the case with InAs–Si heterojunctions, thus making Ge the preferred source material even though InAs has a bandgap (0.36 eV) that is lower still [12, 14]. Further, it has earlier been observed that the use of Gate Metal Engineering (GME) scheme with a Ge-source DG-TFET preserves the performance enhancements obtained by the use of Ge-source while demonstrating improved immunity to Interface Trap Charges (ITCs) [15]. ITCs are unwanted charges that inevitably appear at the semiconductor-dielectric interface while fabricating the device. Improved immunity to ITCs directly translates into improved device reliability [16–18]. Therefore, the GME-Ge-source DG-TFET combines the advantages of both, the SME and GME schemes to exhibit improved device performance and reliability in contrast to the Si-source DG-TFET with only a negligible trade-off in I_{OFF}. Hence, the three devices, viz., Si-source-DG-TFET (D1), Ge-source-DG-TFET (D2) and GME-Ge-source-DG-TFET (D3) form the corpus of study for this comparative analysis (Table 1).

Today, MOSFETs and the CMOS architecture are widely used in RF/microwave applications and communication systems. Even though TFETs are currently in their embryonic stages, it is important to develop TFETs with low linear distortion suitable for use at RF/microwave frequencies. This will allow for the deployment of such devices in analog systems making the systems to not only smaller but also more energy efficient. For this purpose, in this work, a study of the linearity of the aforementioned devices through signal performance metrics like VIP2, VIP3, third-order Input Intercept Point (IIP3) and third-order Intermodulation Distortion (IMD3) has been carried out. Transistors used in RF/microwave applications must necessarily maintain linear operation while receiving a large number of unwanted signals that are stronger than the desired signal. This prevents the weaker desired signal from being drowned out by the stronger undesired signals [16, 19]. A study of this nature is, thus, essential for most modern communication systems where distortion must be minimized through optimization of transistor structure and circuit design. The analog/RF performance of the three devices is explored through analysis of the analog /RF Figures of Merit (FOM) including current gain, unilateral power gain and maximum available power gain (Gma) [20–22]. The objective of this study is to analyze the viability of the use of D2 and D3 for RF/microwave applications.

2 Device structure and simulation

2.1 Device structure

Figure 1a–c illustrates the lateral cross-section of the devices under consideration (DUC), i.e., Si-source-DG-TFET (D1), Ge-source-DG-TFET (D2) and GME-Ge-source-DG-TFET (D3) considered in this work. It consists of three semiconducting regions, i.e., source, channel, and drain. The channel is sandwiched by dielectric layers and metal gates on both sides giving rise to the DG structure. The DG structure provides increased gate controllability over the channel, thereby allowing the transistor to achieve lower OFF-state and higher ON-state currents [23, 24]. The DUC has been designed with a p+ (n+) Si source (drain) with a uniform impurity concentration of 10^{20} $(5\times10^{18})/cm^3$. The lightly p-doped Si channel of the DUC has a length, thickness and uniform impurity concentration of 50 nm, 10 nm and $10^{15}/cm^3$ respectively. The dielectric used in the DUC is SiO_2 which has a relative electrical permittivity of 3.9, while the

Table 1 Detailed description of different device architectures, i.e., D1, D2 and D3 along with their metal work function used during the analysis

Device	Label	Source	Channel	Drain	Gate metal work function
Si-source DG-TFET	D1	Silicon	Silicon	Silicon	$\Phi_m = 4.3$ eV
Ge-source DG-TFET	D2	Germanium	Silicon	Silicon	$\Phi_m = 4.3$ eV
GME-Ge-source DG-TFET	D3	Germanium	Silicon	Silicon	Φ_{m1} (=4.1 eV) near source–channel interface and Φ_{m2}(=4.3 eV) near channel–drain interface

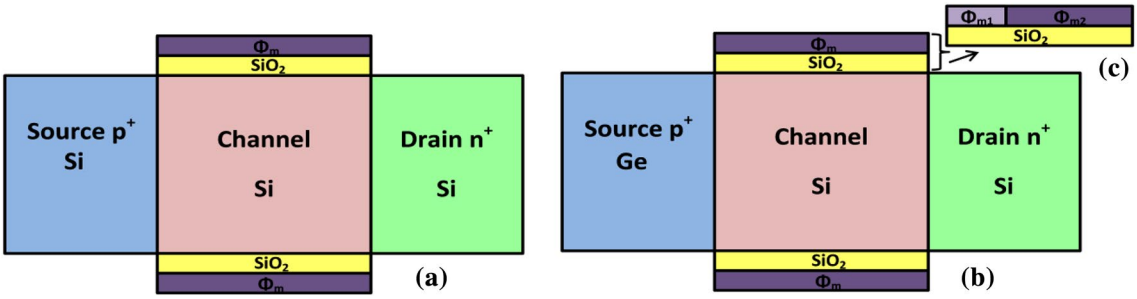


Fig. 1 2D Schematic of the **a** Si-source-DG-TFET (D1), **b** Ge-Source-DG-TFET (D2) and **c** GME-Ge-source-DG-TFET (D3) considered in this work

gate metal workfunction (Φ_m) has been optimized to value of 4.3 eV [10, 15]. All junctions in the DUC are considered to be abrupt. *Ceteris paribus*, two changes are made to the DUC to implement the SME and GME schemes. First, the source of the DUC is altered from conventional Si to Ge to create a heterojunction at the source–channel interface leading to steeper band bending at the tunneling junction (SME scheme). This results in the Ge-source DG-TFET, i.e., D2. Next, a modification is made to the gates of the DUC such that they now consist of two metals with workfunctions (Φ_{m1} and Φ_{m2}) of 4.1 eV and 4.3 eV respectively (GME scheme). The metal with lower workfunction (Φ_{m1}) is placed closer to the tunneling junction at the source–channel interface so as to provide the change in flatband voltage (V_{fb}) necessary to result in the desired reliability improvements. Such an arrangement ensures that the reduction in V_{fb} occurs where the degradation due to ITCs is maximum, i.e., at the source–channel interface (or where the electric field is maximum) [15, 16]. This arrangement gives rise to the GME-Ge-DG-TFET, i.e., D3, thus explaining the structure of the three devices that are central to this study.

2.2 Simulation methodology

All numerical simulations in this analysis have been performed on ATLAS Silvaco TCAD. To perform the numerical simulations accurately in a manner that is closest to the operation of a realizable physical device, the mathematical models used in the simulation environment must be carefully chosen to mimic the physical phenomena that actually eventuate in the working of the device. Since the major source of the drive current in TFETs is BTBT, this phenomenon must be simulated with care. The electric field at the tunneling junction varies dynamically, and an electron must tunnel through this field to get from the source to the drain. For this purpose, here the Nonlocal Band-to-Band Tunneling (N-BTBT) model has been used instead of the Kane model (approximates the potential barrier to have a triangular profile) to obtain precise simulation results. The Shockley–Read–Hall (SRH) recombination model is used to model majority and minority carrier recombination in the device. At large impurity atom concentrations, a phenomenon called Band Gap Narrowing (BGN) causes a reduction in the bandgap of the doped semiconductors. This phenomenon has been simulated by the BGN model, while the Fermi–Dirac statistical model has been used to simulate the charge carrier distribution in the device. The simulator solves the drift and diffusion equations under the mobility model. The aforementioned models are solved using both, Newton's fully coupled and Gummel's decoupled methods. All simulations have been performed at 300 K.

3 Results and discussion

3.1 Enhancement in transfer characteristics of DG-TFET

Figure 2 shows the transfer characteristics of the D1, D2 and D3 at $V_{ds} = 1.0$ V. It can be clearly seen that the use of a Ge

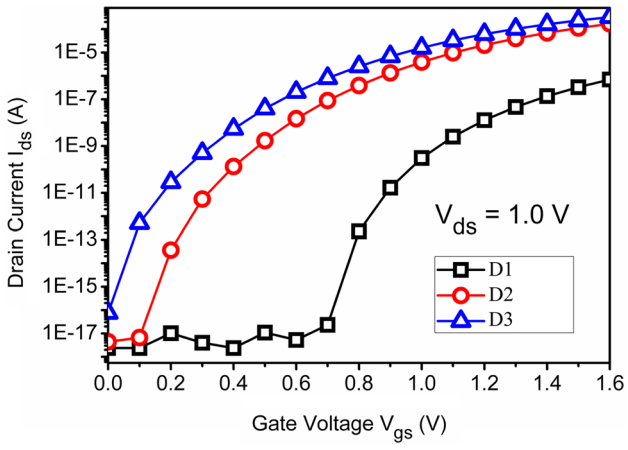


Fig. 2 I_{ds} vs. V_{gs} characteristics of Si-source-DG-TFET (D1), Ge-source-DG-TFET (D2) and GME-Ge-source-DG-TFETs (D3)

source considerably improves the transfer characteristics of the D1 due to reduction in the L_{BW} at the tunneling barrier. This, in turn, allows more electrons to tunnel through the potential barrier at the source–channel interface leading to a higher drive current and faster turn ON of the device (or concomitantly, lower V_{th} and SS) [10, 15]. Specifically, the improvements in V_{th} and SS obtained by changing the source material has been reported earlier by the authors [15]. Moreover, the use of GME scheme in the D3 further improves upon these characteristics of the D2 with lower V_{th} and SS and slightly higher ON-state current. This improvement in V_{th}, SS and I_{ON}/I_{OFF} has also been reported in our earlier work [15]. Furthermore, the GME scheme also imparts improved immunity towards interface trap charges to the D3 [15]. This improves the reliability (which is discussed in detail in [15]) of the device as ITCs are present in all practically realizable devices; their concentration can be reduced but they cannot completely eliminated. However, the OFF-state current of the D3 increases to an order of magnitude of 10^{-17} A from its original range of 10^{-18} A in the case of D2 and D1. Since this degradation is negligible and within tolerable limits for most practical applications, the D3 is a suitable candidate for use in engineering systems subject to propitious analog/RF and linear performance, which is being investigated in detail in this work.

3.2 Study of parasitic capacitances

Parasitic capacitance is the undesired capacitance that arises as a consequence of the proximity of two components in electronic systems. It is of particular importance in devices that have dimensions in nano-scales, since the electrodes of such devices (like the DUC) are inherently in close proximity to each other. In the DUC, there are primarily three parasitic capacitances, viz. gate-source capacitance (C_{gs}), gate-drain capacitance (C_{gd}) and total gate capacitance (C_{gg}), the behaviours of which must be investigated [21, 25, 26]. Even though the construction and nomenclature of components in TFETs is similar to standard MOSFETs, there is one essential difference in the distribution of parasitic capacitance in the two devices. In MOSFETs, C_{gs} is seen to be approximately equal to C_{gd} (as a result of a large potential drop across the channel–drain junction), and both C_{gs} and C_{gd} contribute equally to C_{gg}. However, in TFETs the source–channel junction is reverse biased leading to a large potential drop across this junction and a comparatively smaller one across the channel–drain junction of the device [8, 22, 26]. Consequently, C_{gd} is much larger than C_{gs} in TFETs as charge is mostly decoupled from the source due to the large potential drop across the source–channel junction. By implication, C_{gg} is largely independent of C_{gs} and depends mainly on C_{gd}. Together, C_{gs}, C_{gd} and C_{gg} degrade the performance of the DUC by increasing the response time, active power dissipation and propagation delay of the device [27]. At high frequencies, they may also cause the circuit to execute parasitic oscillations. Hence, it is necessary to limit the parasitic capacitance of a device to a tolerable limit within which the device can be efficiently operated. Since C_{gg} is seen to be approximately equal to C_{gd} in TFETs, the behaviours of all three parasitic capacitances can be probed by merely analysing those of C_{gs} and C_{gg}. Figure 3a, b show the variations of C_{gs} and C_{gg} with V_{gs} (@V_{ds} = 1.0 V) and V_{ds} (@V_{gs} = 1.2 V) respectively. The small signal AC analyses done to study C_{gs} and C_{gg} in this work have been performed at a frequency of 1 MHz. It can be clearly seen from Fig. 3 that the use of the SME scheme in the Ge-source DG-TFET does not significantly change the parasitic capacitances when compared with the standard D1. It can also be observed that the variations of C_{gg} of the D2 with both V_{gs} and V_{ds} remain almost identical to that of the D1 over the entire range of voltages studied. Therefore, the use of the SME scheme in the device under consideration does not adversely impact the inherent parasitic capacitance of the D1. However, there is a minor degradation in the parasitic capacitance of the D3 due to the use of the GME scheme. This can be attributed to the reduction of V_{fb} at the source–channel interface of the D3,

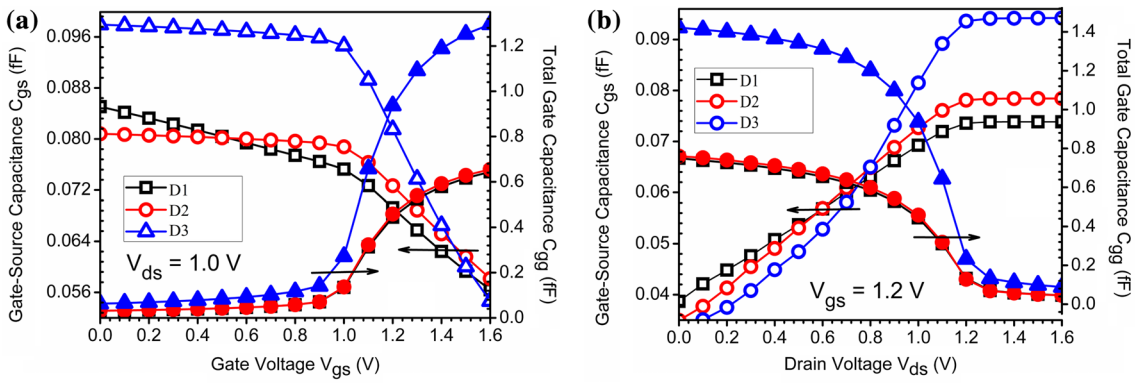


Fig. 3 Variation of **a** C_{gs} and C_{gg} with V_{gs} at constant V_{ds} = 1.0 V, **b** C_{gs} and C_{gg} with V_{ds} at constant V_{gs} = 1.2 V

which, in turn, leads to an increase in the effective voltage ($=V_{gs}–V_{fb}$) at the junction. This increased potential drop at the source–channel interface can be held accountable for the degradation in the parasitic capacitance of the D3. However, the C_{gg} of the D3 at V_{gs} = 1.2 V (@V_{ds} = 1.0 V) and V_{ds} = 1.2 V (@V_{gs} = 1.2 V) only increases by a factor of 0.97 and 0.58, respectively, as compared to the D1. Hence, the degradation caused by the use of GME scheme is within the tolerable limit for most applications, and can be overlooked when improved analog performance and device reliability are taken into account.

3.3 Improvements in linear performance of DG-TFET

MOSFETs have been ubiquitously employed in analog/RF IC design and wireless communication systems as linear amplifiers. Any contender to the MOSFET in this respect must exhibit low distortion and high linearity if the device is to be used in a practical engineering system [19]. In this work, we investigate the linearity of the three devices, labelled as D1, D2 and D3 as they are being proposed as alternatives to well-designed and widely-implemented transistors with the objective of their use in small-size, low-power applications. The Volterra series has been the industry standard for the study of non-linear systems since long [19]. Therefore, it is used for the analysis of both, nature and source of distortion that occurs in analog circuits. Its substantial complexity often impels researchers to shy away from its use even though insights so obtained shed light on the source of the distortion that routinely used numerical simulations do not. In the case of weakly non-linear, memoryless systems, the Volterra series reduces to a Taylor expansion allowing for closed-form expressions of non-linearity. Such a reduction to a Taylor series places a limitation on the nature of applications that this method of analysis can be used for. Consequently, no satisfactory results may be obtained in the case of strongly non-linear devices like power amplifiers and devices which are not memoryless. However, since this paper is not concerned with such limiting cases, the closed-form expression of non-linearity is based on the Taylor series form of the Volterra series, where the coefficients in the series are obtained from the device simulator. Expanding drain current (I_{ds}) in terms of V_{gs} and V_{ds}, we obtain

$$\begin{aligned} I_{ds}(V_{gs}, V_{ds}) = I_0 &+ g_{m1}V_{gs} + g_{d1}V_{ds} + g_{m2}V_{gs}^2 + g_{md}V_{gs}V_{ds} \\ &+ g_{d2}V_{ds}^2 + g_{m3}V_{gs}^3 + g_{m2d}V_{gs}^2V_{ds} \\ &+ g_{md2}V_{gs}V_{ds}^2 + g_{d3}V_{ds}^3 + \ldots \end{aligned} \quad (1)$$

Here, (g_{m1}, g_{m2}, g_{m3}, …), (g_{d1}, g_{d2}, g_{d3}, …) and (g_{md}, g_{m2d}, g_{md2}, …) are the transconductance, output conductance and cross-modulation coefficients, respectively.

If the drain is shorted at the signal frequency, the output conductance and cross-modulation coefficients vanish leaving only the transconductance terms in the equation. In such a scenario, VIP2, VIP3, IIP3, and IMD3 can be mathematically defined as follows [28, 29]:

$$\text{VIP2} = 4 \times \frac{g_{m1}}{g_{m2}} \quad (2)$$

$$\text{VIP3} = \sqrt{24 \times \frac{g_{m1}}{g_{m3}}} \quad (3)$$

$$\text{IIP3} = \frac{2}{3} \times \frac{g_{m1}}{g_{m3} \times R_s} \quad (4)$$

$$\text{IMD3} = \left[\frac{9}{2} \times (VIP3)^3 \times g_{m3}\right]^2 \times R_s \quad (5)$$

$$1 - \text{dB compression point} = 0.22\sqrt{\frac{g_{m1}}{g_{m3}}} \quad (6)$$

where g_{mn} is the n^{th} derivative of I_{ds} with respect to V_{gs} and R_s = 50 Ω. Physically, VIP2 signifies the extrapolated input voltage at which the first and second order harmonic voltages are equal. VIP3 is an indication of the extrapolated input voltage at which the first and third order harmonics are equal. IIP3 is the extrapolated input power at which the first and third order harmonics are equal. Finally, IMD3 represents the extrapolated intermodulation power at which the first and third order intermodulation harmonics are equal [16].

Figure 4a shows the variation of transconductance (g_{m1}) of the D1, D2 and D3 with V_{gs} at V_{ds} = 1.0 V. In essence, g_{m1} can be defined as the change in the drive current through the device or I_{ds} per unit change in the voltage applied at the gate electrodes or V_{gs}. It can be seen clearly in Fig. 4a that the D3 has highest g_{m1} followed by D2 and D1, respectively. There is approximately an 89- and 209-fold increase in the g_{m1} of D2 and D3, respectively, as compared to the D1. This substantially improved g_{m1} can be directly attributed to the increased probability of electron tunneling through the source–channel junctions of the D2 and D3 in comparison to the D1 allowing the former devices to respond better to changes in V_{gs}. Therefore, SME not only improves I_{ds}, I_{ON}/I_{OFF}, V_{th} and SS, but also g_{m1}.

Figure 4b indicates the variation of g_{m2} with V_{gs} at V_{ds} = 1.0 V. The variation of g_{m2} can be explained through variations in g_{m1}; variation of g_{m2} with V_{gs} follows the same trend as that of g_{m1}, g_{m2} of D3 is the highest followed by the D2 and D1. Among the harmonic distortions due to non-linearities of higher-order derivatives of g_{m1}, the third order transconductance (g_{m3}) fixes the lower bound on linear distortion of the device, thus necessitating the minimization of

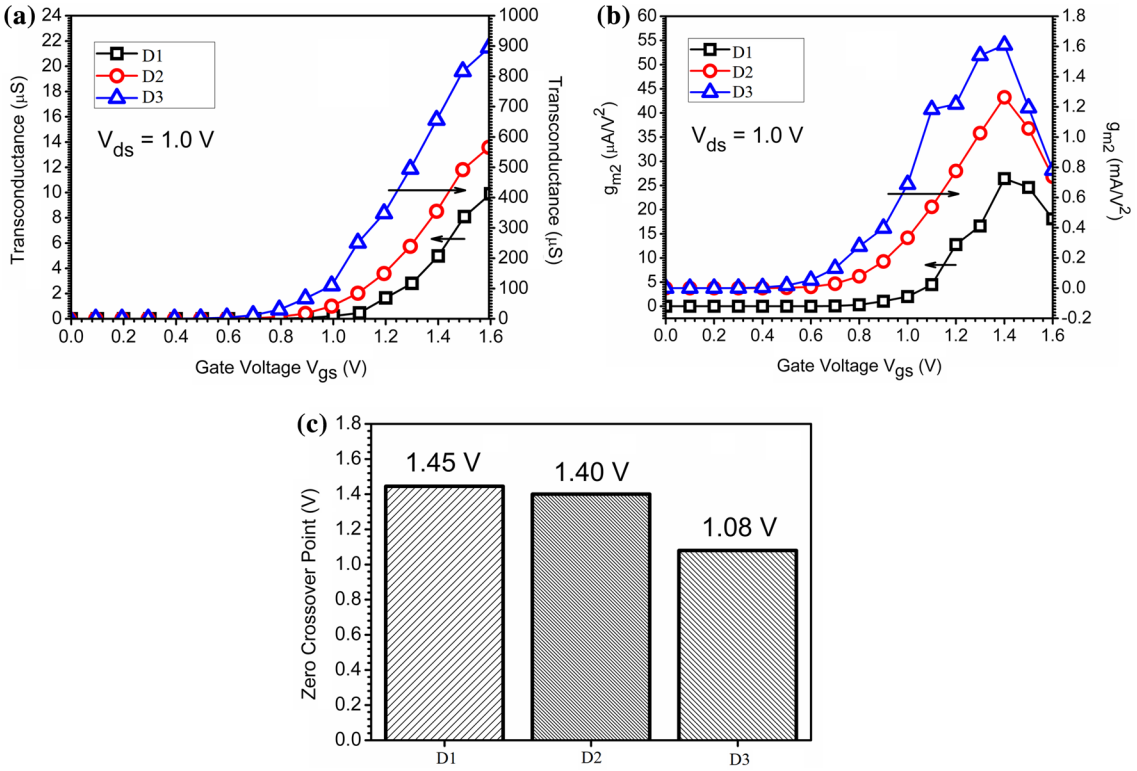


Fig. 4 Plot of **a** g_{m1}, **b** g_{m2}, and **c** zero-crossover point of the D1, D2 and D3

g_{m3}. Further, the point at which g_{m3} crosses the V_{gs} axis is known as the zero-crossover. It determines the bias point for optimum operation of the device. Operation of the device at the zero-crossover bias cause the harmonic distortion for a small signal to be minimized as the variations of g_{m3} on either side of the zero-crossover point are approximately balanced. Figure 4c shows the zero-crossover points of the three DUCs. The D3 has the lowest zero-crossover point followed by the D2 and the D1, respectively. The zero-crossover point of the D3 is lower than that of the D1 and D2 by 34% and 30%, respectively. This reduction in the value of the zero-crossover point of the D3 is due to the faster turn-ON of the D3 as compared to the D2 and D1 allowing low distortion operation of the D3 at lower V_{gs}. This opens avenues for the design of low-power, low-distortion devices.

Next, the linearity FOM, viz., VIP2, VIP3, IIP3, and IMD3 are studied to analyze the linear performance of the D1, D2 and D3. Figure 5a represents the variation of VIP2 with V_{gs} at constant V_{ds} = 1.0 V. A higher value of VIP2 is desirable for low distortion operation. It is clearly seen that there is very little variation in VIP2 of the three devices. VIP2 of the D1 is higher than that of the D2 and D3 by a factor of 0.63 and 0.45 at V_{gs} = 1.2 V, which is negligible for most practical applications. Variation of VIP3 of the three devices with V_{gs} at V_{ds} = 1.0 V has been indicated in Fig. 5b. As is the case with VIP2, devices with higher values of VIP3 display better linearity. Therefore, it can be concluded that, in terms of VIP3, the performance of the D3 is better than that of the D1 and D2. In fact, at V_{gs} = 1.2 V, there is an increase in VIP3 of the D1 by 8% due to the use of the D2, and 91% due to the use of the GME scheme along with the Ge-source. The next linearity FOM under consideration is IIP3, the variation of which for the three devices with V_{gs} at V_{ds} = 1.0 V can be seen in Fig. 6a. Again, the linear performance of a device increase with an increase in IIP3 of the device. In this respect, it can be observed in Fig. 6a that the IIP3 of the D3 is the highest across most values of V_{gs} followed by that of the D2 and D1, respectively. Quantitatively, the values of IIP3 of the D2 and the D3 are higher than the IIP3 of the D1 by 0.16 and 2.70 times. respectively at V_{gs} = 1.2 V. This indicates a large improvement in IIP3 as a result of the inclusion of SME and GME schemes in conventional DG-TFET design. IMD3 is an important linearity FOM in linear amplifier design that represents the amount of intermodulation distortion that occurs as a result of the intermodulation products formed by the mixing of two or more signals within the bandwidth of the amplifier. Devices with lower values of IMD3 are observed to have the ability to withstand higher signal distortion [30]. Figure 6b indicates the variation of IMD3 of the D1, D2 and D3 with V_{gs} at constant V_{ds} = 1.0V. The IMD3 of the D1 is found to be the lowest followed by the D2 and D3, respectively, indicating

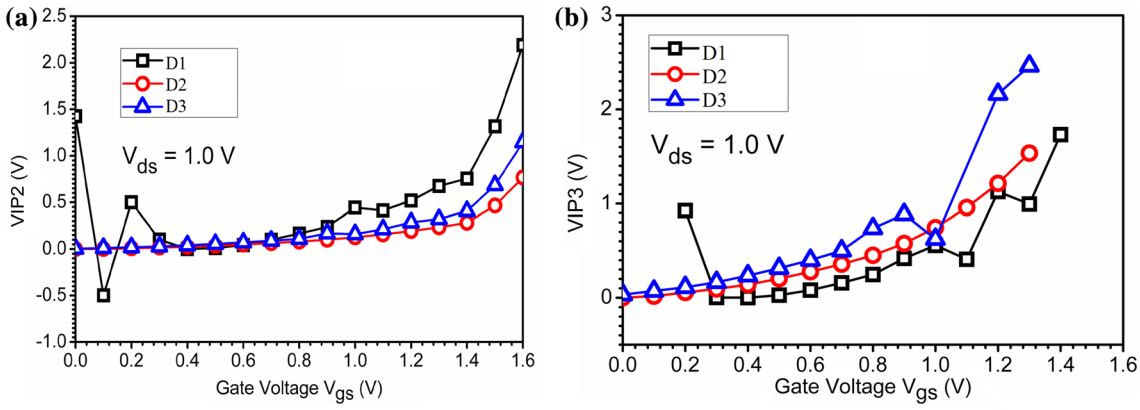


Fig. 5 Variation of **a** VIP2 and **b** VIP3 of the D1, D2 and D3 with V_{gs} at $V_{ds} = 1.0$ V

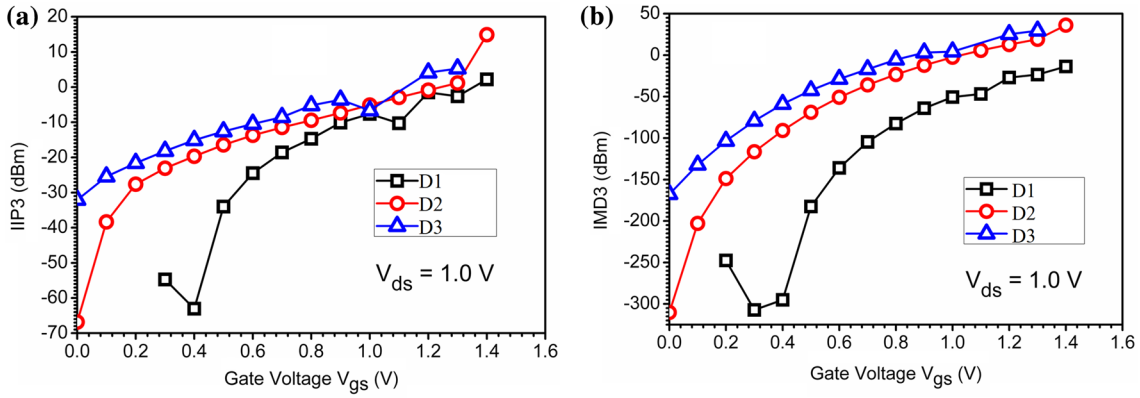


Fig. 6 Plot of **a** IIP3 and **b** IMD3 of the DUCs with V_{gs} at $V_{ds} = 1.0$ V

that the intermodulation distortion performance of the conventional TFET is the best among the three devices. However, linear amplifier designers often try to balance a lower value of IMD3 with a higher value of another linearity FOM called 1-dB compression point.

Beyond a certain value of input power, the gains of most amplifiers begin to saturate leading to a concomitant saturation in the value of output power generated by the amplifier. This means that any further increase in input power leads to little or no increase in the output power of the amplifier. At this stage, the response of the amplifier begins to become non-linear producing signal distortion and intermodulation products. This process of saturation of the gain of the amplifier is known as compression. In this regard, the 1-dB compression point is said to be that value of input power at which the real saturated output power of the amplifier is 1-dB lower than the imaginary value of output power that would have been obtained had the amplifier continued to follow a linear trend in the non-linear domain as well (i.e., 1-dB lower than the value of output power if compression had not taken place). Figure 7 displays the 1-dB compression

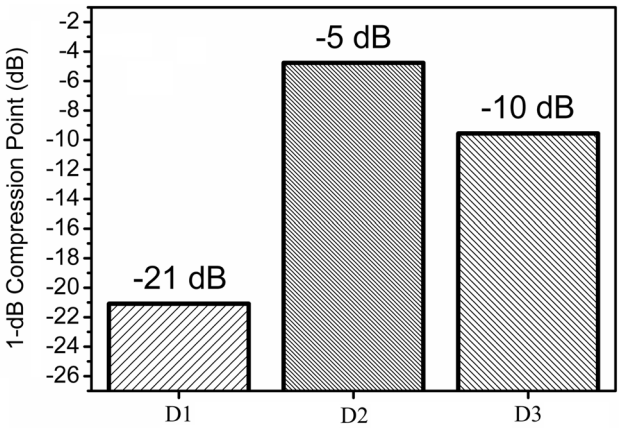


Fig. 7 1-dB compression points of the three DUCs

points of the D1, D2 and D3, respectively. It is observed that the use of SME and GME schemes enhances the value of the 1-dB compression point of the conventional DG-TFET manifold. The 1-dB compression point of the D2 is observed

to be the highest followed by the D3 and D1, respectively. The 1-dB compression point of the D2 and D3 are 78% and 55% higher than that of the D1, respectively. This substantial increase in the 1-dB compression point of the D2 and D3 offsets the degradation values of IMD3 displayed by them as the 1-dB compression point is the real point at which signal distortion starts affecting the output, whereas IMD3 is a set of imaginary extrapolated points at which the device may never be operated in practice. As indicated by Eqs. (2–6) the variations of VIP2, VIP3, IIP3, IMD3 and 1-dB compression point depend on the variations of g_{m1} and g_{m2}, g_{m1} and g_{m3}, g_{m1} and g_{m3}, g_{m1}, g_{m2} and g_{m3}, and g_{m1} and g_{m2}, respectively. Thus, it is the interplay between g_{m1}, $g_{m2,}$ and g_{m3} to which the variations of the linearity FOM can be attributed. These linearity FOM indicate that the D2 and D3 have considerable advantages over conventional D1 for use in low-power, high-linearity applications.

3.4 Analysis of analog/RF performance

It is imperative to study the analog/RF performance of any device being considered as a potential candidate for high-frequency operations, such as in communication systems. Therefore, in this work, an analysis of three analog/RF FOM, namely current gain, unilateral power gain and maximum available power gain (Gma), has been carried out to assess the viability of employment of the three DUCs in high-frequency applications. Figure 8 shows the variation of the current gain of the DUCs with frequency at constant V_{ds} = 1.0 V and V_{gs} = 1.2 V. It is clear that the D3 offers the highest current gain followed by the D2 and D1 respectively over the range of frequencies that constitute the radio spectrum used for most communication applications, i.e., from 10^5 to 10^{11} Hz. At commonly used cellular transmission and reception frequencies of the order of 10^8 Hz (100 MHz), the current gains of the D2 and D3 are 380% and 395% higher than that of the D1, indicating a substantial improvement in the current gain of the conventional DG-TFET through the use of SME and GME schemes. In DG-TFETs, current gain is defined as the ratio of the magnitude of the drain–source current (output current) to the magnitude of the gate-source current (input current). Therefore, this increased gain of the D3 and D2 can be attributed to the higher I_{ON} and faster turn-ON of the two devices as compared to the D1. Current cutoff frequency (f_T) is defined as the frequency at which the current gain becomes unity (0 dB). The values of f_T of the D1, D2 and D3 as obtained from Fig. 8 have been plotted in its inset.

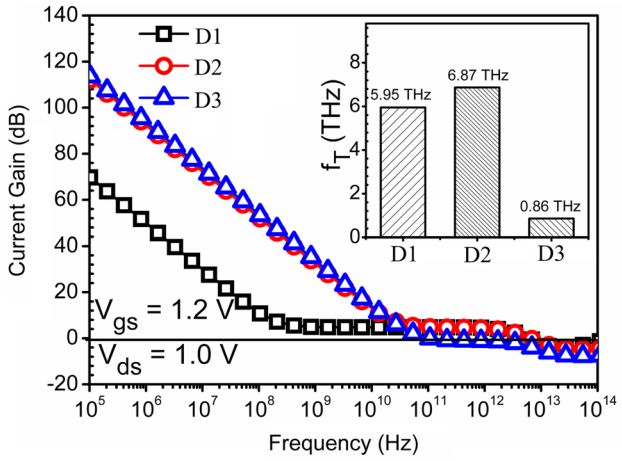


Fig. 8 Variation of current gain with frequency at V_{gs} = 1.2 V and V_{ds} = 1.0 V. Inset: cut off frequency (f_T) of the D1, D2, and D3

f_T Can be considered to be an approximate measure of the intrinsic delay of the transistor and is often deemed to be a first-order FOM of the frequency response of the device. It is evident from the inset of Fig. 8 that f_T of the D2 is higher than that of the D1 indicating that the use of SME scheme leads to an enhancement of f_T. Quantitatively, f_T of the D2 (D3) is 0.15 (0.85) times higher (lower) than that of the D1. This improvement in f_T of the D2 can be attributed to the increased ease with which electrons tunnel through the source–channel junction leading to lower transit times. At the same time, the use of the GME scheme leads to an increase in the transit time of electrons resulting in a reduction of f_T. However, the manifold improvement in other analog/RF FOM of the D3 along with its enhanced reliability more than compensate for this reduction in f_T. Figure 9a is a plot of the variation of the unilateral power gains of the D1, D2 and D3 with frequency at V_{ds} = 1.0 V and V_{gs} = 1.2 V. Here again, it can be observed that the unilateral power gains of the D2 and D3 are higher than that of the D1 for most frequencies of practical importance. However, at frequencies higher than 10^{11} Hz, the unilateral power gain of the D3 decreases rapidly. Therefore, it can be concluded that, over the entire range of frequencies under consideration, the D2 shows more favorable variations in unilateral power gain as compared to both, the D1 and D3. The unilateral power gain is the transducer power gain under the unilateral assumption, i.e., S_{12}=0 for a two-port system. Physically, this indicates the absence of feedback within the device. The unilateral power gain remains invariant under reciprocal, lossless embeddings, and can, therefore, be used as a means to compare the activity or passivity of two devices. Hence, one can say that the D2 and D3 are more active than the D1 for most frequencies under consideration. Another important analog/RF FOM to be considered in the design of low noise amplifiers is Gma. While the unilateral power gain is a measure of the amount of power transferred by the transistor to the terminations, Gma is a measure of the theoretical maximum power gain that a device can offer. The Gma of a device cannot be surpassed. Therefore, it is

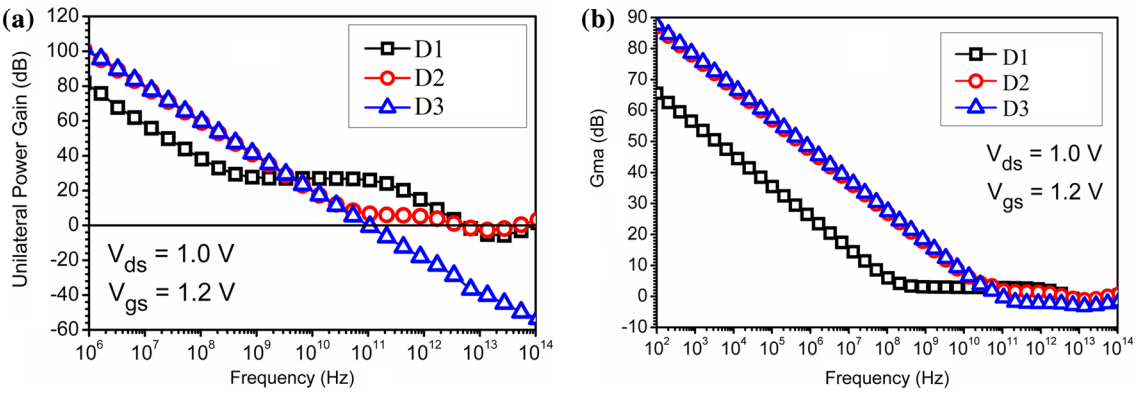


Fig. 9 Plot of **a** unilateral power gain and **b** Gma of the D1, D2 and D3 with frequency at V_{gs} = 1.2 V and V_{ds} = 1.0 V

desirable to design devices with higher Gma as they can, in principle, offer higher power gains when used as actual amplifiers. Figure 9b indicates the variation of Gma with frequency for the three DUCs at constant V_{ds} = 1.0 V and V_{gs} = 1.2 V. It can easily be discerned that the Gma of the D2 and D3 is significantly higher than that of the D1. At a frequency of 100 MHz, there is a 3.45 fold and 3.55 fold increase respectively in Gma of the D2 and D3 as compared to the Gma of the D1. These enhancements in the unilateral power gain and Gma of the D2 and D3 can be attributed to the higher values of I_{ON} (for the same value of V_{gs}) of these devices in contrast to the D1. The aforementioned enhancements in current gain, unilateral power gain and Gma, along with improved transfer characteristics and device reliability, give further impetus to the use of SME and GME schemes to improve the performance of conventional DG-TFETs.

4 Conclusion

In this work, a detailed comparative study of the linear and analog/RF performances of the D1, D2 and D3 is carried out through numerical device simulations. The principal investigation of this study is to evaluate the viability of the D2 and D3 for high-linearity and high-frequency applications. For this purpose, parasitic capacitances, transconductances, linearity FOM including VIP2, VIP3, IIP3, IMD3 and 1-dB compression point, and analog/RF FOM like current gain, f_T, unilateral power gain and Gma are analyzed. These FOM made it apparent that the D2 and D3 are feasible candidates for use in engineering systems. The devices chosen to carry out the analyses in this work also illustrated how the amalgamation of SME and GME schemes with conventional TFET design results in devices suitable for high-linearity and high-frequency operations while preserving the enhancements in device performance parameters that these schemes have been seen to offer in the past. Further, it became evident that the devices considered in this work provide researchers with a direction in which to probe further, in terms of both, ideation and device fabrication. Such research and study is bound to lead to often-envisioned high-speed, low-power devices, which are the need of the future.

Acknowledgements The authors would like to thank the Microelectronics Research Lab, Department of Applied Physics, Delhi Technological University, New Delhi, India.

References

1. T. Sekigawa, Y. Hayashi, Calculated threshold-voltage characteristics of an XMOS transistor having an additional bottom gate. Solid-State Electron. **27**, 827–828 (1984)
2. K. Bernstein, R.K. Cavin, W. Porod, A. Seabaugh, J. Welser, Device and architecture outlook for beyond CMOS switches. Proc. IEEE **98**, 2169–2184, (2010)
3. P.-F. Wang, K. Hilsenbeck, T. Nirschl, M. Oswald, C. Stepper, M. Weis et al., Complementary tunneling transistor for low power application. Solid-State Electron. **48**, 2281–2286 (2004)
4. J. Madan, S. Shekhar, R. Chaujar, PNIN-GAA-tunnel FET with palladium catalytic metal gate as a highly sensitive hydrogen gas sensor, in *Simulation of semiconductor processes and devices (SISPAD), 2017 international conference on*, 2017, pp. 197–200
5. Y. Taur, An analytical solution to a double-gate MOSFET with undoped body. IEEE Electron Device Lett. **21**, 245–247 (2000)
6. N.N. Mojumder, K. Roy, Band-to-band tunneling ballistic nanowire FET: Circuit-compatible device modeling and design of ultra-low-power digital circuits and memories. IEEE Trans. Electron Devices **56**, 2193–2201 (2009)
7. J. Madan, R. Chaujar, Gate drain underlapped-PNIN-GAA-TFET for comprehensively upgraded analog/RF performance. Superlattices Microstruct. **102**, 17–26 (2017)
8. J. Madan, R. Chaujar, Numerical simulation of N + source pocket PIN-GAA-tunnel FET: impact of interface trap charges and temperature. IEEE Trans. Electron Devices **64**, 1482–1488 (2017)
9. A.S. Verhulst, W.G. Vandenberghe, K. Maex, S. De Gendt, M.M. Heyns, G. Groeseneken, Complementary silicon-based heterostructure tunnel-FETs with high tunnel rates. IEEE Electron Device Lett. **29**, 1398–1401 (2008)

10. J. Madan, S. Shekhar, R. Chaujar, Source material assessment of heterojunction DG-TFET for improved analog performance, in *Microelectronic devices, circuits and systems (ICMDCS), 2017 international conference on*, 2017, pp. 1–5
11. A.N. Hanna, H.M. Fahad, M.M. Hussain, InAs/Si hetero-junction nanotube tunnel transistors. Sci Rep **5**, 9843 (2015)
12. K.E. Moselund, D. Cutaia, H. Schmid, M. Borg, S. Sant, A. Schenk et al., Lateral InAs/Si p-type tunnel FETs integrated on Si—part 1: experimental devices. IEEE Trans. Electron Devices **63**, 4233–4239 (2016)
13. A. Hanna, M.M. Hussain, Si/Ge hetero-structure nanotube tunnel field effect transistor. J. Appl. Phys. **117**, 014310 (2015)
14. C.-H. Shih, N.D. Chien, Sub-10-nm tunnel field-effect transistor with graded Si/Ge heterojunction. IEEE Electron Device Lett. **32**, 1498–1500 (2011)
15. J. Madan, S. Shekhar, R. Chaujar, Gate metal engineered heterojunction DG-TFETs for superior analog performance and enhanced device reliability, in *Information and communication technology (CICT), 2017 conference on*, 2017, pp. 1–4
16. J. Madan, R. Chaujar, Interfacial charge analysis of heterogeneous gate dielectric-gate all around-tunnel FET for improved device reliability. IEEE Trans. Device Mater. Reliab. **16**, 227–234 (2016)
17. M. Pala, D. Esseni, F. Conzatti, Impact of interface traps on the IV curves of InAs tunnel-FETs and MOSFETs: A full quantum study, in *Electron devices meeting (IEDM), 2012 IEEE international*, 2012, pp. 6.6. 1-6.6. 4
18. W. Cao, C. Yao, G. Jiao, D. Huang, H. Yu, M.-F. Li, Improvement in reliability of tunneling field-effect transistor with pnin structure. IEEE Trans. Electron Devices **58**, 2122–2126 (2011)
19. R. Chaujar, R. Kaur, M. Saxena, M. Gupta, R. Gupta, Intermodulation distortion and linearity performance assessment of 50-nm gate length L-DUMGAC MOSFET for RFIC design. Superlattices Microstruct. **44**, 143–152 (2008)
20. S. Cho, J.S. Lee, K.R. Kim, B.-G. Park, J.S. Harris, I.M. Kang, Analyses on small-signal parameters and radio-frequency modeling of gate-all-around tunneling field-effect transistors. IEEE Trans. Electron Devices **58**, 4164–4171 (2011)
21. I.M. Kang, J.-S. Jang, W.Y. Choi, Radio frequency performance of hetero-gate-dielectric tunneling field-effect transistors. Jpn. J. Appl. Phys. **50**, 124301 (2011)
22. J. Madan, R.S. Gupta, R. Chaujar, Performance investigation of heterogeneous gate dielectric-gate metal engineered–gate all around-tunnel FET for RF applications, Microsyst. Technol. **23**, 4081–4090
23. K. Boucart, A.M. Ionescu, Double-gate tunnel FET with high-K gate dielectric. IEEE Trans. Electron Devices **54**, 1725–1733 (2007)
24. N. Cui, R. Liang, J. Xu, Heteromaterial gate tunnel field effect transistor with lateral energy band profile modulation. Appl. Phys. Lett. **98**, 142105 (2011)
25. E. Gnani, E. Baravelli, A. Gnudi, S. Reggiani, G. Baccarani, Capacitance estimation for InAs tunnel FETs by means of full-quantum k· p simulation. Solid-State Electron. **108**, 104–109 (2015)
26. S. Mookerjea, R. Krishnan, S. Datta, V. Narayanan, On enhanced Miller capacitance effect in interband tunnel transistors. IEEE Electron Device Lett. **30**, 1102–1104 (2009)
27. J. Madan, R. Gupta, R. Chaujar, in *Advanced Manufacturing, Electronics and Microsystems: TechConnect Briefs*. TCAD Analysis of Small Signal Parameters and RF Performance of Heterogeneous Gate Dielectric-Gate all around Tunnel FET, vol 4 (2015), pp. 189–192
28. S. Kaya, W. Ma, Optimization of RF linearity in DG-MOSFETs. IEEE Electron Device Lett. **25**, 308–310 (2004)
29. P.H. Woerlee, M.J. Knitel, R. Van Langevelde, D.B. Klaassen, L.F. Tiemeijer, A.J. Scholten et al., RF-CMOS performance trends. IEEE Trans. Electron Devices **48**, 1776–1782 (2001)
30. R. Chaujar, R. Kaur, M. Saxena, M. Gupta, R. Gupta, TCAD assessment of gate electrode workfunction engineered recessed channel (GEWE-RC) MOSFET and its multi-layered gate architecture, Part II: analog and large signal performance evaluation". Superlattices Microstruct. **46**, 645–655 (2009)

ISSN 0735-2727, Radioelectronics and Communications Systems, 2018, Vol. 61, No. 9, pp. 394–405. © Allerton Press, Inc., 2018.
Original Russian Text © P. Dawar, A. De, N.S. Raghava, 2018, published in Izvestiya Vysshikh Uchebnykh Zavedenii, Radioelektronika, 2018, Vol. 61, No. 9, pp. 508–521.

S-Shaped Metamaterial Ultra-Wideband Directive Patch Antenna

Parul Dawar[1*], Asok De[2], and N. S. Raghava[2***]**
[1]*Guru Tegh Bahadur Institute of Technology, Delhi, India*
[2]*Delhi Technological University, Delhi, India*
**e-mail: paru.dawar@gmail.com*
***e-mail: asok.de@gmail.com*
****e-mail: nsraghava@gmail.com*
Received in final form November 24, 2017

Abstract—Antenna parameter optimization using S-shaped metamaterial embedded in antenna substrate is elucidated in this paper. Upon incorporation of proposed metamaterial array inside the antenna substrate, the bandwidth of antenna increases by 74% and directivity by about 11%. Results obtained are in good coherence when using the FEM based Ansoft HFSS simulation and MATLAB programming based on CAD formulas using the equivalent circuit analysis of patch antenna. The evolution of 'S-shape' has been explained beginning from the primitive Single Split Ring Resonator's shape. The proposed structure was fabricated and nearly 6% deviation was obtained in comparison with the simulation results. This metamaterial antenna overcomes the low bandwidth limitation of patch antenna and helps in maintaining a low profile by obtaining 81% miniaturization.

DOI: 10.3103/S0735272718090029

1. INTRODUCTION

"Metamaterials" (MTMs) are engineered to modify the bulk permeability and/or permittivity of the medium [1]. They are realized by placing periodically structures that alter the material parameters, with elements of size less than the wavelength of the incoming electromagnetic wave. This results in "meta" i.e. "altered" behavior or behavior unattainable by natural materials [2]. Particle shape has effect on the value of negative permittivity corresponding to the Frohlich resonance. The geometry of negative permittivity particle has a strong effect on its surface plasmonic properties.

The software tool HFSS is used because it is a high performance full wave electromagnetic (EM) field simulator for arbitrary 3D volumetric passive device modeling [3].

This paper abridges the design of RMPA with resonant frequency 4.3 GHz and the operating frequency range 3.8–4.8 GHz having FR4 ($\varepsilon_r = 4.4$) as a substrate material (Section 2) by using the Finite Element Method (FEM) based Ansoft HFSS software and the parametric study using the equivalent circuit analysis in MATLAB. Section 3 describes DNG MTM (S-shaped) having negative refraction in the same frequency range and explains its evolution from the primitive Single Split Resonator structure. Section 4 mathematically demonstrates the resonant frequency of S-shaped MTM using internal inductances and capacitances from circuit diagram. Section 5 deals with MTM application in antenna parameter optimization by embedding it in substrate. Section 6 describes the trade-offs in achieving the bandwidth enhancement in RMPA with DNG material in substrate.

2. RMPA DESIGN

2.1. Design

Rectangular Microstrip Patch Antenna (RMPA) as its name implies consists of a rectangular patch over a microstrip substrate. Its major disadvantage is a relatively low impedance bandwidth, which limits the field of application of these antennas. Metamaterials are embedded in the middle of its substrate to increase the bandwidth of antenna without altering its parameters and without much effecting antenna's radiation properties. Henceforth, metamaterial based antennas are introduced.

REFERENCES

1. H. Schantz, *The Art and Science of Ultra-Wideband Antennas* (Artech House, 2005).
2. C. A. Balanis, *Antenna Theory*, 4th ed. (John Wiley & Sons, Inc., 2016).
3. D. M. Pozar, "Microstrip antennas," *Proc. IEEE* **80**, No. 1, 79 (1992). DOI: 10.1109/5.119568.
4. J. Volakis, *Antenna Engineering Handbook*, 4th ed. (McGraw Hill, 2007).

5. S. Wang. A. P. Feresidis, G. Goussetis, J. C. Vardaxoglou, "Low-profile resonant cavity antenna with artificial magnetic conductor ground plane," *Electronics Lett.* **40**, No. 7, 405 (2004). DOI: 10.1049/el:20040306.
6. M. R. Ahsan, M. T. Islam, M. H. Ullah, N. Misran, "Bandwidth enhancement of a dual band planar monopole antenna using meandered microstrip feeding," *The Scientific World J.* **2014**, ID 856504 (2014). DOI: 10.1155/2014/856504.
7. K. F. Lee, K. M. Luk, K. M. Mak, S. L. S. Yang, "On the use of U-slots in the design of dual-and triple-band patch antennas," *IEEE Antennas Propag. Mag.* **53**, No. 3, 60 (2011). DOI: 10.1109/MAP.2011.6028422.
8. K.-L. Wong, *Compact and Broadband Microstrip Antennas* (Wiley-Interscience, Hoboken, NJ, 2004).
9. Parul Dawar, N. S. Raghava, Asok De, "Ultra wide band, multi-resonance antenna using swastika metamaterial," *Int. J. Microwave Optical Technology* **11**, No. 6, 423 (2016).
10. M. R. Ahsan, M. T. Islam, M. H. Ullah, M. J. Singh, M. T. Ali, "Metasurface reflector (MSR) loading for high performance small microstrip antenna design," *PLOS One*, May (2015). DOI: 10.1371/journal.pone.0127185.
11. S. Linden, C. Enkrich, G. Dolling, M. W. Klein, J. Zhou, T. Koschny, C. M. Soukoulis, S. Burger, F. Schmidt, M. Wegener, "Photonic metamaterials: Magnetism at optical frequencies," *IEEE J. Selected Topics Quantum Electronics* **12**, No. 6, 1097 (2006). DOI: 10.1109/JSTQE.2006.880600.
12. Han Xiong, Jing-Song Hong, Ming-Tao Tan, Bing Li, "Compact microstrip antenna with metamaterial for wideband applications," *Turk. J. Electrical Eng. Comp. Sci.* **21**, 2233 (2013). DOI: 10.3906/elk-1204-6.
13. P. Kaur, S. K. Aggarwal, A. De, "Performance enhancement of rectangular microstrip patch antenna using double H shaped metamaterial," *Radioelectron. Commun. Syst.* **59**, No. 11, 496 (2016). DOI: 10.3103/S0735272716110030.

Volume 6, Issue 3(2018) 175-181 ISSN 2347-3258
International Journal of Advance Research and Innovation

Study for Reduction of Pollution Level in Diesel Engines, Petrol Engines and Generator Sets by Bio signal Ring

SK Singh, Surbhi Jain, Tejasvini Ahuja, Yukti Sharma, Nidhi Pathak
Department of Environmental Engineering, Delhi Technological University, Delhi, India

Article Info:
Article history
Received 18 August 2018
Received in revised form:
28 August 2018
Accepted 2September 2018
Available online 15 September 2018
Keywords: Pollution, Diesel engine, bio signal ring, generator set, atomization, fuel combustion

Abstract
problem in megacities and large urban areas throughout the globe, and transportation is recognized as the major source of air pollution in many cities, especially in developing countries. Contribution of automobiles is reported in the range of 40 to 80% of the total air pollution. The challenge facing megacities is how to reduce the adverse environmental impacts and other negative effects of transportation without giving up the benefits of mobility. This research paper includes the testing of biosignal ring developed by Wellan International Limited for reduction of pollution levels in petrol and diesel vehicles and generator sets. The ring when attached to the inlet of the fuel pipe acts on the fuel through biosignals. The biosignals help in atomization of the fuel and reduces the pollutants by further enhancing the combustion process, and thus reducing the unburnt pollutant particles from the exhaust. The ring was tested on diesel engines, petrol engines and generator sets. For diesel engine and generator set, reduction in smoke was measured, while in the petrol engine parameters such as CO, CO2, SOx and NOx were measured before and after attaching the ring. The percentage reduction of the pollutants obtained from the ring for the diesel engine is about 16.4%, and in petrol engine, significant reduction in the pollutants is observed.The average reduction in the opacity of the diesel generator set exhaust was 54.7%. Therefore, it is evident that the device can effectively control the issue of rising air pollution levels in residential, commercial as well as industrial zones.

1. Introduction

A biosignal is any signal in living beings that can be continually measured and monitored. The term biosignal is often used to refer to bioelectrical signals, but it may refer to both electrical and non-electrical signals. The usual understanding is to refer only to time-varying signals, although spatial parameter variations (e.g. the nucleotide sequence determining the genetic code) are sometimes subsumed as well. There are various types of biosignals. Some of them are stated. Electrical biosignals, or bioelectrical time signals, usually refers to the change in electric current produced by the sum of an electrical potential difference across a specialized tissue, organ or cell system like the nervous system. Thus, among the best-known bioelectrical signals are: Electroencephalogram (EEG) is an electrophysiological monitoring method to record electrical activity of the brain. EEG measures voltage fluctuations resulting from ionic current within the neurons of the brain. Electrocardiography (ECG) is the process of recording the electrical activity of the heart over a period of time using electrodes placed on the skin [6]. Electromyography (EMG) is an electro diagnostic medicine technique for evaluating and recording the electrical activity produced by skeletal muscles. Mechanomyogram (MMG) is the mechanical signal observable from the surface of a muscle when the muscle is contracted. At the onset of muscle contraction, gross changes in the muscle shape cause a large peak in the MMG. Electrooculography (EOG/E.O.G.) is a technique for measuring the cornea-retinal standing potential that exists between the front and the back of the human eye. The resulting signal is called the electrooculogram .Electro dermal activity (EDA) is the property of the human body that causes continuous variation in the electrical characteristics of the skin. Magnetoencephalography (MEG) is a functional neuroimaging technique for mapping brain activity by recording magnetic fields produced by electrical currents occurring naturally in the brain, using very sensitive magnetometers. Biosignals may also refer to any non-electrical signal that is capable of being monitored from biological beings, such as Mechanical signals (e.g. the mechanomyogram or MMG), Acoustic signals (e.g. phonetic and non-phonetic utterances, breathing), Chemical signals (e.g. pH, oxygenation) and Optical signals (e.g. movements).

Biosignals have various applications like water purification, non-invasive diagnosis, potential for direct interfaces and health monitoring, biosignal monitoring system for mobile telemedicine, ways of interaction using biosignlas etc. For water purification,

***Corresponding Author,**
E-mail address: surbhi15.8.jain@gmail.com; sksinghdce@gmail.com
All rights reserved: http://www.ijari.org

biosignals can be used by storing in rings called power rings. POWER RINGS work through biosignals, which have been stored in the rings' interior. These signals penetrate all pipes and tubes to influence the water which is passing through. As a result, the vibrations or oscillations created in the water are changed in such a way that lime, rust, scale or bio-fouling matter no longer accumulate within a piped system and are washed out in minute particles. They can also be used for non-invasive diagnosis. In recent years, a particular challenge has arisen in non-invasive medical diagnostic procedures. Because biosignals recorded on the body surface reflect the internal behaviour and status of the organism or its parts, they are ideally suited to provide essential information of these organs to the clinician without any invasive measures. Potential for direct interfaces and health monitoring: Biosignals are harnessed to enable enhanced computer interfaces. In particular, biosignals can serve for passive health monitoring for the elderly who want to remain at home but need some level of support as they age. Such devices can directly manipulate computer interfaces to give people with limited peripheral mobility some control over their environment. Biosignal monitoring system for mobile telemedicine is yet another application. In this, a prototype integrated mobile telemedicine system has been designed that is compatible with existing public mobile telecommunication network. The mobile telemedicine system consists of two parts. One is a physiological signal measuring part, and the other is a PC system for the signal processing and telecommunication. The system uses NetMeeting to transmit video, audio and patient biosignals from a moving ambulance to a hospital and delivers to the personal computer of the doctor. This PC based mobile telemedicine system is flexible enough to accommodate newer components in wireless communication and portable sensing technologies. There can be ways of interactions using biosignals. A prototype earphone was developed with three kinds of biosignals - pulse wave, electromyogram (EMG) and acceleration sensors. Using this system three new applications were invented, namely, automatic music selection, tactile and visual communication and automatic metadata annotation. Innumerous methods to control pollution have been employed since the awareness of pollution rise has been spread. For example, a muffler is installed in the engine to control the noise and vibrations due to knocking. Earlier, a positive Crankcase Ventilation (PCV) system was used in the automobiles. In this system, the exhaust gasses were sent back into the crankcase so that complete combustion of the un-burnt particles could be carried out. But, this did

not prove to be an effective method since the efficiency was not as desired. Also, the exhaust gasses caused harm to the crankcase. These days, a new technology has been developed. IC engines employ catalytic convertor at their exhaust valves to reduce the smoke emissions. Inside the catalytic convertor, the smoke emissions are adsorbed on the catalyst which makes them unstable, thereby atomizing them. The atomized exhaust gets converted into non-toxic molecules which enter the atmosphere. It proves to be very efficient in reducing the smoke emissions from the engine exhaust valves. In this paper study for reduction of pollution levels in emissions from petrol and diesel vehicles has been done.

2.Pollution Control By Biosignals

Each elemental molecule has its own specific natural atomic/molecular oscillation. In simple terms, oscillations are best described as vibrations or waves that are given off by that element. Molecules are formed when atoms of different elements fuse together because of natural affinity. This attraction is caused by the oscillations emitted by the atoms. In other words, elements recognize each other by their unique oscillations. If the oscillations of various atoms are neutralized their attraction ceases. This is the basis of pollution control by biosignals. The molecules of the fuel are targeted in order to reduce pollution cause by combustion of these particles. When the fuel particles are acted upon by the biosignals, it leads to the atomization of the fuel. So, when the fuel is burned in engine, it leads to the better and uniform combustion. High frequency biosignals are modulated and stored into an appropriate device which is basically high-quality information storage materials (base material crystals) which are capable of storing such vibrations for indefinite periods of time and releasing them again. These biosignals penetrate any kind of material and they are absorbed by the fuel as interference or resonance vibrations. Whenever extra atoms are forcibly fused to form a molecule by the use of a catalyst, the resultant molecule is unstable and will break apart. Biosignal emissions cause such a situation, resulting in a breakdown within molecules. The dispersed atoms are then targeted by the ring to keep from restructuring. The high frequency oscillations that are generated by the energy field interact with the molecular structure of the fuel, removing harmful elements from your fuel and oil. This causes the fuel to become cleaner and finer, leading to better combustion and lesser number of unburnt particles or pollutants and hence reducing pollution.

2.1 Functioning of Power Ring

Power ring consists of high-quality information storage materials (base material crystals) which are capable of storing such vibrations for indefinite periods of time and releasing them again. These vibrations penetrate any kind of material and they are absorbed by the water as interference or resonance vibrations (hydrogen bridges or water molecules). Within fractions of a second, this information passes through the pipe into the water and is effective even with high flow rates. Power rings can be deployed for water of any quality. No chemical substances, magnetic fields or electromagnetic radiations are released into the environment. The basis of this technology is the realization of the latest knowledge of modern quantum theory.

Frequency patterns from the ultra-fine range (quantum vibrations) are modulated onto Power Rings by means of a laser technology. Elements recognize each other by their unique oscillations. Thus, a water molecule is formed when oxygen is attracted to combine with 2 atoms of hydrogen. If the oscillations of various atoms are neutralized their attraction ceases. That is the function of Power ring emissions. Whenever extra atoms are forcibly fused to form a molecule by the use of a catalyst, the resultant molecule is unstable and will break apart. Power ring emissions cause such a situation, resulting in a breakdown within molecules. The dispersed atoms are then targeted by the ring to keep from restructuring.

2.2 Harnessing Signals

The molecular fundamental oscillations of various substances has been isolated, mapped and stored in the Ring. This technology works using specifically modulated molecular oscillations not dependent upon magnetic forces or electrical pulses. Based on these oscillations of elements and molecules, a power ring has been developed. The aim is to influence the original oscillation of the element that has been targeted through new active oscillations in such a way that the physical properties of the element or of the molecules are modified. These manufactured oscillations are then programmed into the Power Ring and super imposed by a proprietary Laser Technology. This alloy can store an almost unlimited number of active oscillations and emit these to a liquid in a constant and stable format, independent of the ambient temperature. The active oscillations create a field within the ring, which penetrates all piping material and thus passes into the water or fluid

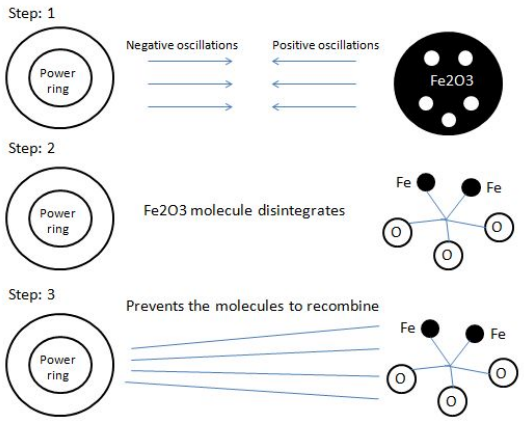


Fig. 1. Functioning of a bio signal ring

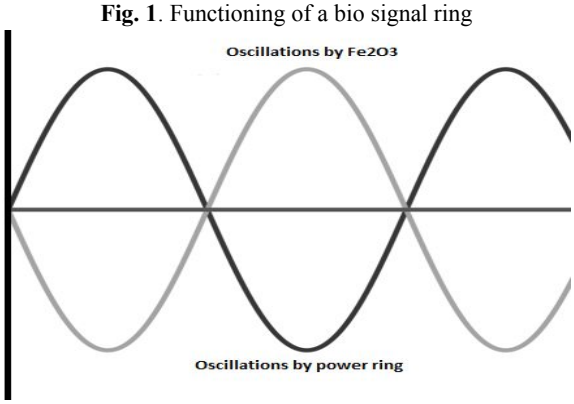


Fig. 2 Interference of natural oscillations emitted by pollutants

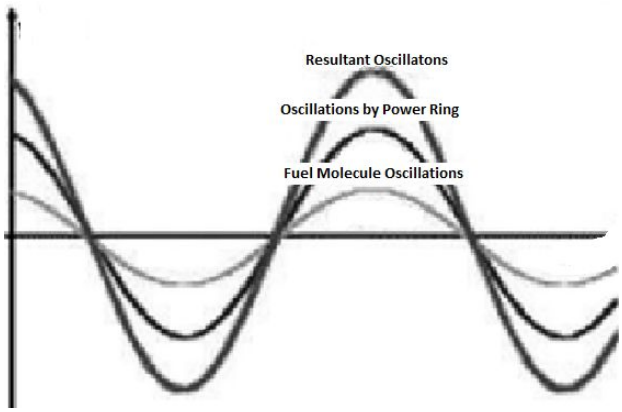


Fig. 3. Overlaying of natural oscillations emitted by pollutants

This Principle results in an 'Interference' or 'Overlaying' of natural oscillations emitted by pollutants in the liquid. Interference and Overlaying can best be described as cancelling sound using equal and opposite waves or playing two adjacent musical notes together to create another note.

3.Vehicular Pollution

The large majority of today's cars and trucks travel by using internal combustion engines that burn gasoline or other fossil fuels. The process of burning gasoline to power cars and trucks contributes to air pollution by releasing a variety of emissions into the atmosphere. Emissions that are released directly into the atmosphere from the tailpipes of cars and trucks are the primary source of vehicular pollution. But motor vehicles also pollute the air during the processes of manufacturing, refueling, and from the emissions associated with oil refining and distribution of the fuel they burn.

The principal air-quality pollutant emissions from petrol, diesel, and alternative-fuel engines are carbon monoxide, oxides of nitrogen, un-burnt hydrocarbons and particulate matter. It is emissions of these pollutants that are regulated by the Euro emissions standards. Modern cars, if kept in good condition, produce only quite small quantities of the air quality pollutants, but the emissions from large numbers of cars add to a significant air quality problem. Carbon monoxide, oxides of nitrogen, and un-burnt hydrocarbons are gases, and are generally invisible. Particulate matter is usually invisible although under certain operating conditions diesel engines will produce visible particles, appearing as smoke. Petrol engines will also produce visible particles if they are burning engine oil or running rich.

3.1 Parameters of Vehicular Pollution

Cars and trucks produce air pollution throughout their life, including pollution emitted during vehicle operation, refueling, manufacturing, and disposal. Additional emissions are associated with the refining and distribution of vehicle fuel. Air pollution from cars and trucks is split into primary and secondary pollution. Primary pollution is emitted directly into the atmosphere; secondary pollution results from chemical reactions between pollutants in the atmosphere. Some of the major pollutants from motor vehicles are particulate matter, hydrocarbons, nitrogen oxides, carbon mono oxide and sulphur dioxide. Particulate matter(PM) are particles of soot and metals which give smog its murky colour. Fine particles — less than one-tenth the diameter of a human hair — pose the most serious threat to human health, as they can penetrate deep into lungs. PM is a direct (primary) pollution and a secondary pollution from hydrocarbons, nitrogen oxides, and sulphur dioxides. Diesel exhaust is a major contributor to PM pollution. Hydrocarbons (HC) pollutants react with nitrogen oxides in the presence of sunlight to form ground level ozone, a primary ingredient in smog. Hydrocarbons contribute to ground-level ozone formation leading to risk of damage to the human respiratory system. Some kinds of hydrocarbons, in addition, are both carcinogenic and indirect greenhouse gases. Nitrogen oxides (NOx) pollutants cause lung irritation and weaken the body's defenses against respiratory infections such as pneumonia and influenza. Oxides of nitrogen include nitrogen dioxide (NO_2) High levels of exposure have been linked with increased hospital admissions due to respiratory problems, while long-term exposure may affect lung function and increase the response to allergens in sensitive people Carbon monoxide (CO) when inhaled, blocks oxygen from the brain, heart, and other vital organs. Fetuses, new-born children, and people with chronic illnesses are especially susceptible to the effects of CO. Carbon monoxide reduces the blood's oxygen-carrying capacity which can reduce the availability of oxygen to key organs. Sulphur dioxide (SO_2) Power plants and motor vehicles create this pollutant by burning sulphur-containing fuels, especially diesel. Sulphur dioxide can react in the atmosphere to form fine particles and poses the largest health risk to young children and asthmatics.

3.2 Standards of Vehicular Pollution

The standards are set by Central Pollution Control Board under the ministry of Environment and Forests and climate change. These are called Bharat Stage Emission Standards that regulate the output of the air pollutants from internal combustion engine equipment. These standards based on European regulations were first introduced in 2000.The standard active now is Bharat Stage-IV norms which has been implemented since 2012.These norms help in bringing down pollution levels. Exposure to air pollution can lead to a number of cardiovascular and respiratory diseases.

Table 1: Emission norms for passenger cars

Norms	CO(g/km)	HC+ NOx(g/km)
1991Norms	14.3-27.1	2.0(Only HC)
1996 Norms	8.68-12.40	3.00-4.36
1998Norms	4.34-6.20	1.50-2.18
India stage 2000 norms	2.72	0.97
Bharat stage-II	2.2	0.5
Bharat Stage-III	2.3	0.35(combined)
Bharat Stage-IV	1.0	0.18(combined)

Table 2: Emission norms for two or three wheelers

Norms	CO(g/km)	HC+ NOx(g/km)
1991Norms	12-30	8-12 (only HC)
1996 Norms	4.5	3.6
India stage 2000 norms	2.0	2.0
Bharat stage-II	1.6	1.5
Bharat Stage-III	1.0	1.0

Table 3: Emission norms for heavy diesel vehicles

Norms	CO (g/kmhr)	HC (g/kmhr)	NOx (g/kmhr)	PM (g/kwhr)
1991Norms	14	3.5	18	-
1996 Norms	11.2	2.4	14.4	-
India stage 2000 norms	4.5	1.1	8.0	0.36
Bharat stage-I	4.0	1.1	7.0	0.15
Bharat Stage-III	2.1	1.6	5.0	0.10
Bharat Stage-IV	1.5	0.96	3.5	0.02

3.3 Measurement of Vehicular Pollution

There are three approaches available to measure vehicle emissions. The first approach involves bringing vehicles into a laboratory, placing the vehicles on a dynamometer, and measuring their emissions as they are operated. The second approach is to place emission measurement and exhaust flow measurement equipment onto the vehicles and measure emissions from the vehicles as they are operated in the field. The third option is to use remote sensing to detect emissions from vehicles as they drive by. The traditional way to measure emissions from vehicles is to bring them into the laboratory and measure emissions as they are operated on a dynamometer. In the better laboratories, the dynamometers can vary the load that they place on a vehicle on a second by second basis. Some older dynamometers are only capable of maintaining a steady load that can be changed, but only while the vehicle is not being tested. In either case, a load is placed on the vehicle that is equivalent to a load that the vehicle might experience while it is operating on a roadway and the emissions are measured from the vehicle. The actual emission measurements are typically made using a Constant Volume Sampling (CVS) system. This system works by pulling a constant volume of air that is larger than the amount of gases emitted by the vehicle being tested through a sampling duct. The exhaust gases from the vehicle being tested are mixed with this makeup air. The makeup air must, of course, be free of gases and

particulate matter in order to not distort the testing process. On-road vehicle emissions measurement equipment has only become realistically available since about 2003. The equipment is often referred to as portable emissions measurement equipment or PEMs.

4.Pollution From Generator Sets

Generator sets are the machines which uses diesel or gas as fuels for converting mechanical energy into electrical energy. Due to continuous development throughout the globe and increasing urbanization, electricity supply has become a basic necessity in day to day life. They are also seen as an alternative source of electricity in those developing regions where there is little or no power supply. This has increased the use of generator sets at an exponential level thereby creating an alarming level of pollution in all regions of the world. Emissions from a diesel generator set engine contain approximately 40% of toxic chemicals such as formaldehyde, arsenic and benzene which are carcinogenic in nature. Other contaminants such as nitrogen are emitted which are responsible for ozone layer depletion. Such compounds pose a threat to both humans as well as the environment. Generator sets in India may range from as low as 60kW to 1MW and above. Generator sets have a large cumulative effect on the ecosystem.

4.1 Parameters of Generator Set Emissions

A number of harmful products are emitted from generator set. The most significant harmful products are HC, NOx, PM and CO. Many reasons such as air-fuel ratio, combustion temperature, turbulence in the combustion chamber and air-fuel concentration make this out of question. NOx has the highest proportion of generator emission with a rate of more than 50%. After NOx emission, the second highest emitted pollutant is PM. The following are the major pollutants from diesel generator set:

Particulate matter (PM) - Particulate matter emissions are resulted from combustion process. They may be generated from partly burned lube oil, ash content of fuel oil, agglomeration of very small particles of partly burned fuel. Most particulate matter is resulted from incomplete combustion of the hydrocarbons in the fuel and lube oil. It is reported that PM consists of elemental carbon (31%), sulfates and moisture (14%), unburnt fuel (7%), unburnt lubricating oil (40%) and remaining may be metals and other substances (Agarwal 2007). These particles cause various health problems such as asthma, lung cancer, premature death and other cardiovascular issues. Nitrogen oxides (NOx) - major influences of the formation of NOx are the concentration of oxygen and temperature in the combustion. The amount of NOx produced is a function of maximum temperature in the cylinder, oxygen concentration, and residence time. Most of the NOx is formed at the starting of the combustion process. NOx emissions lead to formation of ozone, acidification, smog formation etc. NO2 has a level of toxicity five times greater than NO and also it causes direct human lung diseases. Hydrocarbons (HC) - Hydrocarbon emissions are due to unburned fuels as a result of insufficient temperature at the cylinder wall. Hydrocarbons consist of various species such as alkenes, alkanes and aromatics. Diesel engines generally emit low levels of hydrocarbons. In diesel engine the type of fuel, engine adjustment and design affect the hydrocarbon content. Hydrocarbons have various harmful effects on both human health and on environment. They play a significant role in the formation of ground level ozone. Hydrocarbons are toxic with the potential to respiratory track irritation and cause cancer. Carbon monoxide (CO) - Carbon monoxide results from incomplete combustion where oxidation process does not occur completely. Diesel engines are lean combustion engines having a consistently high air-fuel ratio. So, the formation of CO is minimal. Carbon monoxide is an odorless and colorless gas. In humans, CO is inhaled by the lungs and transmitted into the bloodstream. Carbon monoxide reduces the blood's oxygen-carrying capacity which can reduce the availability of oxygen to key organs

4.2 Standards of Generator Set Emissions

Ministry of Environment and Forests, Government of India regulates the Emissions from new diesel engines used in generator set. The regulations impose type approval testing, production conformity testing and labeling requirements. The regulations also include a list of authorized agencies for type approval testing Engine.

Table 4: Emission Standards for Diesel Engines ≤ 800 kW for Generator Sets (2004/2005)

Power (P)	Date	CO	HC	NOx	PM	Smoke
		g/KWh				1/m
P ≤ 19 Kw	2004.01	5.0	1.3	9.2	0.6	0.7
	2005.07	3.5	1.3	9.2	0.3	0.7
19 kW < P ≤ 50 kW	2004.01	5.0	1.3	9.2	0.5	0.7
	2004.07	3.5	1.3	9.2	0.3	0.7
50 kW < P ≤ 176 kW	2004.01	3.5	1.3	9.2	0.3	0.7
176 kW < P ≤ 800 kW	2004.11	3.5	1.3	9.2	0.3	0.7

Table 5: Emission limits for Diesel Engines > 800 kW for Generator Sets

Date	CO mg/Nm3	NMHC mg C/Nm3	NOx ppm(v)	PM mg/Nm3
Until 2003.06	150	150	1100	75b
2003.07 2005.06	150	100	970 (710a)	75c
2005.07	150	100	710 (360a)	75c
a For engines in plants of total power rating above 75/150 MW located in urban/rural areas, respectively. b 150 mg/Nm3 for engines fuelled with furnace oil. c 100 mg/Nm3 for engines fuelle with furnace oil.				

5.Materials and Methods

Opacity of smoke and smoke density factor has been measured to study the vehicular pollution from diesel engine and generator set with the help of smoke meter. Testing has been done as per IS: 8118:2008. Smoke meter, Model no: 437C manufactured by AVL India Private Ltd was used and the analysis was done in the CASRAE lab, Delhi Technological University, Delhi. Opacity of smoke and smoke density factor has been measured to study the emission from generator set with the help of smoke meter. The analysis was done on generator set in electrical station No.1, Delhi Technological University, Delhi. CO, NO_X, C_xH_y, SO_2 and Fuel efficiency has been measured from a Wagon R petrol Engine with the help of digas analyser Model no: 444 manufactured by AVL India Private Ltd. To test the increase in efficiency of the fuel by power ring, first the emissions of the engine without installing the power ring was tested. The engine was allowed to warm up for 30 minutes and then the emissions were tested by the respective instrument. The test was repeated by installing the power ring at the inlet of the fuel pipe.

6.Results and Discussion

mission Test for Petrol Engine

The exhaust from the WagonR petrol engine, model: ED2 2009/10 having engine dynamometer was tested. The test was performed using Low Gas Analyzer in Biodiesel Laboratory. The various parameters recorded are listed in Table 6 and 7.

CO	NO	C_xH_y	SO_2	NO_x	Fuel Efficiency (%)	T_c
277.1	42.3	1306	1.5	160.1	90.40	144s

Table 6: Parameters of petrol emission tested without ring

Table 7: Parameters of petrol emission tested with ring

CO	NO	C_xH_y	SO_2	NO_x	Fuel Efficiency (%)	T_c
265.1	87.5	975.4	0	114.9	95.30	158s

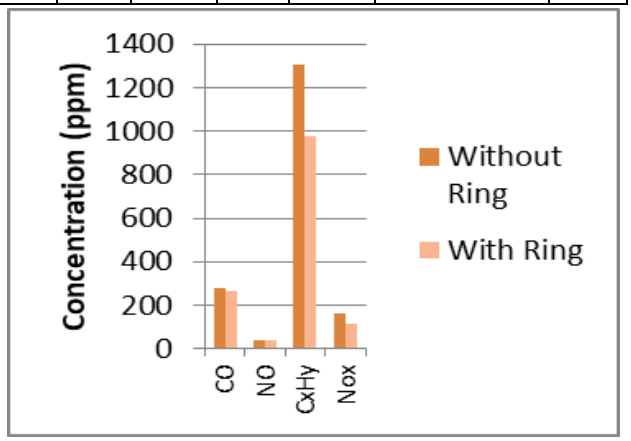


Fig.4. Parameters of Petrol engine emissions

The exhaust from Honda City Car having 4 stroke engine was also tested at Pollution Checking centre in Chandigarh, Haryana (PUC center code P087) with and without the ring. These results would help in realizing the benefits of the ring for practical purposes in the consumer vehicles. The results obtained from the test are as follows:

Table 8. Parameters of Petrol Emission in vehicle without ring

RPM	CO	CO_2	O_2	λ
2647	0.185	13.60	0.38	1.1548

Table 9. Parameters of Petrol Emission in vehicle with ring

RPM	CO	CO_2	O_2	λ
2526	0.0003	17.31	0.07	1.002

The exhaust from the Indica diesel engine, model: AG80ED2 2010/11 having engine dynamometer was tested. The test was performed using Smoke Meter in the Biodiesel Laboratory. Opacity of the smoke measured with and without ring is listed in Table 10 and 11.

Table 10. Parameters of diesel engine emission tested without ring

N(%) Opa-city	K(m^{-1}) Smoke density facto	T(°C) Heat-ing Chamber Temperatur	n(min^{-1}) rpm
17.3	0.42	52	972
16.9	0.41	55	972

Table 11. Parameters of diesel engine emissions tested with ring

N(%) Opa-city	K(m^{-1}) Smoke density factor	T(°C) Hea-ting Chamb Temperature	n(min^{-1}) rpm	t(°C) O Temperature
14.4	0.36	55	970	24
14.6	0.37	56	970	24

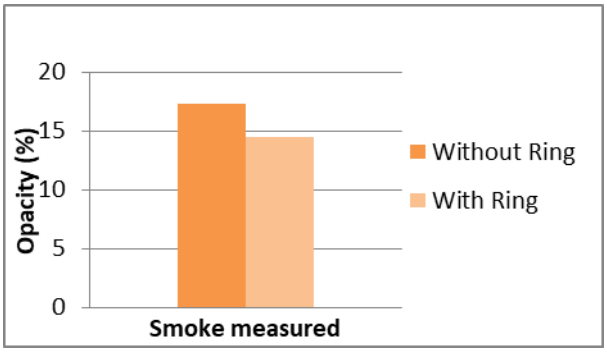


Fig. 5. Comparison of diesel engine emissions

There is a reduction in smoke emissions after installation of the Power Ring. The reductions in Petrol and Diesel engine thus calculated have been summed up in Table 12 and 13.

Table 12. Percentage reduction in emissions in diesel engine

Parameters	Without Ring	With ring	Reduction (%)
CO	277.1	265.1	4.3
NO	42.3	87.5	11.3
C_xH_y	1306	975.4	25.3
SO_2	1.5	0	100
No_x	160.1	114.9	28.3

Table 13: Percentage reduction in emissions in diesel engine

Parameter	Without Ring	With ring	Reduction (%)
Opacity	17.3	14.4	16.76
	16.9	14.6	13.6

The exhaust from the generator set of 320kVh was tested. The test was performed using smoke meter in electrical station No.1, Delhi Technological University, Bawana. The opacity was measured at an interval of 15 minutes and recorded for 30 as well as 45 minutes both with ring and without ring. The AVL opacity meter was used to measure the perimeters. The various parameters that were recorded are listed in Table 3 and 4. Table 3 shows the parameters without the ring whereas the table 4 shows parameters with ring. There is a significant reduction discussed in table 14 and table 15.

Table 14. Parameters of generator set emission without ring

Duration After warming (minutes)	Opacity (N)	(K)	Oil temperature (T)	RPM (n)	Engine temperature (t)
30	77	1.27	91	742	22
45	88.6	1.25	91	742	23

Table 15. Parameters of generator set emission with ring

Duration After warming (minutes)	Opacit (N)	(K)	Oil temperature (T)	RPM (n)	Engine temperature (t)
30	77	1.27	91	742	22
45	88.6	1.25	91	742	23

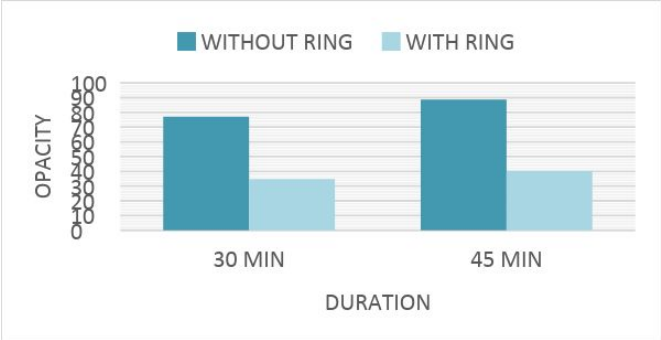


Fig. 6. Generator set emission comparison

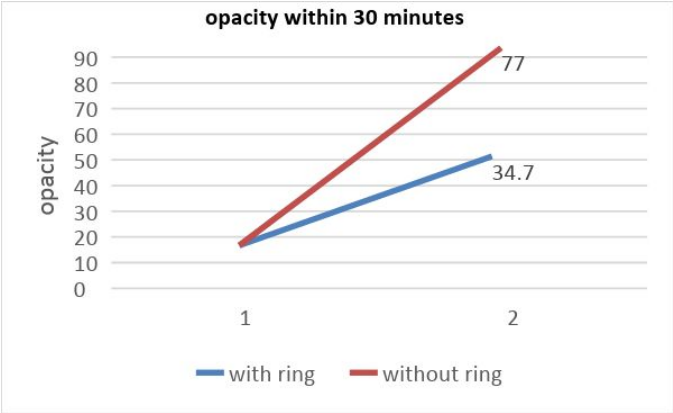


Fig. 7. Opacity witing 30 minutes

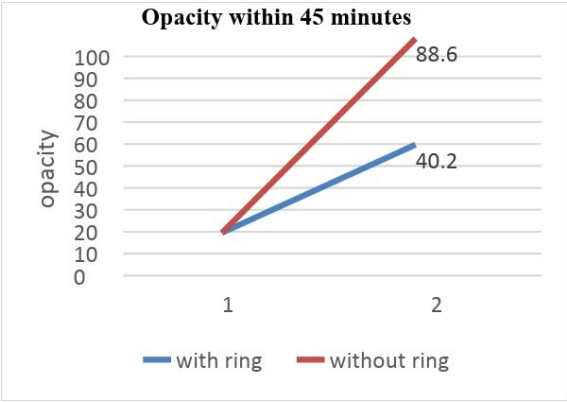


Fig.8. Opacity within 45 minutes

There is a reduction in smoke emissions after installation of the Power Ring. The reduction in generator set engine thus calculated has been summed up in the graphs below. The reduction from the starting till 30 and 45 minutes is compared with and without ring. The graph clearly shows the reduced emissions of ring installation.

There is a reduction in smoke emissions after installation of the Power Ring. The reduction in generator set engine thus calculated has been summed up in Table 14 and 15.

This purpose of the study was to find the performance evaluation of biosignal ring to reduce the level of pollution in diesel and petrol engines. According to the tests performed on the diesel and petrol engine, it can be observed that a significant decrease in the pollutants released by the engine can be achieved with the help of the ring. The overall reduction in diesel engines with the aid of the ring was found to be about 16.18%, which shows that if these rings are fitted in diesel engine it can give a remarkable outcome in alleviating pollution levels. In case of petrol engine, considerable reduction in the value of various parameters was observed. The reduction percentage in hydrocarbons is about 25.3% and that in nitrogen oxides is 28.3%. It is also noteworthy to see that besides reducing pollution levels, the fuel efficiency of the petrol engine has also increased from 90.40% to 95.30% as recorded from the instrument. It was also observed that time taken by the engine to consume 100 mL of fuel without the ring was 144 seconds, and on the application of the ring the time was reduced to 158 seconds. This goes on to show that the application of the ring is of importance not only in reducing pollutants from engine but also in increasing the efficiency of the engine. The reduction percentage is summed up in table 5 and later an average can be calculated for both the values The purpose of the study was to find the performance evaluation of biosignal ring to reduce the level of pollution in generator sets. According to the tests performed on the generator sets, it can be observed that 57.4% (on an average) decrease in the pollutants released by the engine can be achieved with the help of the ring.

5.Conclusions

From the study it is concluded that the Power Ring is very efficient in reducing the pollutants from the atmosphere. Various parameters constituting the smoke in have significantly decreased in the engine exhaust after installation of the ring. The ring works effectively on diesel as well as petrol engines. In petrol engines, maximum reduction is in the oxides of nitrogen The ring was found to be very efficient in reducing the pollution level from the generator exhaust and can reduce a significant amount of pollution from the country if used effectively in all areas. According to the tests performed on the generator set, it can be observed that there is a considerable reduction in the value of smoke emissions. The reduction in percentage of opacity was 54.7% as recorded by the smoke meter. . Besides reducing the pollution, it proves to be economical to install the ring into to engine inlets of the motor vehicles and generator sets. It is so because an increase in burning capacity of the fuel is observed. This indicates increase in fuel efficiency which makes the fuel last longer. Also, as per our previous study, there is a significant increase in efficiency of the fuel after installation of the ring in addition to reduction in pollutants in vehicle engine. This goes on to show that the application of the ring is of importance in two ways, firstly, reducing pollutants from engine and secondly, increasing the efficiency of engines.

The Power Ring is the much needed means to get the rising levels of pollution in control and further also proves to be economical. We recommend the incorporation of Power ring in private and govt. vehicles and generator set to be made compulsory. The idea of the Power Ring is to atomize the fuel at the very inlet of the engine. This gives two positive results, firstly, burning the fuel much more efficiently and secondly, minimizing the emission of the un-burnt carbon particles. This not only helps in reducing the pollution at the exhaust valve, but also increases the efficiency of the fuel by about 5-30%, making it last longer. The Power Ring can be installed in petrol engines, diesel engines and generators. The installation process being simple, it can be attached on the fuel pipe by any individual. Since, automobiles contribute to 80-90% of the air pollution; there is tremendous scope of this Power Ring in the present as well as in the future as it is high time for mankind to act on the rising levels of air pollution especially in megacities.

References

[1]. SK Singh, S Jain, T Ahuja, Y Sharma, N Pathak. Performance evaluation of Biosignal Ring for reduction of Pollution Level in Diesel and petrol Engines. IJIRSET, 6(7), 2017

[2]. New non-invasive sensor can detect brainwaves remotely. University of Sussex. 24 October 2002.

[3]. T J Sullivan, SR Deiss, G Cauwenberghs. A Low-Noise, Non-Contact EEG/ECG Sensor. IEEE Biomedical Circuits and Systems Conference, BIOCAS 2007,. 154–157. doi:10.1109/BIOCAS.2007.4463332.

[4]. YM Chi, P Ng, E Kang, J Kang, J Fang, G Cauwenberghs. Wireless non-contact cardiac and neural monitoring. Proceedings of Wireless Health 2010, 15–23. doi:10.1145/1921081.1921085.

[5]. CJ Harland, TD Clark, R, Prance. Electric potential probes - new directions in the remote sensing of the human body. Measurement Science and Technology. 13 (2), 2002, 163. doi:10.1088/0957-0233/13/2/304.

[6]. CJ Harland, TD Clark, NS Peters, M. Everitt, PB Stiffell. A compact electric potential sensor array for the acquisition and reconstruction of the 7-lead electrocardiogram without electrical

charge contact with the skin. Physiological Measurement. 26 (6), 939. doi:10.1088/0967-3334/26/6/005.
[7]. M Oehler, V Ling, K Melhorn, M Schilling. A multichannel portable ECG system with capacitive sensors. Physiological Measurement. 29 (7), 783 ff. doi:10.1088/0967-3334/29/7/007.
[8]. A Brouse. A Young Person's Guide to Brainwave Music: Forty years of audio from the human EEG. eContact! 14.2 — Biotechnological Performance Practice/Pratiques de performance biotechnologique, 2012.
[9]. M Ortiz. A Brief History of Biosignal-Driven Art: From biofeedback to biophysical performance. eContact! 14.2 — Biotechnological Performance Practice / Pratiques de performance biotechnologique,2012.
[10]. P Lopes, Jefchippewa. Performing Biological Bodies: An open conversation with Marco Donnarumma, Claudia Robles and Peter Kirn at Body Controlled #4 — Bio Interfacing. eContact! 14.2 — Biotechnological Performance Practice / Pratiques de performance biotechnologique 2012.
[11]. eContact! 14.2 — Biotechnological Performance Practice / Pratiques de performance biotechnologique, 7, 2012.
[12]. For EG, GümşelEB, OzgonO. Prediction of ABC Removal by Different Materials from Landfill Leachate: Modeling of Experimental Results XYZ Technique, The Scientific World Journal, 52(2), 2013, DOI: 10.1155/2013/240158
[13]. Rosen (Ed.), Erwin M. The Peterson automotive troubleshooting & repairmanual. Grosset& Dunlap, Inc. ISBN 978-0-448-11946-5, 1975.
[14]. Infrared Remote Sensing of on-Road Motor Vehicle Emissions In Washington State.
[15]. HE Kebin, Z Qiang, HOU Hong. Point source of pollution: Local effects and its control,Types and amount of vehicular pollution,1 2010.
[16]. M Jones. Advanced opacitymeters: their potential role in future emission testing legislation for Diesel vehicles, Proc. of the 6th ETH Conference on Nanoparticle Measurement, Zurich, August 2002, http://nanoparticles.ch/archive/2002_Jones_PR.pdf
[17]. Website of Central pollution control board, Ministry of environment and forests, cpcb.nic.in/Vehicular/Exhaust.php.
[18]. J Lents, Ml Walsh, K He, N Davis, STolvett, Haunliu. Handbook of air quality management.
[19]. Central pollution control board 2008 vehicular emissions. cpcb.nic.in/vehicular/exhaust.php.
[20]. ND ISSRC, JL ISSRC, M Walshe ICCT, www.aqbook.org/read/?page=88.
[21]. A Sano, T Tomita, H Oba. Application using earphone with biosignal sensors, 2009
[22]. MAHasan. Detection and processing techniques of FECG Signal for fatal monitoring, biological procedures Online, 33 pages2009.
[23]. J Yang. A New Filtering algorithm for removing salt and pepper Noise,", international conference on environmental science and information application technology,1,2009, 355-358.
[24]. Website of biomedical engineering, application of biosignal, http://www.embs.org/about-biomedical-engineering/our-areas-of-research/biomedical-signal-processing/.
[25]. Ml Benvoy, JR Cooperstock. Centre for Intelligent Machines, McGill University St. Montreal, Canada,Biosignal analysis and its applications.
[26]. S Anders, M Lotze, W Grodd, N Birbaumer. Brain activity underlying emotional valence and arousal:A response-related Fmri study. Human brain mapping, 23, 2004, 200-209.
[27]. PL Dhar, SK Singh. Studies on solid desiccant based hybrid air-conditioning systems. Applied Thermal Engineering, 21(2), 2001, 119–134
[28]. SK Singh, DN Rao, M Agrawal, J Pandey, D Naryan. Air pollution tolerance index of plants Journal of Environmental Management, 32(1), 1991, 45–55
[29]. SK Singh, DN Rao Evaluation of plants for their tolerance to air pollution - Proceedings of symposium on air pollution control, 1983
[30]. M Agrawal, SK Singh, J Singh, DN Rao Biomonitoring of air pollution around urban and industrial sites.- Journal of Environmental Biology, 1991
[31]. AM Aenab, SK Singh, AJ Lafta. Air Quality Assessment: A Statistical Approach to Stationary Air Monitoring Stations, International Journal of Advanced Research, 3(3), 2015, 68-80.
[32]. AA Gour, SK Singh, SK Tyagi, A Mandal. Variation in Parameters of Ambient Air Quality in National Capital Territory (NCT) of Delhi (India), Atmospheric and Climate Sciences; 5, 2015,13-22. http://dx.doi.org/10.4236/acs.2015.51002
[33]. AM Aenab, SK Singh, AJ Lafta. Critical Assessment of Air Pollution by ANOVA Test and Human Health Effects, Atmospheric Environment, 71, 2013, 84-91
[34]. A Seaton, D Godden, W MacNee, K Donaldson. Particulate air pollution and acute health effects, 345(8943), 1995, 176–178
[35]. JR Goldsmith, LT Friberg - Effects of air pollution on human health Air pollution, 1977 - Academic Press New York
[36]. St. Fleur, Nicholas. Atmospheric Greenhouse Gas Levels Hit Record, Report Says. New York Times, 2015.
[37]. EM Saber, G Heydari. Flow patterns and deposition fraction of particles in the range of 0.1–10 μm at trachea and the first third generations under different breathing conditions. Computers in Biology and Medicine. 42 (5), 2012, 631–638. doi:10.1016/j.compbiomed.2012.03.002. PMID 22445097.
[38]. Remote heartbeat monitor will outperform current technology. University of Sussex Bulletin, 2002.

DOI: 10.20472/IAC.2018.039.042

KAPIL SHARMA
Delhi Technological University, India

TECHNOLOGICAL ASPECT IN REAL TIME BIDDING: A PROBABILISTIC APPROACH

Abstract:

This paper aim is to study the process of real time bidding or real time auctions for online digital advertising. Real time bidding drives the focus of bidding strategy from the user's profile by calculating a bid for each impression in real time. Real Time Bidding uses computers and multiple software's which implements multiple algorithms to display ads per impression via real time auction. It has been seen that by taking different parameters (e.g. conversion rates for a targeted audience), those account for varied prices at different market segments or pricing schemes. The data mining model implemented is the Statistical Arbitrage Mining (SAM). The campaigns use the CPA (cost per action) method on the meta-bidder to accomplish CPM (cost per mille-impressions) ad inventories paradigm thereby reducing the advertiser's risk. In SAM, trying to seek the optimal bidding price to maximize the expected arbitrage net profit is the net goal. A modern portfolio base is implemented to manage the risk. The Expectation - Maximization (EM) fashion is used to estimate the profit of each campaign and thereby maximize it. By using this, the meta-bidder successfully catches the statistical arbitrage opportunities in RTB. Also using the concepts of finance, the calculation of risk is done for each campaign.

Keywords:

Expectation - Maximization; Bidding; Statistical Arbitrage Mining

1.Introduction

Real Time Bidding marks the new era of digital advertising. It is similar to financial markets but follows an instantaneous programmatic approach. In this approach we bought advertising inventory and that is sold on the basis of per-impression basis. It is autonomous and completes a full transaction in milliseconds. The main advantage of Real-time bidding is that it lets advertisers to manage and optimize ads collected from multiple ad-networks. Users are granted multitude of various networks, which allows them to launch and create advertising campaigns. The other advantages are that users can prioritize networks. They are also allocated percentages of unsold inventory. This unsold inventory is known as backfill. RTB is also known as programmatic bidding. The advantage of programmatic bidding is to allow companies which represents sellers and buyers to bid. This bid is on the price to show an ad to a user. This bid happens every time a banner ad gets loaded. When a page or a site gets loaded thousands of bidders bids to serve an ad to that user. That bid is based on individual algorithms for each company. The buying and selling of the advertising inventory takes place on a technology platform known as ad-exchange. This is technology driven as opposed to the earlier idea of negotiating on media inventory. The bid is submitted within a very small amount of time, usually less than 100ms. Real time bidding has led to a significant Return On Investment(ROI) for the advertisers and focuses on behavioral targeting of user data.

The problem statement of this paper is how to determine the bid price while achieving advertiser target such that the net arbitrage profit is maximised while the risk involved is minimized (Both in a single campaign and across multiple campaigns).

A primary motivation for following RTB today is that is autonomous and takes place in milli-seconds. Online Advertising takes place on a large scale today and is a novel topic and thus it made us interested in learning about the various approaches and propose new ideas that can further optimize the process of real time bidding for online advertising. However, the biggest challenge involved is how to estimate and divide the budget across multiple campaigns by minimizing the losses and maximising the profits and also avoiding to exhaust the whole budget.

This rest of the paper has been organized as follows. Section 2 gives the literature review of this project Section 3 explains the methodology used in this project. Section 4 provides the results obtained, Section 5 includes conclusions and Section 6 introduces proposed future work. In the end references are also provided.

2.Realted Literature

In [1] , the whole process of real time bidding is explained which is different to the conventional asserting a predefined advertisement for a specified website . Real time bidding takes into consideration the user's profile and then an auction for the ads are done. This paper explains about the main components of the whole real time bidding process and how the elements are interlinked with each other. For an advertisement to be published on a website, the publisher side sends a request to the RTB exchange which further sends this to the Demand Side Platform (DSP). An auction takes place and whichever advertisement has the highest bid price wins it.

In [2] advertisers buy spots in ad exchanges. Before the reaching of the ad to the audience, it passes through many exchanges and data analysts. To avoid this overhead, the use of advertisers or campaigners are seen. Using their own algorithms to calculate bid prices, they ferret out underpriced ads and resell them.

In [3] , the features of an online exchange are mentioned . It explains how normal advertisements have different goals than online ads. The online ad process has an effect on both sides. It has to prove beneficial for the customer who is viewing it. If it is of relevance and the user clicks on the ad, profit is then only generated by the advertiser side. This is in correlation with the whole basic concept of online advertisements as per the need and choice of the user. Focusses on the goals of an exchange design such as efficiency, expression and simplicity.

In [4], the detailed statistical analysis of this dataset is presented. Real time bidding is an emerging technology which lacked publicly available datasets that could be used to conduct the experiments and create solutions for the real time bidding research problems. In the year 2013 iPinYou organized a real time bidding algorithm competition globally for the duration of three seasons. There were two stages in each season: one was the online stage and the other was the offline stage. A dataset for modelling training and testing later was kept in the offline stage. The dataset consists of number of clicks, number of conversions, final conversions and logs of Ad bidding etc. After the completion of the competition iPinYou released the dataset for public use for all the three seasons. [5] gives a detailed explanation of these datasets that could be used for conducting research in bidding strategies, CTR estimation etc. We have also utilized this dataset in our thesis.

In [6] , the impact of publisher determines the conversion rate or at least the probability of an advertisement to be converted. This helps in further bid optimization so that the appropriate price is placed for the advertisements.

While placing an advertisement on any impression it is imperative to focus on its ability to convert a user into a customer. Due to large amount of data available and uncertainty present in the user data, finding probability for conversion is a difficult task. [7] establishes an approach based on data hierarchical model, where the levels are user, publishers and, advertiser whose previous performances are being analyzed, to estimate the conversion rate. Different levels possess different binomial distribution and the distribution parameters are estimated individually which are ultimately combined using logistic regression approach.

3.Methodology

Our aim is to calculate the highest bid price of a given advertisement such that the auction by that bidder is one, keeping in mind not to exhaust the budget suggested.

To solve this, certain terms have to be kept in mind such as total cost and total profit of a campaign

Total Cost = Sum of {Bidding Price * Probability of Bid Winning * Probability of Campaign Being Selected} for all campaigns

Total Profit = Sum of {(Conversion Rate * Payoff Ratio - Bidding Price) * Probability of Bid Winning * Probability of Campaign being selected} for all campaigns

For single campaign, the ‘Probability of Campaign Being Selected ‘will be 1. So the bid price is found on the basis of total cost and the profit generated by an entire campaign. Depending upon the algorithm used by the particular campaign, the price is deciphered.

3.1. Single Campaign

In single campaign we assume all the bids received in a time interval T, which are represented with the help of a feature vector. The meta-bidder calculates a bidding price using a bidding price with the aim of net profit maximization over the time period. The data is initially a 10 column relational table which holds various parameters of the data set such as the timestamp, region, IP, url, url id, city code etc. The data is then pre-processed to reduce the data to keep only the data columns which will be useful. Only 3 columns are required and they are the number of clicks, market price and conversion rate. The probability of selecting a particular campaign in the single campaign procedure is 1.

We want to maximise the target that is the total profit achieved, so the bidding function is estimated and we try to estimate the bidding value given the following constraints:

E[C] <= B (3.1)

The total expected cost spent in time T by the bidder should be less than or equal to the total budget allocated.

Var[R] <= h (3.2)

The variance to measure the total risk should be less than or equal to a pre-defined upper limit h.

0 <= v <=1 (3.3)

The probability of selection of campaign should be between 0 and 1. In single campaign this value is fixed at 1.

$V^T 1 = 1$ (3.4)

The bidding value is calculated using the derivative of the expected value of bidding function and equating it to zero subject to the constraints mentioned above.

$E[R(v,b(\Theta,r))]=T\sum_{i=1}^{M} v_i \int \left(\theta r_i - b(\theta, r_i)\right) w\left(b(\theta, r_i)\right) p_\theta^i(\theta) d\theta$ (3.5)

- We take Lagrangian of the bidding function: L(b(Ɵ,r),λ)
- Take functional derivative of the above equation with respect to bidding function b(Ɵ,r)
- For different winning functions we can obtain optimal values of bidding function.

$$\left(\frac{\theta r_i}{1+\lambda} - b(\theta, r_i)\right)\frac{\partial w(b(\theta, r_i))}{\partial b(\theta, r_i)} = w(b(\theta, r_i)) \quad (3.6)$$

3.2Multiple Campaigns

Here we assume that there are M campaigns for which the bidder can bid for in time interval T. Basically the procedure is decided in 2 steps:

1. E-step – We fix the value of bidding function and calculate the optimal value of probability of selection of campaign subject to the constraints:
    - Var[R] <= h (3.7)
    - 0 <= v <=1 (3.8)

- $V^t 1 = 1$ (3.9)

2. M-step – We fix the campaign selection probably and estimate the optimal value of bidding function with the aim of profit maximization subject to the budget constraint:

    - $E[C] <= B$ (3.10)

We repeat these steps iteratively until the results converge, all the constraints are satisfied and the target is maximized. In expectation maximization algorithm the campaign selection act as a latent factor which helps to calculate the optimal value of bidding function with all constraints satisfied. The M-step is same as the one explained in single campaign; in addition to that we also have to calculate the campaign selection probability to evaluate where to reallocate the volume for each campaign. We use a new parameter γ called as net profit margin. The net profit margin is the ratio of the net profit of the advertising, either from one campaign or a set of them (meta-bidder), divided by the advertising cost during the corresponding period.

$$\gamma = {}^{R}/_{C} = \text{ROI} - 1 \quad (3.11)$$

Where, ROI is return on investment hence γ is a random variable with variance and expectation. To solve this problem, we use portfolio based risk management methods as described below:

1. Single Campaign: If there is only one single campaign then we calculate the expectation and variance of the net profit margin γ_i for each campaign i. Both $\mu_i(b)$ and $\sigma^2_i(b)$ can be estimated from MCMC methods:

    (i) Repeat N times on sampling T bid requests from the training data and calculate $R_i(v_{i=1}; b)$ and $C_i(v_{i=1}; b)$, then

    (ii) Calculate the expectation and variance using these N observations of $R_i(v_{i=1}; b)$ and $C_i(v_{i=1}; b)$

$$\mu_i(b) = E[\gamma_i] \quad (3.12)$$

$$\sigma^2_i(b) = E\left[\frac{R_i(v_{i=1},b)^2}{C_i(v_{i=1},b)^2}\right] - E\left[\frac{R_i(v_{i=1},b)}{C_i(v_{i=1},b)}\right]^2 \quad (3.13)$$

These are the equations used to calculate the mean and variance of the expected value of net profit margin γ.

2. Campaign Portfolio: If we have M campaigns then rather than a single value of mean we would have a vector of expected net profit margins and a covariance matrix for the net profit margins of M campaign where each entry is the correlation factor between campaign i and j, which can be calculated by routine given the net profit margin time series of the two campaigns i and j. This probabilistic campaign combination is called as campaign portfolio, the mean and variance in this case is defined as:

$$\mu_{p\,(v,\,b)} = v^{T}\ \mu\,(b) \qquad (3.14)$$

$$\sigma 2_{p\,(v,\,b)} = v^{T} \Sigma\,(b)\,v \qquad (3.15)$$

For simplicity, we assume that the net profit margin distribution does not change much w.r.t. the auction volume allocated to the campaign during a short period.

3. Campaign Portfolio Optimization: The E- step can be re written with the help of Lagrangian as:

$$\text{Max } v^{T}\mu\,(b) - \alpha\ v^{T}\Sigma\,(b)\,v \qquad (3.16)$$

$$\text{s. t. } v^{T}1 = 1,\quad 0 <= v <= 1$$

Where, the Lagrangian multipliers α acts as a risk-averse parameter to balance the expected net profit margin and its variance. This optimization framework is widely used as portfolio optimization. When the risk, i.e., the variance of the net profit margin, is not considered, α is set as 0. Then the campaign i with the highest $\mu_i(b)$ will be always selected, i.e., $v_i = 1$, while $v_j = 0$ for all other campaigns j. For E-step, the computationally costly parts are the MCMC methods for evaluating the margin of M individual campaign, the complexity is O (MNT) and the campaign correlation calculation which is, O (M^2NT). The downside risk that we have calculated in the project is based on Post Modern Portfolio Theory which has not been used in the previous research paper to the best of our knowledge.

4. Results and Discussion

4.1. Single campaign portfolio result

The following tables show a comparison of the profit for different algorithms (const, rand, truth, linear, ortb, sam1, sam1c, sam2, sam2c) when the budget proportion is 1/16 and pay-off ratio is 0.2

Campaign: 1458

For this particular campaign maximum profit is generated when sam2 algorithm is used.

```
$ bash single_campaign_arbitrage_demo.sh
import the parameters setting from config file
single campaign arbitrage driver
test campaign: ['1458']
test algorithms:['const', 'rand', 'truth', 'lin', 'ortb', 'sam1', 'sam2']
payoff ratio:0.2
read in train data...
read in test data...
start to train/test
```

prop	alpha	algo	profit	cnvs	bids	imps	budget	cost	rratio	para	up
16	0.00100	const	-2.96	0	80000	740	339072.7	2958.0	0.2	5.0	0.0000
16	0.00100	rand	0.00	0	80000	0	339072.7	0.0	0.2	5.0	0.0000
16	0.00100	truth	44.72	3	80000	601	339072.7	11862.0	0.2	1.0	0.0000
16	0.00100	lin	-93.89	13	59404	10843	339072.7	339097.0	0.2	75.0	0.0000
16	0.00100	ortb	-131.61	11	55468	12556	339072.7	339091.0	0.2	290.0	0.0000
16	0.00100	sam1	45.77	6	80000	3148	339072.7	67405.0	0.2	25.0	0.0000
16	0.00100	sam1c	59.14	6	80000	1303	339072.7	54032.0	0.2	25.0	8.2433
16	0.00100	sam2	63.92	7	80000	4146	339072.7	68116.0	0.2	7.1	0.0000
16	0.00100	sam2c	60.63	7	80000	4146	339072.7	71408.9	0.2	7.1	0.7942

Campaign: 2259

For this particular campaign maximum profit is generated when sam1 algorithm is used. No particular algorithm generates better results for every data set.

```
$ python ../python/driver_single_campaign_demo.py 2259 0.2
import the parameters setting from config file
single campaign arbitrage driver
test campaign: ['2259']
test algorithms:['const', 'rand', 'truth', 'lin', 'ortb', 'sam1', 'sam2']
payoff ratio:0.2
read in train data...
read in test data...
start to train/test
```

prop	alpha	algo	profit	cnvs	bids	imps	budget	cost	rratio	para	up
16	0.00100	const	35.35	5	80000	16187	440672.6	212350.0	0.2	23.0	0.0000
16	0.00100	rand	-0.40	0	80000	197	440672.6	396.0	0.2	5.0	0.0000
16	0.00100	truth	65.32	4	80000	8920	440672.6	132844.0	0.2	1.0	0.0000
16	0.00100	lin	89.55	8	80000	13883	440672.6	306771.0	0.2	30.0	0.0000
16	0.00100	ortb	-2.69	8	80000	18565	440672.6	399011.0	0.2	110.0	0.0000
16	0.00100	sam1	111.34	8	80000	13322	440672.6	284979.0	0.2	29.0	0.0000
16	0.00100	sam1c	52.70	7	80000	12629	440672.6	294076.4	0.2	29.0	1.9456
16	0.00100	sam2	27.61	6	80000	14908	440672.6	269629.0	0.2	4.3	0.0000
16	0.00100	sam2c	23.38	6	80000	14079	440672.6	273862.2	0.2	4.3	1.6672

4.2. Multiple Campaign Portfolio Results

This test includes 3 campaigns with id: 1458, 2259 and 2261. The budget proportion is 1/16 and pay-off ratio is 0.2. We have compared profits for 2 alpha values: 0.1 and 2.

```
$ bash multiple_campaign_arbitrage_demo.sh
import the parameters setting from config file
multiple campaign arbitrage driver
test campaign: ['1458', '2259', '2261']
test algorithms:['const', 'rand', 'truth', 'lin', 'ortb', 'sam1', 'sam2']
payoff ratio:0.2
read in train data...
read in test data...
average campaign original cost test: 6567701.0
test volume: 80000
prop    alpha    algo    profit  cnvs     bids     imps      budget        cost  rratio    para      up
  16  0.10000   const     -4.26     0    80000     1780    410481.3      4264.0     0.2     5.0  0.0000
  16  0.10000    rand    -39.98     1    80000    10304    410481.3    106134.0     0.2    29.0  0.0000
  16  0.10000   truth     44.72     3    80000      601    410481.3     11862.0     0.2     1.0  0.0000
  16  0.10000     lin   -184.14    12    40581    10983    410481.3    410491.0     0.2   105.0  0.0000
  16  0.10000    ortb   -184.16    12    60043    14348    410481.3    410511.0     0.2   290.0  0.0000
  16  0.10000    sam1     45.77     6    80000     3148    410481.3     67405.0     0.2    25.0  0.0000
  16  0.10000   sam1c     49.06     6    80000     2202    410481.3     64108.8     0.2    25.0  3.9500
  16  0.10000    sam2     66.64     7    80000     3905    410481.3     65391.0     0.2     6.5  0.0000
  16  0.10000   sam2c     50.21     6    80000     3100    410481.3     62967.4     0.2     6.5  2.6962
  16  2.00000   const     -3.79     0    80000     1552    410481.3      3787.0     0.2     5.0  0.0000
  16  2.00000    rand     -0.48     0    80000      310    410481.3       477.0     0.2     5.0  0.0000
  16  2.00000   truth     12.26     2    80000     4573    410481.3     72756.0     0.2     1.0  0.0000
  16  2.00000     lin   -184.14    12    40581    10983    410481.3    410491.0     0.2   105.0  0.0000
  16  2.00000    ortb   -184.16    12    60043    14348    410481.3    410511.0     0.2   290.0  0.0000
  16  2.00000    sam1     45.77     6    80000     3148    410481.3     67405.0     0.2    25.0  0.0000
  16  2.00000   sam1c     49.06     6    80000     2202    410481.3     64108.8     0.2    25.0  3.9500
  16  2.00000    sam2     45.56     7    80000     4758    410481.3     86471.0     0.2     6.5  0.0000
  16  2.00000   sam2c     28.05     6    80000     3937    410481.3     85128.1     0.2     6.5  2.6962
```

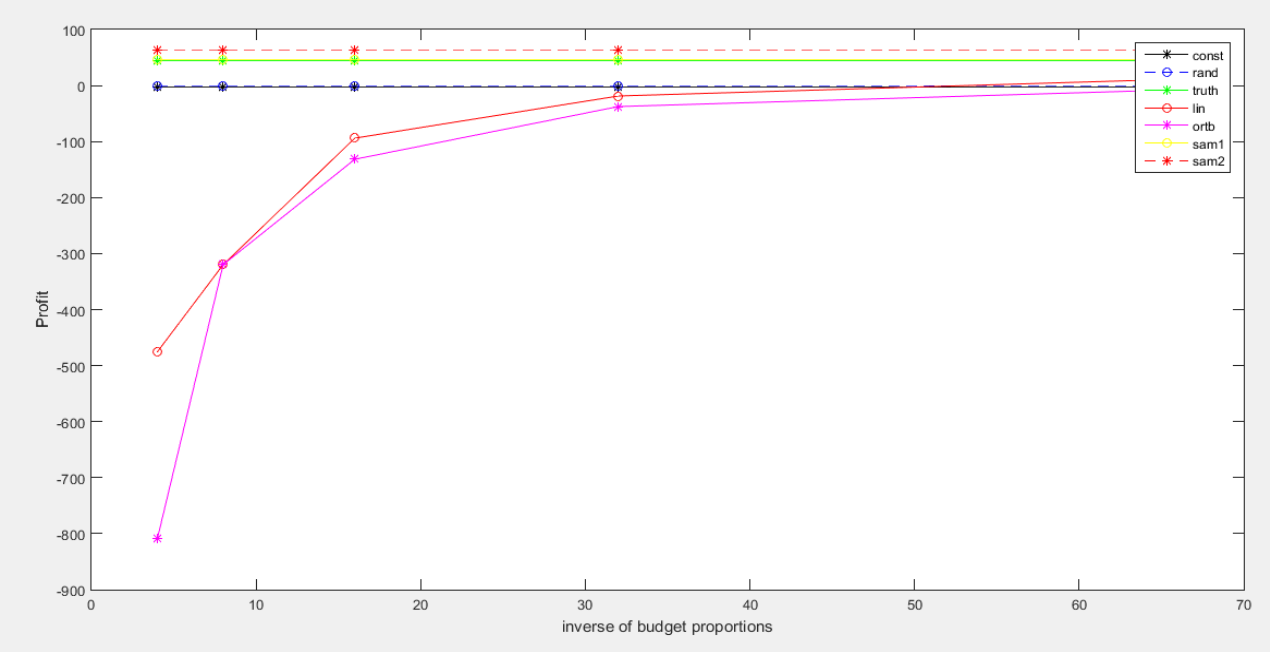
Table 4.1 Profit/Risk for different proportions of budget for SAM1 Algorithm

S.NO	BUDGET PROPORTIONS	PROFIT/RISK
1.	1/4	24.90
2.	1/8	24.90
3.	1/16	24.90
4.	1/32	24.90
5.	1/64	13.51

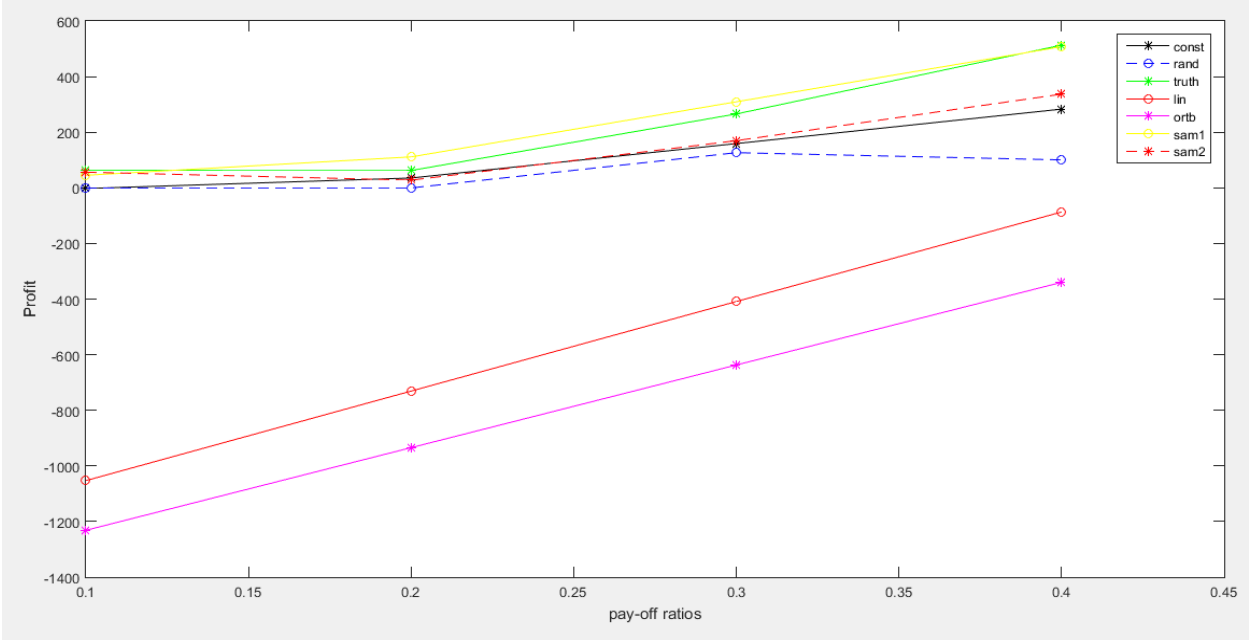
Table 4.2 Profit/Risk for different proportions of budget for SAM2 Algorithm

S.NO	BUDGET PROPORTIONS	PROFIT/RISK
1.	1/4	24.82
2.	1/8	24.82
3.	1/16	24.82
4.	1/32	26.52
5.	1/64	11.11

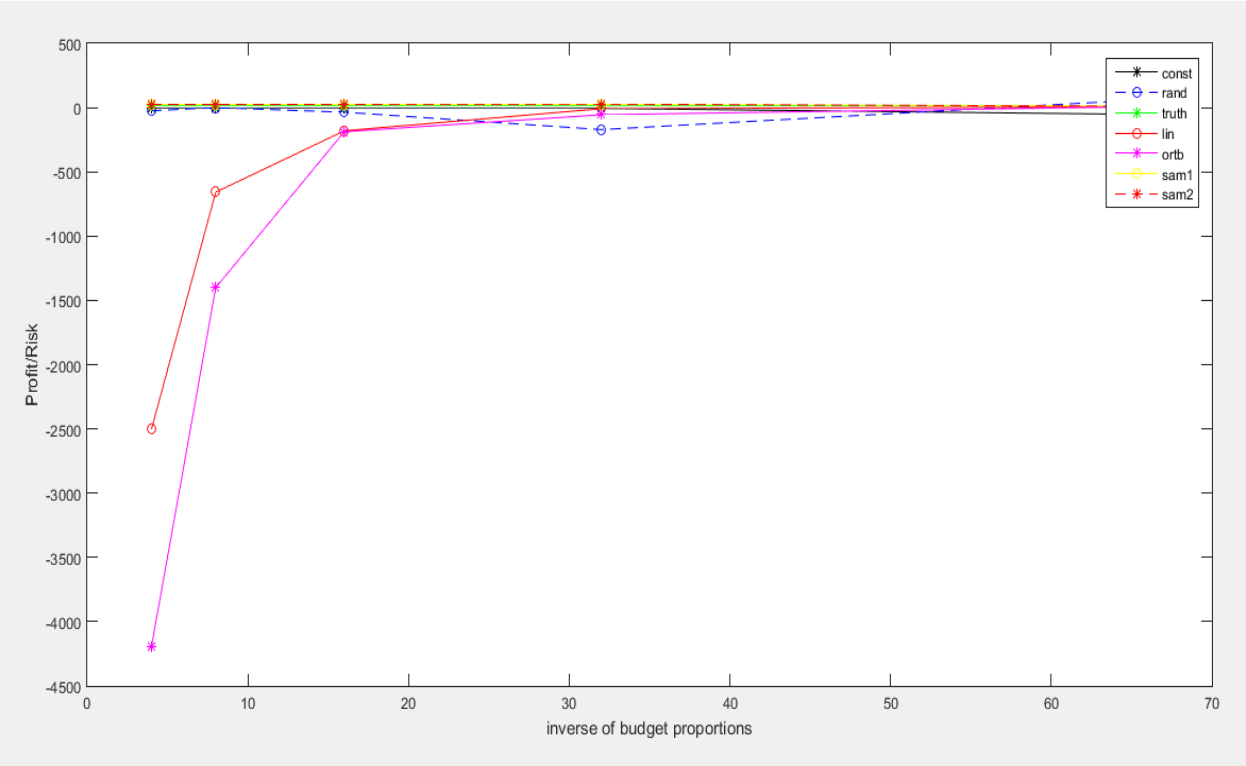
PROFIT FOR DIFFERENT ALGORITHMS AS A FUNCTION OF DIFFERENT BUDGET PROPORTIONS



PROFIT FOR DIFFERENT ALGORITHMS AS A FUNCTION OF DIFFERENT PAY-OFF RATIOS



PROFIT/RISK FOR DIFFERENT ALGORITHMS AS A FUNCTION OF DIFFERENT BUDGET PROPORTIONS



5. Conclusion and Future Scope

In our paper, we have demonstrated and tried to optimize the bidding strategy from the advertisers' side or from the DSP point of view in Real Time Bidding. For the basic single campaign approach, we have optimized the bidding strategy given the constraint of the auction budget by maximizing the profit by maximizing the effective costs per click or the conversion rate. In the more advanced approach that is dealing with multiple campaigns in addition to maximizing the utility we also try to allocate the budget in such a way so as to avoid exhausting the whole budget and investing a larger proportion in the campaign that will lead to a greater revenue. The whole process is based on Modern-Portfolio Theory which takes the advantage of multiple campaigns serves in the DSP. However, for a greater accuracy in the real world (in the process of real time bidding), the downside risk involved in real time bidding is calculated precisely by the Post-modern portfolio theory also proposed in this project.

The concept of Real time bidding came into picture in 2009 and is still an emerging technology which does not only have scope for advertisers and publishers but also

possess a lot of scope for improvements. Thus in the future work we would try to find algorithms that maximizes the profit much more and simultaneously reduces the risk as much as possible. We would also try to incorporate the feedback mechanism that is put in action as quickly as possible. This involves updating the bidding price in accordance to the risk found in the previous investment. We would try to make this task iterative and sequential in nature.

6.References

S. a. K. D. Adikari, "Real time bidding in online digital advertisement," in *International Conference on Design Science Research in Information Systems*, 2015.

D. Bloomfield, "High-speed ad traders profit by arbitraging your eyeballs," https://www.bloomberg.com/news/articles/2014-11-07/high-speed-ad-traders-profit-by-arbitraging-your-eyeballs, 2015.

R. P. a. S. V. McAfee, "An overview of practical exchange design," *Current Science(Bangalore),* vol. 109, no. 9, pp. 1056-1063, 2012.

S. Y. J. W. X. S. Weinan Zhang, "Real-time bidding benchmarking with ipinyou dataset," arXiv preprint, 2014.

L. P. Z. L. X. S. Hairen Liao, "iPinYou global rtb bidding algorithm competition dataset," in *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising. ACM*, 2014.

B. D. R. H. O. S. T. R. a. F. P. Claudia Perlich, "Bid optimizing and inventory scoring in targeted online advertising," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, Beijing, China, 2012.

B. O. A. D. a. W. L. Kuang-chih Lee, "Estimating conversion rate in display advertising from past erformance data," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, Beijing, China, 2012.

Received: 12 May 2018 | Revised: 11 July 2018 | Accepted: 21 August 2018

DOI: 10.1002/ctpp.201800058

Check for updates

Contributions to Plasma Physics

REVIEW ARTICLE

Theoretical model for the effect of dust grains on self-filamentation of a gaussian electromagnetic beam in a fully ionized plasma

Ruchi Sharma | **Suresh C. Sharma***

Department of Applied Physics, Delhi Technological University (DTU), Delhi, India

***Correspondence**
Suresh C. Sharma, Department of Applied Physics, Delhi Technological University (DTU), Shahbad Daulatpur, Bawana Road, Delhi 110042, India.
Email: suresh321sharma@gmail.com

A theoretical model for the effect of dust grains on the self-filamentation of a Gaussian electromagnetic beam propagating in a fully ionized plasma has been developed by employing the energy balance of the plasma constituents, perturbed electron and ion concentrations, and temperature. In this model, neutral atom ionization, re-integration and accumulation of electrons and ions, photoelectric emission of electrons from the surface of dust grains, as well as elastic and charging collisions have also been considered. The effective dielectric constant in the presence of dust grains has been constructed. The effect of temporal growth of dust grains on various plasma parameters for different values of the dust density has been explored. The variation of the beam width with the normalized channel of propagation has been observed for distinct dust densities and dust charge states. It is observed that the non-linearity induced by the effective dielectric constant in the presence of dust grains increases the self-filamentation of the beam, thus enhancing the effective critical power with the dust density. Some of the outcomes of our approach are in line with experimental observations. These outcomes may be useful for explaining space and laboratory plasma experiments as well as for future studies in complex plasmas.

KEYWORDS

critical power, dusty plasmas, fully ionized plasma, propagation of beam, self-filamentation

1 | INTRODUCTION

Several investigations[1–5] have been carried out on complex plasmas because of their exclusive characteristics. Especially, self-filamentation has been attracting a lot of of interest in recent years because of the non-linear interaction of an electromagnetic (EM) beam in complex plasmas. Non-linearity appears because of the non-linear radial distribution of the intensity of the beam in complex plasmas and modifies the electron concentration, electron temperature, and dielectric constant. Collisional, relativistic, and ponderomotive non-linearity are the leading factors for carrying out research in the field of plasma applications. Collisional non-linearity appears as a result of heating of the plasma components due to the dissipation of energy of the EM beam propagating through the plasma. The energy gained by the electrons during this process is lost by thermal conduction as well as by electron collisions with ions, dust, and neutral particles. In a fully ionized plasma, the former mechanism becomes predominant because of the higher degree of ionization as $\delta_0(\rho_0^2/l_m^2) << 1$, where $\delta_0(=2m_e/m_i)$ is the fraction of energy loss due to collisions, m_e and m_i are the electronic and ionic masses, ρ_0 is the initial width of the beam, and l_m is the mean free path of collisions of the electrons.

The importance of the subject has been confirmed by various researchers.[6–19] Sharma et al.[20] analysed the non-linear self-modulation of an EM beam in a dusty plasma. They measured the non-linear current density and change in the modulation

 © 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

index in the presence of dust grains in the plasma and found that the dust grains reduce the self-filamentation of microwaves and decrease their amplitude as compared to a dust-free plasma.

Mishra et al.[21] considered the propagation of EM waves in a complex plasma and observed the behaviour of the beam for the cases of convergence, oscillatory divergence, and steady divergence.

Sodha et al.[22] observed the variation in the various parameters of the plasma with applied electric field in the presence of dust grains.

Verheest and Meuris[23] developed the non-linear Schrödinger wave equation and observed the transformation in dust charge due to the non-linear interaction of the EM wave in the plasma.

Prasad and Tripathi[24] reported the self-filamentation characteristics of an EM beam in a weakly ionized plasma in their model.

Sambandan et al.[25] developed a theoretical model for the non-linear propagation of an EM beam in a dusty plasma and found that the EM beam gets self-focused for powers greater than a threshold value.

Sodha et al.[26] developed a theoretical model for the self- and cross-filamentation of an EM beam in a fully ionized collisional magneto-plasma. They looked at the heating of electrons through the plasma due to the beam and energy decay via collisions and thermal conduction. In addition, they observed the self-/cross-filamentation of the beam for higher and lower values of the magnetic field due to ionic thermal conduction.

Sodha et al.[27] investigated the self-focusing of a laser beam in a fully ionized plasma due to non-linearity, which appears as a result of energy loss by thermal conduction. In this case, they studied the non-linear dielectric constant for oscillatory and self-focused modes of the beam. They predicted an enhancement of the axial intensity by a factor of 25 for a 10^{10} W laser over a length of 0.6 cm.

Sodha et al.[28] investigated dust-particle charging in an illuminated open complex plasma. They included neutral atom ionization, ion–electron recombination, electron/ion accretion and electron emission by dust grains, and binary collision between different constituents of the complex plasma. They illustrated the parametric study of Cs, LaB_6, and CeO_2 in the daytime ionosphere at an altitude of 150 km assuming that the radius of dust cloud is greater than the mean free path of accretion of the dust grains.

Sodha et al.[7] also presented a model for thee evaluation of dust charge fluctuation in a complex plasma considering cases such as photoelectric emission, thermionic emission, and no emission of electrons from the dust grain surface.

Jia et al.[29] studied the modified dielectric constant for an EM wave propagating in fully ionized plasma using the Boltzmann's distribution law. They considered both charging and collisional effects of the dust grains and studied the effect of larger dust radii and charge on the propagation of the EM wave in fully ionized and weakly ionized plasmas.

Dan et al.[30] investigated the transmission properties of an EM wave propagating in a fully ionized plasma with and without dust grains. They considered the Fokker-Planck-Landau (FPL) and Bernstein-Greene-Kruskal (BGK) models and compared the results with those of a weakly ionized plasma using the matrix method. They observed that the dust radii and density affect the reflection and transmission coefficients of fully ionized dusty plasmas in the laboratory.

Stoffels et al.[31] demonstrated laser–particle interaction in an Ar/CCl_2F_2 RF plasma. In this case, they monitored the white light emitted during this process as a function of time and wavelength, and analysed it with the spectra of blackbody curves. They obtained a spectral temperature of ~3500 K for various dusty plasma states and ascribed it to the decomposition temperature of the particulate material.

Li et al.[32] studied the effect of the charge and size of dust particles on the propagation of EM waves in a dusty plasma using a power-law dust size distribution and observed the variation in the attenuation coefficient with dust size distribution.

Zharova et al.[33] developed an electrodynamic model for laser–particulate interaction in an ionized gas cluster. They studied the self-filamentation of laser radiation using the Schrodinger wave envelope and observed that for laser powers exceeding a critical power, the wave field met with self-compression.

In this paper, we study the role of dust grains in the self-filamentation of a highly intense Gaussian EM beam in a fully ionized complex plasma. Electrons are heated non-linearly when a Gaussian beam passes through complex plasma; i.e. bonding of dust particles with the electrons is maximum along the axial direction of beam because of the maximum temperature of electrons in the paraxial region. This non-linear heating causes the variation in electron concentration around the axis of the beam even when there is an ambipolar diffusion of the plasma. This inhomogeneous distribution of electron density creates a channel which guides the beam propagation. The energy and momentum acquired by the electrons during the propagation of the EM beam through the complex plasma are balanced by the loss of energy via collisions and thermal conduction. In a fully ionized complex plasma, the latter effect dominates over the first. In the present analysis of the non-linear interaction of an EM beam in fully ionized complex plasma, the following assumptions were made:

1. The dust grains are spherical and closely packed.
2. The amplitude of the EM beam has a Gaussian profile.

3. The propagation of the Gaussian beam follows the paraxial ray approximation, i.e. all the related parameters can be expanded in the surrounding of the axis of beam ($\rho = 0$) up to the term ρ^2/ρ_0^2 except the coefficients of thermal conduction, diffusion, and mobility, which adapt the behaviour of the axial temperature of the electrons and ions.
4. Heating of the electrons is non-linear ohmic because of the perturbed electron concentration and temperature due to the non-linear irradiance of the EM beam. Ohmic heating of ions may be ignored because of their heavier masses.
5. We assume the ionization and re-integration of the electrons and ions.
6. Collisions of electrons with the constituent particles are elastic and sticking.
7. Ions and electrons accumulate on the dust particle surfaces, and the emission of electrons from the surface of dust grains is photoelectric.
8. In the steady state, there is number and energy balance of the electrons and ions.
9. The complex plasma is charge-neutral.
10. Mobility and diffusion of the dust grains are zero because of their heavy mass.
11. Space charge is present because of the inhomogeneous electron temperature.
12. In the steady state, there is balance of the gradient of electron/ion pressure by force due to the space charge.
13. The change in electron and ion concentration is the same in the steady state because of the space charge.

For the photoelectric emission of electrons from the surface of dust particles, the density and temperature of electrons should be high enough and the irradiance of light should be low enough to keep the particle negatively charged as in case of Saturn's rings. Thus the calculations were carried out by processing the data corresponding to the metallic dust of high work function of 7.8 eV (stainless steel) in the near-space region of atmosphere where photoemission occurs on account of Lyman alpha spectral lines (1215.7 Å) of extreme ultraviolet (EUV) radiation.[34]

The remainder of the paper is structured as follows: In Section 2, we formulate the problem using a basic set of equations. We construct the effective concentration and temperature of electrons and ions using the energy balance to give the effective dielectric constant in the presence of dust grains. In Section 3, the wave equation is solved for the propagation of the Gaussian EM beam in the fully ionized complex plasma in terms of the variation of the beam width parameter f with the normalized channel of propagation ξ. In Section 4, the effective critical power of self-filamentation is derived using the initial boundary conditions. Outcomes of the present work are discussed in Section 5, and, finally, the conclusions are given in Section 6.

2 | THEORETICAL MODEL OF INTERACTION TECHNIQUES IN FULLY IONIZED COMPLEX PLASMAS

We assume a three-species plasma composed of ions, electrons, and dust grains with charge $-Z_d e$ and radius r_d, where Z_d is the dust charge state and e is the electronic charge. The dynamics of fully ionized complex plasma include re-integration of electrons and ions, ionization, neutralization of the ions, accumulation of electrons on the dust grain surfaces, ohmic heating, and thermal conduction.

The following equations explain the dynamics of the complex plasma:

2.1 | Ion and neutral-atom conservation

$$n_i + n_0 = n_{i0} + n_{00}, \tag{1}$$

where n_i, n_{i0}, and n_0, n_{00} are the densities of the ions and neutral atoms with and without the dust grains, respectively.

2.2 | Dust grain charging

The charge on dust grains is controlled by the accumulation of electrons and ions and emission of electrons from dust grain surfaces; thus

$$\frac{dZ_d}{d\tau} = n_{ic} + n_{ph} - \Upsilon n_{ec}, \tag{2}$$

where $n_{ic} = \pi r_d^2 (8k_B T_i/\pi m_i)^{1/2} n_i[1 - Z_d\alpha_i]$ is the ion accumulation current to the dust grain surface and $n_{ec} = \pi r_d^2 (8k_B T_e/\pi m_e)^{1/2} n_e \exp[Z_d\alpha_e]$ is the electron accumulation current to dust grain surface, and $n_{ph} = \pi r_d^2 n_p$ is the photocurrent emitted from the surface of negatively charged dust grains, where

$$n_p = \chi(\nu)\Lambda(\nu) = (4\pi m_e k_B^2 T^2/h^3)\mu(\nu)\Lambda(\nu)\Phi_{ph}(\eta),$$

$\Phi_{ph}(\eta)$ is tabulated integral (Fowler 1936)[49] defined by

$$\Phi_{ph}(\eta) = \int_0^{\exp(\eta)} \frac{\ln(1+\Omega)}{\Omega} d\Omega,$$
$$\eta = (h\nu - \phi_0)/k_B T_d.$$

n_p is rate of photoelectric emission per unit area from a neutral or negatively charged dust surface, $\chi(\nu)$ is the photoelectric efficiency, $\Lambda(\nu)$ is the number of photons incident per unit area per unit time on a surface normal to the direction of incidence, h is the Planck' constant, ν is the frequency of the radiation, ϕ_0 is work the function of dust, T_d is temperature of dust particle, $\mu(\nu)$ is the probability of absorption of a photon by an electron hitting the surface from inside.[34]

$\alpha_i = (e^2/r_d k_B T_i)$, $\alpha_e = (e^2/r_d k_B T_e)$, where r_d is the radius of dust grain, Υ is the coefficient of sticking of electrons on surface of a dust grain, k_B is the Boltzmann constant, T_i and T_e are the ion and electron temperature, m_i and m_e are the ionic and electronic masses, respectively.

2.3 | Dynamics of electron and ions

$$\frac{dn_e}{d\tau} = \beta_i n_0 - \alpha_r n_e n_i - n_d(\Upsilon n_{ec} - n_{ph}) - \nabla\Phi_p \tag{3}$$

and

$$\frac{dn_i}{d\tau} = \beta_i n_0 - \alpha_r n_e n_i - n_d n_{ic} - \nabla\Phi_p, \tag{4}$$

where β_i is the ionization coefficient, $\alpha_r = (\alpha_{r0}(300/T_e)^{k'})$ [cm^3/s] is the re-integration coefficient of ions and electrons, k' and α_{r0} are non-variables, and n_d is the number density of dust.

Equations 3 and 4 represent the growth rate of electron and ion densities in the complex plasma, respectively. The first term on the RHS of 3 and 4 related to the rate of gain of electron and ion densities, respectively, due to neutral atom ionization. The succeeding two terms in both equations refer to decay rate of electrons and ions on account of re-integration and the net accumulation of electrons and ions on the dust grain surface. The last term $\nabla\Phi_p$ represents the density decay due to ambipolar diffusion in both the equations, obtained by solving the momentum equations as shown in Appendix A.

2.4 | Neutrality of charge

Neutrality of charge may be shown by the following equation:

$$Z_d n_d + n_i - n_e = 0. \tag{5}$$

2.5 | Energy balance equations of electrons and ions

$$\frac{d}{d\tau}\left(\frac{3}{2}n_e k_B T_e\right) = \beta_i n_0 \varepsilon_e - \alpha_r n_e n_i \left(\frac{3}{2}k_B T_e\right) - n_d(\Upsilon n_{ec}\varepsilon_{ec} - n_{ph}\varepsilon_{ph}) + \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho k_{pe}\frac{dT_e}{d\rho}\right) + \frac{n_e e^2 \nu_{eff} U_0^2}{4m_e(\omega_d^2 + \nu_{eff}^2)}, \tag{6}$$

and

$$\frac{d}{d\tau}\left(\frac{3}{2}n_i k_B T_i\right) = \beta_i n_0 \varepsilon_i - \alpha_r n_e n_i \left(\frac{3}{2}k_B T_i\right) - n_d n_{ic}\varepsilon_{ic} + \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho k_{pi}\frac{dT_i}{d\rho}\right), \tag{7}$$

where ε_e and ε_i refer to the average energy of electrons and ions due to neutral atom ionization;

$\varepsilon_{ec}(=2k_B T_e - Z_d(e^2/r_d))$ and $\varepsilon_{ic}(=(2 - Z_d\alpha_i)/(1 - Z_d\alpha_i))$ represent the average energy of electrons and ions accumulating on the dust grain surfaces, respectively,

$$\varepsilon_{ph} = \left(\frac{k_B T_d}{\Theta(\eta - \xi)}\int_0^{\infty} 2\xi \ln(1 + \exp(\eta - \xi)d\xi\right) - Z_d\alpha_d$$

is concerned with the energy of photoemission from the dust grain surfaces, where Θ $(=Z_d e/r_d)$ is the potential of charged dust grains, $\eta(=h\nu - \phi_0/k_B T_d)$, h is the Planck' constant, ϕ_0 is the work function of dust grains material, $\alpha_d = (e^2/r_d k_B T_d)$, ξ is the normalized channel of propagation, and T_d is the dust temperature supposed to be non-variable on account of large heat capacity.

$k_{ph'}$ is the coefficient of thermal conductivity of electrons and ions of the fully ionized complex plasma given by Spitzer[35], and the subscript h' refers to electrons and ions, respectively.

$$k_{ph'} = \alpha {T_{h'}}^{5/2},$$

and

$$\alpha = 19.6 \times 10^{-5} / \ln \left\{ \frac{3}{2h^3} (k_B^3 T_{h'}^3 / \pi n_{h'})^{1/2} \right\}.$$

$\nu_{eff} (= \nu_{ei} + \nu_{ed} + \nu_{en} + \nu_{sc})$ is the effective collision frequency of electrons,where $\nu_{ei} = \nu_{ei0} \left(\frac{n_e}{n_{e0}} \right) \left(\frac{T_e}{T_{e0}} \right)^{-\frac{3}{2}}$ is the rate of electron–ion collisions,[36] and n_{e0} and T_{e0} refer to unperturbed density and temperature of the electrons, respectively. $\nu_{ei0} = \left(\frac{5.5 n_{e0}}{T_{e0}^{3/2}} \right) \ln \left(\frac{220 T_{e0}}{n_{e0}^{1/3}} \right)$ is the unperturbed rate of electron–ion collisions, $\nu_{ed} = \nu_{ed0} \left(\frac{T_e}{T_{e0}} \right)^{-\frac{3}{2}} Z_d^2$ is the rate of electron–dust collisions,[22] $\nu_{ed0} = (2.9 \times 10^{-6}) n_{d0} T_{e0}^{-3/2} \ln \Lambda_e$ is the rate of unperturbed electron–dust collisions [$\ln \Lambda_e$(coulomb logarithm) = 10–13 for fully ionized plasma], $\nu_{en} = \nu_{en0} \left(\frac{n_0}{n_{00}} \right) \left(\frac{T_e}{T_{e0}} \right)^{\frac{1}{2}}$ is the rate of electron–neutral particles collisions,[36] $\nu_{en0} = (8.3 \times 10^5) \sigma_0 n_{00} T_{e0}^{1/2}$ is the rate of unperturbed electron–neutral particles collisions, where $n_{00}$is unperturbed number density for neutral particles, $\sigma_0 (= 6.6 \times 10^{-15} [(T_e/4) - 0.1/\{1 + (T_e/4)^{1.6}\}])$ is the cross-sectional area of collision between electrons and neutral particles averaged over a Maxwellian electron distribution, where T_e is in eV and σ_0 is in cm^2. This indicates that $\sigma_0$increases to a threshold value of electron temperature or energy and then starts to decrease and $\nu_{sc} = n_d (n_{ec}/n_e)$ is the average count of sticking collisions of electrons with dust grains.

The initial two terms of 6 and 7 correspond to the final average energy gain of electrons and ions because of neutral atom ionization and re-integration of electrons and ions in the complex plasma. The succeeding term in both the equations refers to the final energy loss of electrons and ions because of their accumulation and photoemission of electrons from the surface of dust grains. The fourth term in both the equations represents the final gain in energy density of electrons and ions due to thermal conduction. The last term in 6 corresponds to the gain in energy density of the electrons from the field because of ohmic heating. Ohmic heating of ions can be neglected because of their heavier mass. The term corresponding to the power loss per unit volume due to binary elastic collisions between the components can be neglected in a fully ionized complex plasma because

$$R' = \text{Loss of energy via collisions/loss of energy via thermal conduction.}$$

According to Prasad and Tripathi,[24] in a fully ionized plasma, $R' \ll 1$, i.e. the rate of gain in energy due to electron–ion collisions is less than the work done by the electrons against the thermal forces. The direction of the current flow against the gradient of electron temperature will define the positive or negative sign of latter term.

A comparison of the term R' shows that the latter mechanism is more predominant in a fully ionized complex plasma because of the larger mean free path of the electrons than the Debye length; hence the former can be neglected.

2.6 | Estimation of the ionization coefficient (β_i)

The temperature of the dust grains and neutral atoms is supposed to be constant because of the sufficient energy exchange between them due to their large thermal expansion. The coefficient of ionization β_i can be estimated by executing the dynamics of electrons in a dust-free environment using the initial condition as

$$\beta_i n_{00} - \alpha_r n_{e0} n_{i0} = 0, \tag{8}$$

$$\beta_i n_{00} = \alpha_r n_{e0} n_{i0} = \alpha_r n_{e0}^2. \tag{9}$$

2.7 | Estimation of mean energy of electrons and ions

The mean energies of electrons and ions appearing on account of ionization can be estimated by applying the initial conditions in the energy balance equations for a dust-free plasma; thus

$$\varepsilon_e = \left(\frac{3}{2} k_B T_{e0} \right), \tag{10}$$

and

$$\varepsilon_i = \left(\frac{3}{2} k_B T_{i0} \right). \tag{11}$$

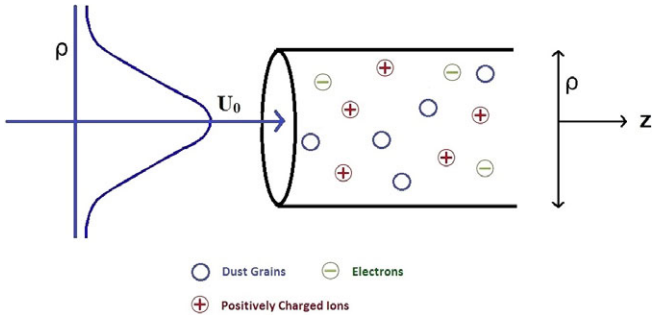


FIGURE 1 Schematic diagram of Gaussian irradiance distribution in a complex cylindrical plasma

2.8 | Paraxial ray approximation

Now we consider an EM beam having Gaussian irradiance distribution along its wave front. It is convenient to analyse this in the cylindrical coordinates system, with the propagation direction assumed to be the z-axis (cf. Figure 1).

The distribution of intensity of the beam at $z=0$ is given by

$$U_0^2 = U_{00}^2 \mathrm{e}^{(-\rho^2/\rho_0^2)}, \tag{12}$$

where U_0 is the amplitude of electric vector of the Gaussian beam at a distance ρ from axis, U_{00} is the amplitude of the beam along the axis at $\rho=0$, ρ_0 is the width of the beam at which amplitude of the beam is $1/e$ of its value on the axis, and ρ is the radial part of cylindrical coordinate system.

For a finite value of z, the intensity distribution under the paraxial theory is taken as

$$U_0^2 = \frac{U_{00}^2}{f^2} e^{(-\rho^2/\rho_0^2 f^2)}, \tag{13}$$

where f is the beam width, which depends on the propagation distance z, and $\rho_0 f$ is the beam spot size.

As is clear from the above expression, almost all of the intensity of the beam lies in the vicinity of the central axis of the beam. Therefore, the small-angle approximation can be adequately used to analyse the Gaussian beam in the paraxial ray approximation (i.e. $\rho/\rho_0 f \ll 1$). Related to this concept, all the parameters (electron/ion temperature, electron concentration, dielectric function, eikonal function, and irradiance profile of the beam) are expanded in the surrounding of central axis ($\rho=0$) only up to the term ρ^2/ρ_0^2 as

$$P(\rho, z) = P_{\mathrm{a}}(z) - (\rho^2/\rho_0^2) P_{\rho}(z), \tag{14}$$

where P_{a} and P_{ρ} are the components of the variables represented by $P(\rho,z)$ along the axis and radius, respectively.

The Gaussian profile of the beam remains unaffected during the propagation of the EM beam in a fully ionized complex plasma.

Hence, by inserting the relevant parameters in the dynamic Equations 2–4, 6, and 7, and comparing the coefficient ρ^0 and ρ^2/ρ_0^2 terms, we get the following.

2.9 | Charging of dust grains

$$\frac{dZ_{\mathrm{da}}}{d\tau} = n_{\mathrm{ica}} + n_{\mathrm{pha}} - \Upsilon n_{\mathrm{eca}}, \tag{2a}$$

and

$$\frac{dZ_{d\rho}}{d\tau} = n_{\mathrm{ic}\rho} + n_{\mathrm{ph}\rho} - \Upsilon n_{\mathrm{ec}\rho} \tag{2b}$$

2.10 | Dynamics of electron and ions

$$\frac{dn_{\mathrm{ea}}}{d\tau} = \beta_{\mathrm{ia}} n_0 - \alpha_{\mathrm{ra}} n_{\mathrm{ea}} n_{\mathrm{ia}} - n_{\mathrm{d}}(\Upsilon n_{\mathrm{eca}} - n_{\mathrm{pha}}) - \frac{(2/\rho_0^2)}{n_{\mathrm{ia}} + c_2 n_{\mathrm{ea}}} [D_{1\mathrm{i}}(c_1 n_{\mathrm{ia}} n_{\mathrm{e}\rho} + c_2 n_{\mathrm{ea}} n_{\mathrm{i}\rho}) + D_{2\mathrm{i}} n_{\mathrm{ea}} n_{\mathrm{ia}} (c_2 T_{\mathrm{ia}}^{-1} T_{\mathrm{i}\rho} + c_3 T_{\mathrm{ea}}^{-1} T_{e\rho})], \tag{3a}$$

and

$$\frac{dn_{e\rho}}{d\tau} = \beta_{i\rho}n_0 - \alpha_{r\rho}n_{ea}n_{ia} - n_d(n_{ec\rho} - n_{pha}) - \frac{(6/\rho_0^2)}{n_{ia}+c_2n_{ea}}\begin{bmatrix} D_{1i}(c_1+c_2)n_{i\rho}n_{e\rho} - \frac{(c_1n_{ia}n_{e\rho}+c_2n_{ea}n_{i\rho})(n_{i\rho}+c_2n_{e\rho})}{n_{ia}+c_2n_{ea}} \\ +D_{2i}[(n_{ea}n_{i\rho}+n_{ia}n_{e\rho})(c_2T_{ia}^{-1}T_{i\rho}+c_3T_{ea}^{-1}T_{e\rho}) \\ -n_{ea}n_{ia}(c_2T_{ia}^{-2}T_{i'\rho}^2+c_3T_{ea}^{-2}T_{e\rho}^2) \\ -\frac{n_{na}n_{ia}(c_2T_{ia}^{-1}T_{i\rho}+c_3T_{ea}^{-1}T_{e\rho})(n_{i\rho}+c_2n_{e\rho})}{n_{ia}+c_2n_{ea}} \end{bmatrix}. \tag{3b}$$

$$\frac{dn_{ia}}{d\tau} = \beta_{ia}n_0 - \alpha_{ra}n_{ea}n_{ia} - n_dn_{ica} - \frac{(2/\rho_0^2)}{n_{ia}+c_2n_{ea}}[D_{1i}(c_1n_{ia}n_{e\rho}+c_2n_{ea}n_{i\rho}) + D_{2i}n_{ea}n_{ia}(c_2T_{ia}^{-1}T_{i\rho}+c_3T_{ea}^{-1}T_{e\rho})], \tag{4a}$$

and

$$\frac{dn_{i\rho}}{d\tau} = \beta_{i\rho}n_0 - \alpha_{r\rho}n_{ea}n_{ia} - n_dn_{ic\rho} - \frac{(6/\rho_0^2)}{n_{ia}+c_2n_{ea}}\begin{bmatrix} D_{1i}(c_1+c_2)n_{i\rho}n_{e\rho} - \frac{(c_1n_{ia}n_{e\rho}+c_2n_{ea}n_{i\rho})(n_{i\rho}+c_2n_{e\rho})}{n_{i'a}+c_2n_{e'a}} \\ +D_{2i}[(n_{ea}n_{i\rho}+n_{ia}n_{e\rho})(c_2T_{ia}^{-1}T_{i\rho}+c_3T_{ea}^{-1}T_{e\rho}) \\ -n_{ea}n_{ia}(c_2T_{ia}^{-2}T_{i\rho}^2+c_3T_{ea}^{-2}T_{e\rho}^2) \\ -\frac{n_{ea}n_{ia}(c_2T_{ia}^{-1}T_{i\rho}+c_3T_{ea}^{-1}T_{e\rho})(n_{i\rho}+c_2n_{e\rho})}{n_{ia}+c_2n_{ea}} \end{bmatrix}. \tag{4b}$$

2.11 | Energy balance equations for electrons and ions

$$\frac{d}{d\tau}\left(\frac{3}{2}n_{ea}k_BT_{ea}\right) = \beta_{ia}n_0\varepsilon_e - \alpha_{ra}n_{ea}n_{ia}\left(\frac{3}{2}k_BT_{ea}\right) - n_d(\Upsilon n_{eca}\varepsilon_{eca} - n_{pha}\varepsilon_{pha}) - \frac{4k_{pe}}{\rho_0^2}T_{e\rho} + \left(\frac{3}{2}k_BT\right)\frac{\beta_0U_{00}^2}{f^2}n_{ea}\nu_{effa}\delta_{comp} \tag{6a1}$$

and

$$\frac{d}{d\tau}\left(\frac{3}{2}k_B(n_{ea}T_{e\rho}+n_{e\rho}T_{ea})\right) = \beta_{i\rho}n_0\varepsilon_e - \frac{3}{2}k_B(\alpha_{r\rho}n_{ea}n_{ia}T_{ea} + \alpha_{ra}(n_{ea}n_{ia}T_{e\rho}+n_{ea}n_{i\rho}T_{ea}+n_{e\rho}n_{ia}T_{ea})) - n_d(\Upsilon(n_{ec\rho}\varepsilon_{eca}+n_{eca}\varepsilon_{ec\rho}) - (n_{pha}\varepsilon_{ph\rho}+n_{ph\rho}\varepsilon_{pha})) + \left(\frac{3}{2}\delta_{comp}k_BT\right)\frac{\beta_0U_{00}^2}{f^2}\left(\frac{n_{ea}\nu_{effa}}{f^2} + (n_{e\rho}\ \nu_{effa} + n_{ea}\ \nu_{eff\rho})\right) \tag{6b1}$$

Substituting $\frac{dn_{ea}}{d\tau}$ and $\frac{dn_{e\rho}}{d\tau}$ from 3a and 3b, we obtain

$$\frac{3}{2}n_{ea}k_B\left(\frac{dT_{ea}}{d\tau}\right) = \begin{bmatrix} \beta_{ia}n_0\left(\varepsilon_e - \frac{3}{2}k_BT_{ea}\right) - n_d\Upsilon n_{eca}\left(\varepsilon_{eca} - \frac{3}{2}k_BT_{ea}\right) - n_dn_{pha}\left(\varepsilon_{pha} - \frac{3}{2}k_BT_{ea}\right) \\ +\frac{3}{2}k_BT_{ea}\left(\frac{(2/\rho_0^2)}{n_{ia}+c_2n_{ea}}\right)[D_{2i}(c_1n_{ia}n_{e\rho}+c_2n_{ea}n_{i\rho}) \\ +D_{1i}n_{ea}n_{ia}(c_2T_{ia}^{-1}\ T_{i\rho}+c_3T_{ea}^{-1}\ T_{e\rho})] \\ -\frac{4k_{pe}}{\rho_0^2}T_{e\rho} + \left(\frac{3}{2}k_BT\right)\frac{\beta_0U_{00}^2}{f^2}n_{ea}\nu_{effa}\delta_{comp} \end{bmatrix}, \tag{6a2}$$

where $\beta_0 = \frac{e^2}{3m_e\delta_{comp}k_BT\omega^2}$, and $\delta_{comp}\left(=\nu_{sc}/\nu_{eff}\right)$[25,37] is the mean part of collisional energy loss of electron with the constituents of plasma; c_1, c_2, c_3, D_{1i}, and D_{2i} are defined in Appendix A; and

$$\frac{3}{2}n_{ea}k_B\left(\frac{dT_{e\rho}}{d\tau}\right) = \beta_{ia}n_0\left[\frac{3}{2}k_BT_{e\rho} - \frac{n_{e\rho}}{n_{ea}}\left(\varepsilon_e - \frac{3}{2}k_BT_{ea}\right)\right] + \beta_{i\rho}n_0\left(\varepsilon_e - \frac{3}{2}k_BT_{e\rho}\right) - n_d\Upsilon n_{eca}\left[\left(\varepsilon_{eca} - \frac{3}{2}k_BT_{ea}\right) + \left(\varepsilon_{ec\rho} - \frac{3}{2}k_BT_{e\rho}\right)\right] - n_dn_{ec\rho}\left(\varepsilon_{eca} - \frac{3}{2}k_BT_{ea}\right) + \frac{3}{2}k_B\left(T_{e\rho} + \frac{n_{e\rho}}{n_{ea}}T_{ea}\right)\left(\frac{(2/\rho_0^2)}{n_{ia}+c_2n_{ea}}\right)[D_{2i}(c_1n_{ia}n_{e\rho}+c_2n_{ea}n_{i\rho}) + D_{1i}n_{ea}n_{ia}(c_2T_{ia}^{-1}\ T_{i\rho}+c_3T_{ea}^{-1}\ T_{e\rho})]$$

$$-\frac{3}{2}k_{\rm B}T_{\rm ea}\frac{(6/\rho_0^2)}{n_{\rm ia}+c_2n_{\rm ea}}\begin{bmatrix} D_{2\rm i}(c_1+c_2)n_{\rm i\rho}n_{\rm e\rho}-\frac{(c_1n_{\rm ia}n_{\rm e\rho}+c_2n_{\rm ea}n_{\rm i\rho})(n_{\rm i\rho}+c_2n_{\rm e\rho})}{\rm n_{ia}+c_2n_{ea}} \\ +D_{1\rm i}[(n_{\rm ea}n_{\rm i\rho}+n_{\rm ia}n_{\rm e\rho})(c_2T_{\rm ia}^{-1}\ T_{\rm i\rho}+c_3T_{\rm ea}^{-1}\ T_{e\rho}) \\ -n_{\rm ea}n_{\rm ia}(c_2T_{\rm ia}^{-2}\ T_{\rm i\rho}^2+c_3T_{\rm ea}^{-2}\ T_{\rm e\rho}^2)-\frac{n_{\rm ea}n_{\rm ia}(c_2T_{\rm ia}^{-1}\ T_{\rm i\rho}+c_3T_{\rm ea}^{-1}\ T_{e\rho})(n_{\rm i\rho}+c_2n_{\rm e\rho})}{n_{\rm ia}+c_2n_{\rm ea}} \end{bmatrix}.$$

$$-\frac{4k_{\rm pe}}{\rho_0^2}\frac{n_{\rm e\rho}}{n_{\rm ea}}T_{e\rho}+\left(\frac{3}{2}\delta_{\rm comp}\,k_{\rm B}T\right)\frac{\beta_0U_{00}^2}{f^2}n_{\rm ea}\nu_{\rm effa}+\left(\frac{3}{2}\delta_{\rm comp}\ \ k_{\rm B}T\right)\frac{\beta_0U_{00}^2}{f^2}(n_{\rm ea}\nu_{\rm eff\rho}) \tag{6b2}$$

Similarly

$$\frac{d}{d\tau}\left(\frac{3}{2}n_{\rm ia}k_{\rm B}T_{\rm ia}\right)=\beta_{\rm ia}n_0\varepsilon_{\rm i}-\alpha_{\rm ra}n_{\rm ea}n_{\rm ia}\left(\frac{3}{2}k_{\rm B}T_{\rm ia}\right)-\Upsilon n_{\rm d}n_{\rm ica}\varepsilon_{\rm ica}-\frac{4k_{\rm pi}}{\rho_0^2}T_{\rm i\rho}, \tag{7a1}$$

and

$$\frac{d}{d\tau}\left(\frac{3}{2}k_{\rm B}(n_{\rm ia}T_{\rm i\rho}+n_{\rm i\rho}T_{\rm ia})\right)=\beta_{\rm i\rho}n_0\varepsilon_{\rm i}-\frac{3}{2}k_{\rm B}(\alpha_{\rm r\rho}n_{\rm ea}n_{\rm ia}T_{\rm ia}+\alpha_{\rm ra}(n_{\rm ea}n_{\rm ia}T_{\rm i\rho}+n_{\rm ea}n_{\rm i\rho}T_{\rm ia}+n_{\rm e\rho}n_{\rm ia}T_{\rm ia}))$$
$$-n_{\rm d}(\Upsilon(n_{\rm ic\rho}\ \ \varepsilon_{\rm ica}+n_{\rm ica}\ \ \varepsilon_{\rm ic\rho})) \tag{7b1}$$

Substituting $\frac{dn_{\rm ia}}{d\tau}$ and $\frac{dn_{\rm i\rho}}{d\tau}$ from 4a and 4b, we obtain

$$\frac{3}{2}n_{\rm ia}k_{\rm B}\left(\frac{dT_{\rm ia}}{d\tau}\right)=\begin{bmatrix} \beta_{\rm ia}n_0\left(\varepsilon_{\rm i}-\frac{3}{2}k_{\rm B}T_{\rm ia}\right)-n_{\rm d}n_{\rm ica}\left(\varepsilon_{\rm ica}-\frac{3}{2}k_{\rm B}T_{\rm ia}\right) \\ +\frac{3}{2}k_{\rm B}T_{\rm ia}\left(\frac{(2/\rho_0^2)}{n_{\rm ia}+c_2n_{\rm ea}}\right)[D_{1\rm i}(c_1n_{\rm ia}n_{\rm e\rho}+c_2n_{\rm ea}n_{\rm i\rho})+D_{2\rm i}n_{\rm ea}n_{\rm ia}(c_2T_{\rm ia}^{-1}\,T_{\rm i\rho}+c_3T_{\rm ea}^{-1}\,T_{\rm e\rho}) \\ -\frac{4k_{\rm pi}}{\rho_0^2}T_{\rm i\rho} \end{bmatrix}, \tag{7a2}$$

and

$$\frac{3}{2}n_{\rm ia}k_{\rm B}\left(\frac{dT_{\rm i\rho}}{d\tau}\right)=\beta_{\rm ia}n_0\left[\frac{3}{2}k_{\rm B}T_{\rm i\rho}-\frac{n_{\rm i\rho}}{n_{\rm ia}}\left(\varepsilon_{\rm ia}-\frac{3}{2}k_{\rm B}T_{\rm ia}\right)\right]+\beta_{\rm i\rho}n_0\left(\varepsilon_{\rm i\rho}-\frac{3}{2}k_{\rm B}T_{\rm ia}\right)$$
$$-n_{\rm d}n_{\rm ica}\left[\frac{n_{i\rho}}{n_{\rm ia}}\left(\varepsilon_{\rm ica}-\frac{3}{2}k_{\rm B}T_{\rm ia}\right)+\left(\varepsilon_{\rm ic\rho}-\frac{3}{2}k_{\rm B}T_{\rm i\rho}\right)\right]-n_{\rm d}n_{\rm ic\rho}\left(\varepsilon_{\rm ica}-\frac{3}{2}k_{\rm B}T_{\rm ir}\right)$$
$$+\frac{3}{2}k_{\rm B}\left(T_{\rm i\rho}+\frac{n_{\rm i\rho}}{n_{\rm ia}}T_{\rm ia}\right)\left(\frac{(2/\rho_0^2)}{n_{\rm ia}+c_2n_{\rm ea}}\right)[D_{2\rm i}(c_1n_{\rm ia}n_{\rm e\rho}+c_2n_{\rm ea}n_{\rm i\rho})+D_{1\rm i}n_{\rm ea}n_{\rm ia}(c_2T_{\rm ia}^{-1}\,T_{\rm i\rho}+c_3{\rm T}_{\rm ea}^{-1}\,T_{\rm e\rho})]$$
$$-\frac{3}{2}k_{\rm B}T_{\rm ia}\frac{(6/\rho_0^2)}{n_{\rm ia}+c_2n_{\rm ea}}\begin{bmatrix} D_{2\rm i}(c_1+c_2)n_{\rm i\rho}n_{\rm e\rho}-\frac{(c_1n_{\rm ia}n_{\rm e\rho}+c_2n_{\rm ea}n_{\rm i\rho})(n_{\rm i\rho}+c_2n_{\rm e\rho})}{\rm n_{ia}+c_2n_{ea}} \\ +D_{1\rm i}[(n_{\rm ea}n_{\rm i\rho}+n_{\rm ia}n_{\rm e\rho})(c_2T_{\rm ia}^{-1}\,T_{\rm i\rho}+c_3{\rm T}_{\rm ea}^{-1}\,T_{\rm e\rho}) \\ -n_{\rm ea}n_{\rm ia}(c_2T_{\rm ia}^{-2}T_{\rm i\rho}^2+c_3T_{\rm ea}^{-2}T_{\rm e\rho}^2)-\frac{n_{\rm ea}n_{\rm ia}(c_2T_{\rm ia}^{-1}\,T_{\rm i\rho}+c_3T_{\rm ea}^{-1}\,T_{e\rho})(n_{\rm i\rho}+c_2n_{\rm e\rho})}{n_{\rm ia}+c_2n_{\rm ea}} \end{bmatrix}$$
$$-\frac{4k_{\rm pi}}{\rho_0^2}\frac{n_{\rm i\rho}}{n_{\rm ia}}T_{\rm i\rho}. \tag{7b2}$$

The stationary solution (i.e. at $t\to\infty$) of 2–7(a2, b2) jointly gives the dependence of the different fully ionized complex plasma parameters on the axial part of radiation by choosing the axial and radial components of the relevant set of parameters.

2.12 | Effective dielectric constant

Using the above set of parameters, the non-linear effective dielectric constant may be calculated as follows:

$$\varepsilon_{\rm eff}=1-\left(\frac{\omega_0^2}{\omega_{\rm d}^2}\right)\frac{\omega_{\rm d}^2}{(\nu_{\rm eff}^2+\omega_{\rm d}^2)}\left(\frac{n_{\rm e}}{n_{\rm e0}}\right)=\varepsilon_{\rm a}-\left(\rho^2/\rho_0^2\right)\varepsilon_\rho. \tag{15}$$

Substituting $n_{\rm e}$ and $\nu_{\rm eff}$ corresponding to 14 and comparing the coefficient of ρ^0 and ρ^2/ρ_0^2 terms, we obtain

$$\varepsilon_{\rm a}=1-\left(\frac{\omega_0^2}{\omega_{\rm d}^2}\right)\left(1-\frac{\nu_{\rm effa}^2}{\omega_{\rm d}^2}\right)\left(\frac{n_{\rm ea}}{n_{\rm e0}}\right), \tag{15a}$$

and

$$\varepsilon_\rho=\left(\frac{\omega_0^2}{\omega_{\rm d}^2}\right)\left[\left(1-\frac{\nu_{\rm effa}^2}{\omega_{\rm d}^2}\right)\left(\frac{n_{e\rho}}{n_{\rm e0}}\right)-\left(\frac{n_{\rm ea}}{n_{\rm e0}}\right)\left(\frac{2\nu_{\rm effa}\nu_{\rm eff\rho}}{\omega_{\rm d}^2}\right)\right], \tag{15b}$$

where ν_{effa} and $\nu_{\mathrm{eff}\rho}$ are the axial and radial part of the effective collision frequencies, n_{ea} and $\mathrm{n_{e\rho}}$ are the axial and radial electron densities, ω_0 and ω_{d} are the electron plasma frequencies in the absence and presence of dust grains, respectively.

Using the axial and radial parts of electron densities and the effective collision frequencies by concurrent solution of 2–7(a2, b2), we get the axial and radial components of the effective dielectric constants (ϵ_{a} and ϵ_{ρ}), in terms of $\beta_0 U_{00}^2$.

3 | SELF-FILAMENTATION OF A GAUSSIAN ELECTROMAGNETIC BEAM IN FULLY IONIZED COMPLEX PLASMA

Consider a uniformly polarized Gaussian EM beam propagating along the z-axis in a fully ionized complex plasma with its polarized electric vector along the y-axis. In cylindrical coordinates, the electric field vector U for above-defined beam may be expressed as

$$U(z) = \hat{j} U_0(\rho, z) e^{i\omega_d \mathrm{t}}, \tag{16}$$

where $(U_0)_{z=0} = U_{00} e^{(-\rho^2/\rho_0^2)}$, and $\hat{\mathrm{j}}$ is the unit vector along the y-axis.

The corresponding wave equation is given by

$$\nabla^2 U - \nabla(\nabla \cdot U) + \frac{\epsilon_{\mathrm{eff}}(\rho, z)}{c^2} \frac{\partial^2 U}{\partial t^2} = 0. \tag{17}$$

Since the divergence of the electric vector of the transverse beam is zero, 17 reduces to

$$\nabla^2 \mathrm{U}_0 + (\omega_{\mathrm{d}}^2/\mathrm{c}^2)\epsilon_{\mathrm{eff}}(\rho, \mathrm{z})\mathrm{U}_0 = 0. \tag{18}$$

For a cylindrically symmetric beam (U_0 being independent of θ), from 18 we obtain

$$\frac{\partial^2 \overrightarrow{U_0}}{\partial \mathrm{z}^2} + \frac{1}{\rho} \frac{\partial \overrightarrow{U_0}}{\partial \rho} + \frac{\partial^2 \overrightarrow{U_0}}{\partial \rho^2} = \frac{\omega_{\mathrm{d}}^2 \epsilon_{\mathrm{eff}}(\rho, z)}{c^2} \overrightarrow{U}_0. \tag{19}$$

The solution of 19, according to Akhmanov et al.,[38] can be expressed as

$$U_0(\rho, z) = A_0(\rho, z) e^{-\mathrm{ikz}}, \tag{20}$$

$$A_0(\rho, z) = A_{00}(\rho, z) e^{-\mathrm{iks}(\rho, z)}, \tag{21}$$

where $\mathrm{A}_0(\rho,\mathrm{z})$ is the complex amplitude of the electric field, $k = (\omega_{\mathrm{d}}/c)\epsilon_{\mathrm{a}}^{1/2}$, A_{00} and s are real.

Following Sodha et al.,[39] substituting U_0 and A_0, respectively, from 20 and 21 in 19 and comparing the real and imaginary parts of consequent equations, we obtain

$$2\frac{\partial s}{\partial z} + \left(\frac{\partial s}{\partial \rho}\right)^2 = \frac{1}{\epsilon_{\mathrm{a}}} \Phi\left(\frac{1}{2} A_0 A_0^*\right) + \frac{1}{k^2 A_{00}} \left\{ \frac{\partial^2 A_{00}}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial A_{00}}{\partial \rho} \right\}, \tag{22}$$

and

$$\frac{\partial A_{00}^2}{\partial z} + \frac{\partial s}{\partial \rho} \frac{\partial A_{00}^2}{\partial \rho} + A_{00}^2 \left(\frac{\partial^2 s}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial s}{\partial \rho} \right) = 0, \tag{23}$$

where

$$\Phi\left(\frac{1}{2} A_0 A_0^*\right) = (\rho^2/\rho_0^2)\epsilon_{\rho}. \tag{23a}$$

Following earlier analysis,[38,40,41] the solution of A_{00}^2 may be written as

$$A_{00}^2 = \frac{U_{00}^2}{f^2} e^{-\frac{\rho^2}{\rho_0^2 f^2}}. \tag{24}$$

Substituting the expression of Φ and A_{00}^2 from 23a and 24, respectively, in 22, we obtain

$$2\frac{\partial s}{\partial z} + \left(\frac{\partial s}{\partial \rho}\right)^2 = \frac{1}{2} \frac{\epsilon_{\rho}}{\epsilon_{\mathrm{a}}} A_{00}^2 - \frac{2}{k^2 \rho_0^2 f^2} + \frac{\rho^2}{k^2 \rho_0^4 f^4}. \tag{25}$$

The eikonal function $s(\rho, z)$ in the paraxial ray approximation may be represented as

$$s = \frac{\rho^2}{2} \beta(z) + \Psi(z), \tag{26}$$

where $\Psi(\mathrm{z})$ is any function of z and $\beta(z) = \frac{1}{f} \frac{df}{dz}$.

The shape and power of the beam is preserved by the solution of A_{00}^2 when the beam propagates.

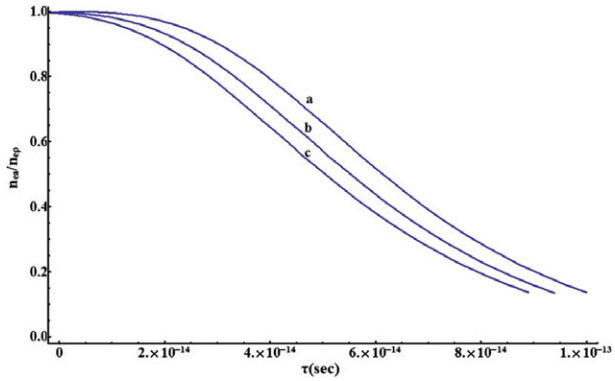


FIGURE 2 Variation of the number density of electrons with the parameter τ for different values of the number density of dust grains (n_d). The labels a, b, and c refer to $n_d = 2\times10^2$, 2×10^3, 2×10^4 cm^{-3}, respectively

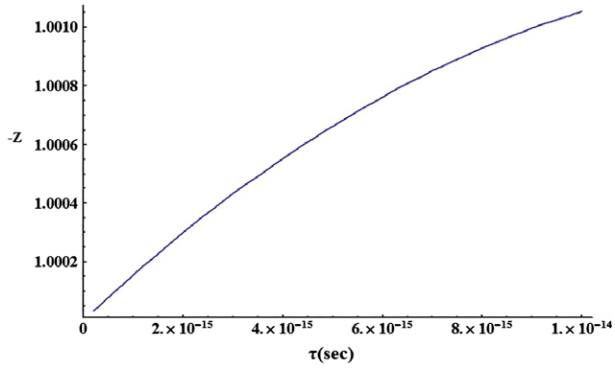


FIGURE 3 Dependence of the negative dust charge state on time τ. The other parameters are given in the text

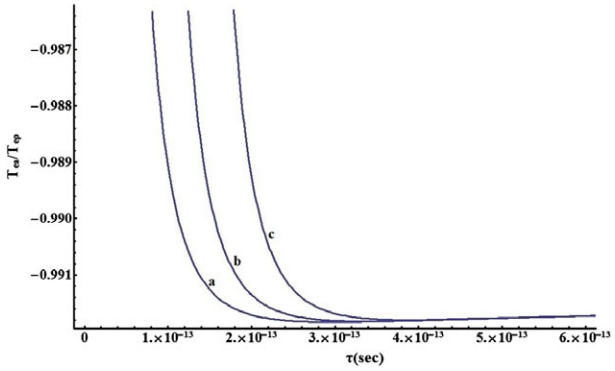


FIGURE 4 Variation of the electron temperature with the parameter τ for different values of number density of dust grains (n_d). The labels a, b, and c refer to $n_d = 2\times10^2$, 2×10^3, 2×10^4 cm^{-3}, respectively

Employing the expression for $\beta(z)$ in 26 and comparing the coefficient of ρ^2/ρ_0^2 of the resulting equation yields[32] the equation for beam width parameter of the form

$$\frac{d^2f}{dz^2} = \frac{1}{R_d^2}\frac{1}{f^3} - \frac{1}{2\rho_0^2}\frac{\varepsilon_\rho}{\varepsilon_a}\frac{U_{00}^2}{f^3} \tag{27}$$

or

$$\frac{d^2f}{d\xi^2} = -\frac{1}{R_{ns}^2}\frac{1}{f^3} + \frac{1}{f^3}, \tag{28}$$

where $R_{ns}\left(=\frac{2\varepsilon_a\rho_0^2}{\varepsilon_\rho U_{00}^2}\right)^{1/2}$ is the characteristic length of focusing, $\xi = \frac{z}{R_d}$, and $R_d = k\rho_0^2$.

The first term on the right-hand side in 28 represents the non-linear self-filamentation term, while the last term is from the divergence due to diffraction.

4 | EFFECTIVE CRITICAL POWER IN THE PRESENCE OF DUST GRAINS

From 28, the effective critical power P_{cr} (evaluated in Appendix B) of self-filamentation for $R_{ns} = R_d$ (for which $\frac{d^2f}{d\xi^2} = 0$, f = 1, for all values of ξ) can be written as

$$P_{cr} = \frac{c^3}{4\omega_d^2}\frac{\varepsilon_a^{1/2}}{\varepsilon_\rho}. \tag{29}$$

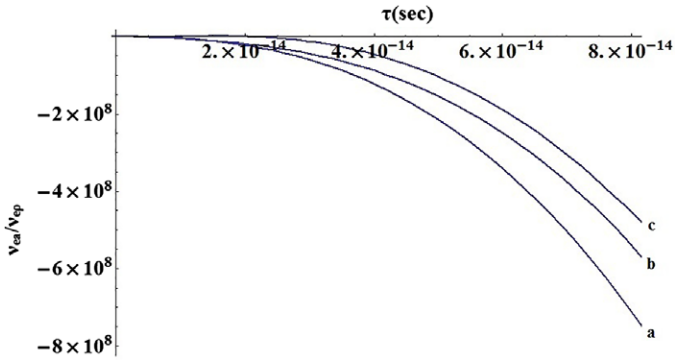


FIGURE 5 Variation of the effective collision frequency with time for different values of n_d. The labels a, b, and c refer to $n_d = 2\times10^2$, 2×10^3, 2×10^4 cm^{-3}, respectively

Thus, an incident EM plane wave front will travel independent of the beam spot size, resulting in linear waveguide propagation. The effective critical power comes out to be 7.1094×10^{12} (ergs s^{-1}) for the given set of parameters.

5 | COMPUTING OUTCOMES AND DISCUSSION

In this paper, we presented a theoretical model for the dynamics of temporal growth of dust grain in a fully ionized complex plasma and expressed the features of non-uniform correspondence of a Gaussian EM beam with fully ionized complex plasma having electrons, ions, and dust grains as the constituents. Diffusion, electron–ion/electron–dust collisions in plasma, electron/ion deposition on the dust grain surfaces, and thermal conduction are the specific processes that have been included in this theoretical model. In a fully ionized complex plasma, the energy loss of electrons via thermal conduction dominates the over the collisional effect, so we ignored energy loss by collision. The number density of constituents and energy balance equations were used to study the non-uniform dependence of the different variables on the dimensionless axial intensity of the beam; also, the computation was carried out to see the dependence of the dust charge, electron and ion density, their temperatures, and effective collision frequency on the time τ for different values of dust density by solving the coupled equations simultaneously. MATHEMATICA AND MATLAB software were employed to solve the equations numerically. Using these equations, ε_a and ε_ρ can be obtained by substituting the appropriate values. From the values of ε_a and ε_ρ, we can gain the beam width parameter as a function of the normalized channel of propagation $\xi(=z/R_d)$, satisfying the boundary conditions $f = 1$, $\frac{df}{d\xi} = 0$ at $\xi = 0$ for different values of beam intensities and dust densities. For the numerical solution of this analytical model, the curve representing the self-filamentation was obtained by plotting beam width parameter (f) as a function of normalized channel of propagation of the beam $\xi(=z/R_d)$ for a selected set of variables, picturing the region of propagation for self-filamentation.

The set of variables used in the calculations are as follows:

$\alpha_{r0} = 5.5\times10^{-7}$ cm^3/s, $n_{e0} = n_{i0} = 2.45\times10^6$ cm^{-3}, $n_0 = n_{00} = 0.9\times10^{12}$ cm^{-3}, $k' = 1.1$, $n_d = 10^2$ cm^{-3}, $\rho_0 = 5.5\times10^{-2}$ cm, $k_B = 1.38\times10^{-16}$ $ergsk^{-1}$, $T_{e0} = 1001$ K, $m_e = 9.1\times10^{-28}$g, $\nu_{e0} = 6.4\times10^3$ s^{-1}, $T_{i0} = 399$K, $m_i = 29\times1.6\times10^{-24}$g, $\nu_{i0} = 5\times10^2$ s^{-1}, $m_d = 39\times1.6\times10^{-24}$ g, $\omega_d = 1.1\times10^8$ s^{-1}, $e = 4.8\times10^{-10}$ statcoloumb, $T_d = 300$K, $r_d = 10^{-5}$ cm, $\sigma_0 = 1.5\times10^{-16}$ cm^2, $U_{00} = 10^5$ stat volt cm^{-1} and 6×10^5 stat volt cm^{-1}.

The variation of the electron number density, $n_{ea}/n_{e\rho}$, has been plotted against time τ [s] for different values of the dust grain number density n_d in Figure 2. The figure shows that $n_{ea}/n_{e\rho}$ decreases with time, and for larger values of n_d, $n_{ea}/n_{e\rho}$ decays faster because of the increasing availability of dust grains for the accumulation of electrons. Figure 3 displays the variation of dust charge state with time, which indicates that the negative dust charge state decreases with time on account of the decreasing number density of electrons with increased dust density. Figure 4 presents the dependence of electron temperature on the variable τ [s] for distinct values of dust densities. The figure shows that the temperature of electrons decays faster for larger values of the dust density n_d.These results are in line with the experimental measurements of Couëdel et al.[42] and others.[43–47] Figure 5 shows the variation of the effective collision frequency with time. It indicates that the effective collision frequency decreases with time because of the greater accumulation of electrons on the dust grains. The higher the dust density, the faster the mitigation of collision frequency. Figure 6 shows that the electron density decreases with the dielectric constant and hence with the dust density. These outcomes are in good agreement with the conclusion of Schneider et al.[48] Figures 7 and 8 show the dependence of the beam width parameter f on the normalized channel of propagation for different dust densities n_d and dust charge state Z_d, respectively. From Figure 7, it is seen that the beam width parameter f quickly diminishes with ξ and leads to self-filamentation in the presence of dust grains in a fully ionized plasma. It shows that the self-filamentation of a Gaussian EM beam increases in the presence of dust grains, which is in contrast with the result obtained by Sodha et al.[27] in their earlier analysis of the propagation of the beam in a fully ionized plasma without dust. The extent of self-filamentation decreases with the dust density because of the reduced non-linearity due to the presence of dust grains, whereas the inverse result is obtained

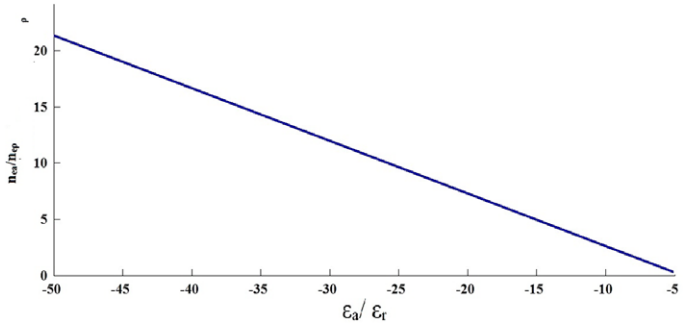


FIGURE 6 Vthe ariation of electron density, $n_{ea}/n_{e\rho}$, with the effective dielectric constant, $\varepsilon_a/\varepsilon_\rho$, for a fully ionized plasma in the presence of dust grains

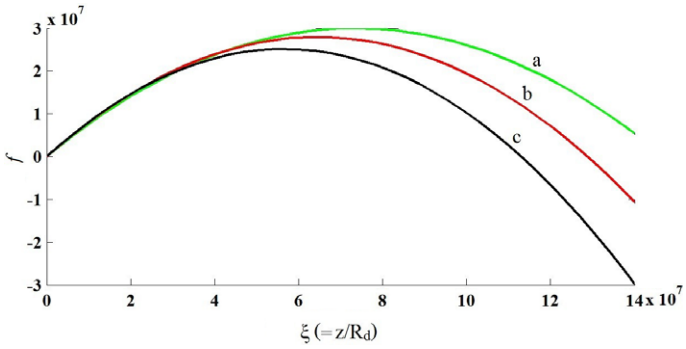


FIGURE 7 Dependence of the beam width parameter f on the normalized channel of propagation ξ (= z/R_d) for the transverse of Gaussian EM beam as a function of the number density of dust grains. The labels a, b, and c refer to $n_d = 2\times10^4$, 2×10^3, 2×10^2 cm^{-3}, respectively

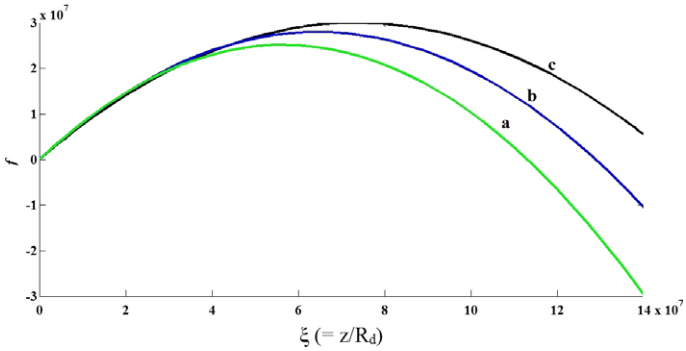


FIGURE 8 Dependence of f on ξ for different values of dust charge state Z_d. The other parameters are given in the text. The labels a, b, and c refer to $Z_d = -4\times10^3$, -6×10^3, -8×10^3, respectively

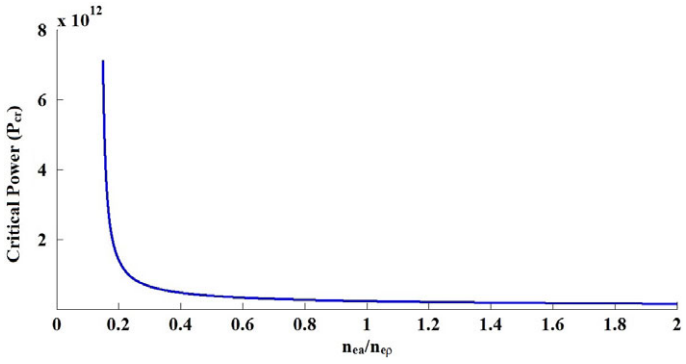


FIGURE 9 Variation of the effective critical power P_{cr} (ergs/s) with the normalized electron density $n_{ea}/n_{e\rho}$ in a fully ionized plasma in the presence of dust grains

in the case of the dust charge state. Figure 9 expresses the variation of the effective critical power P_{cr} (ergs/s) with electron density $n_{ea}/n_{e\rho}$ in a fully ionized plasma in the presence of dust grains. It indicates that the critical power increases with decaying electron density, i.e. enhanced dust density. So the power of the Gaussian beam must be increased for the beam to be focused for lower values of $n_{ea}/n_{e\rho}$.

From the above discussion, we can conclude that by changing the various dust/plasma parameters, an effective non-linearity can be obtained in a plasma. Moreover, the effective dielectric constant and self-filamentation behaviour of the beam can be estimated by modifying the various parameters.

6 | CONCLUSIONS

In this paper, we investigated the effect of dust grains on the self-filamentation of a Gaussian EM beam in a fully ionized plasma. An analytical model to study the dynamics of a fully ionized complex plasma was established. Under the charge neutrality condition, energy balance equations of the plasma constituents, perturbed number density of electrons, effective dielectric constant, and the effect of growth of dust grains on the various plasma parameters were examined. Moreover, the propagation of the Gaussian EM beam and its critical power in a fully ionized plasma in the presence of dust grains were studied. To analyse the behaviour of the beam, paraxial ray approximation was used, and the behaviour of self-filamentation was described by the curves. Our results may be useful for space and laboratory plasma experiments as well as for future studies in complex plasmas.

ACKNOWLEDGEMENTS

The authors wish to thank Prof. M.S. Sodha for valuable discussions and useful suggestions. One of the authors (R.S.) is grateful to Ms Jyotsna Panwar for her kind support and encouragement during the preparation of this manuscript.

REFERENCES

[1] M. S. Sodha, S. K. Mishra, S. Misra, *Phys. Plasmas* **2009**, *16*, 123701.
[2] A. Gahlot, R. Walia, S. C. Sharma, R. P. Sharma, *J. Plasma Phys.* **2012**, *78*, 33.
[3] Z. X. Wang, G. Lu, P. Duan, J. Liu, S. Zheng, Y. Liu, *Phys. Scr.* **2010**, *82*, 045501.
[4] I. Denysenko, J. Berndt, E. Kovacevic, I. Stefanovic, V. Selenin, J. Winter, *Phys. Plasmas* **2006**, *13*, 073507.
[5] M. Kundu, K. Avinash, A. Sen, R. Ganesh, *Phys. Plasmas* **2014**, *21*, 103705.
[6] M. Mikikian, L. Boufendi, A. Bouchoule, H. M. Thomas, G. E. Morfill, A. P. Nefedov, V. E. Fortov, the PKE–Nefedov team, *New J. Phys.* **2003**, *5*, 19.1.
[7] M. S. Sodha, S. K. Mishra, S. Misra, S. Srivastava, *Phys. Plasmas* **2010**, *17*, 073705.
[8] V. N. Tsytovich, U. de Angelis, A. V. Ivlev, G. E. Morfill, S. Khrapak, *Phys. Plasmas* **2005**, *12*, 112311.
[9] Y. Hong, C. Yuan, J. Jia, R. Gao, Y. Wang, Z. Zhou, X. Wang, H. Li, J. Wu, *Phys. Sci. Technol.* **2017**, *19*, 055301.
[10] J. H. Chu, I. Lin, *Phys. Rev. Lett.* **1994**, *72*, 4009; H. Thomas, G. E. Morfill, V. Demmel, J. Goree, B. Feuerbacher, D. Molmann, *Phys. Rev. Lett*. **1994**, *73*, 652; H. Thomas, G. E. Morfill, Nature (London) **1996** *379*, 806.
[11] S. V. Vladimirov, K. Ostrikov, *Phys. Rep.* **2004**, *393*, 175.
[12] G. S. Selwyn, *Jpn. J. Appl. Phys.* **1993**, *32*, 3068.
[13] J. Jia, C. Yuan, R. Gao, Y. Wang, Y. Liu, J. Gao, Z. Zhou, X. Sun, J. Wu, H. Li, S. Pu, *J. Phys. D: Appl. Phys.* **2015**, *48*, 465201.
[14] M. S. Sodha, A. Dixit, S. Srivastava, S. K. Mishra, M. P. Verma, L. Bhasin, *Plasma Sources Sci. Technol.* **2010**, *19*, 015006.
[15] M. S. Sodha, S. Misra, S. K. Mishra, *Plasma Sources Sci. Technol.* **2010**, *19*, 045022.
[16] M. S. Sodha, S. Mishra, S. K. Mishra, S. Srivastava, *J. Appl. Phys.* **2010**, *107*, 103307.
[17] M. S. Sodha, S. K. Mishra, S. Misra, *J. Appl. Phys.* **2011**, *109*, 013303.
[18] C. Zafiu, A. Melzer, A. Piel, *Phys. Plasmas* **2003**, *10*, 1278.
[19] M. S. Sodha, S. Mishra, S. K. Mishra, *Phys. Plasmas* **2010**, *17*, 113705.
[20] S. C. Sharma, A. Gahlot, R. P. Sharma, *Phys. Plasmas* **2008**, *15*, 043701.
[21] S. K. Mishra, S. Misra, M. S. Sodha, *Phys. Plasmas* **2011**, *18*, 043702.
[22] M. S. Sodha, S. K. Mishra, S. Misra, *Phys. Plasmas* **2011**, *18*, 023701.
[23] F. Verheest, P. Meuris, *Phys. Lett. A* **1996**, *210*, 198.
[24] S. Prasad, V. K. Tripathi, *J. Appl. Phys.* **1973**, *44*, 4595.
[25] G. Sambandan, V. K. Tripathi, J. Parashar, R. Bharuthram, *Phys. Plasmas* **1999**, *6*, 762.
[26] M. S. Sodha, S. K. Mishra, S. K. Aggarwal, *Phys. Plasmas* **2007**, *14*, 112302.
[27] M. S. Sodha, R. K. Khanna, V. K. Tripathi, *Opto-electronics* **1973**, *5*, 533.
[28] M. S. Sodha, S. Misra, S. K. Mishra, *Phys. Plasmas* **2009**, *16*, 123705.
[29] J. Jia, C. Yuan, S. Liu, F. Yue, R. Gao, Y. Wang, Z.-X. Zhou, J. Wu, H. Li, *Phys. Plasmas* **2016**, *23*, 043302.
[30] D. Li, L.-X. Guo, J.-T. Li, *Phys. Plasmas* **2018**, *25*, 013707.
[31] E. Stoffels, W. W. Stoffels, D. vender, G. M. W. Kroesen, F. J. de Hoog, *IEEE Trans. Plasma Sci.* **1994**, *22*, 2.
[32] H. Li, J. Wu, Z. Zhou, C. Yuan, *Phys. Plasmas* **2016**, *23*, 073702.
[33] N. A. Zharova, A. G. Litvak, V. A. Mironov, *JETP Lett.* **2003**, *78*, 619.
[34] M. S. Sodha, A. Dixit, *Appl. Phys. Lett.* **2009**, *95*, 101502.
[35] L. Spitzer Jr., *Physics of Fully Ionized Gases*, Inter Science Publishers, John Wiley and Sons, New York **1967**.
[36] A. V. Gurevich, *Nonlinear Phenomena in the Ionosphere*, Springer, New York **1978**.
[37] J. D. Huba, *NRL Physics Formularly*, Naval Research Laboratory, Washington, DC **1998**, p. 28.
[38] S. A. Akhamanov, A. P. Sukhorukov, R. V. Khokhlov, *Sov. Phys. Usp.* **1968**, *10*, 619.
[39] M. S. Sodha, R. S. Mittal, S. Kumar, V. K. Tripathi, *Opto-electronics (London)* **1974**, *6*, 167.
[40] M. S. Sodha, A. K. Ghatak, V. K. Tripathi, *Progress in Optics*, Vol. *13*, North Holland and Co., Amsterdam, **1976**, p. 169.
[41] M. S. Sodha, A. K. Ghatak, V. K. Tripathi, *Self-filamentation of LaserBeams in Dielectrics, Semiconductors and Plasmas*, Tata-McGraw Hills, Delhi **1974**.
[42] L. Couëdel, A.A. Samarian, M. Mikikian, L. Boufendi, in *Fifth Int. Conf. on the Physics of Dusty Plasmas*, Ponta Delgada, Azores, Portugal **2008**. https://doi.org/10.1063/1.2996828.
[43] X. Liang, J. Zheng, J. X. Ma, W. D. Liu, J. Xie, G. Zhuang, C. X. Yu, *Phys. Plasmas* **2001**, *8*, 1459.
[44] A. Douglass, V. Land, L. Matthews, T. Hydea, *Phys. Plasmas* **2011**, *18*, 083706.
[45] A. Barken, N. D'Angelo, R. L. Merlino, *Planet. Space Sci.* **1995**, *43*, 905.
[46] L. Couëdel, M. Mikikian, A. A. Samarian, L. Boufendi, *Phys. Plasmas* **2010**, *17*, 083705.
[47] M. Mikikiana, L. Couëdel, M. Cavarroc, Y. Tessier, L. Boufendi, *Eur. Phys. J. Appl. Phys.* **2010**, *49*, 13106.
[48] R. S. Schneider, L. F. Ziebell, *Braz. J. Phys.* **2006**, *36*, 759.
[49] R. Fowler, *Statistical Mechanics*, Cambridge University Press, **1936**, p. 358.

How to cite this article: Sharma R, Sharma SC. Theoretical model for the effect of dust grains on self-filamentation of a gaussian electromagnetic beam in a fully ionized plasma. *Contributions to Plasma Physics* 2018;1–15. https://doi.org/10.1002/ctpp.201800058

APPENDIX A: AMBIPOLAR DIFFUSION AND TRANSPORT COEFFICIENTS

We assume the fully ionized complex plasma to be made up of the constituent electrons, ions, and charged dust grains. The net particle flux or density decay of the constituent species is given by

$$\Phi_e = -n_e\mu_e U - D_{1e}n_e(\nabla T_e/T_e) - D_{2e}\nabla n_e, \tag{A1}$$

$$\Phi_i = -n_i\mu_i U - D_{1i}n_i(\nabla T_i/T_i) - D_{2i}\nabla n_i, \tag{A2}$$

$$\Phi_d = -n_d\mu_d \mathrm{U} - D_{1d}\nabla n_d, \tag{A3}$$

where Φ_g, n_g, U, T_g, $D_{1g}(=5k_BT_g/2m_g\nu_{effg})$, $D_{2g}(=k_BT_g/m_g\nu_{effg})$, $\mu_g(=eD_{2g}/k_BT_g)$, are the flux, number density, space charge field, temperature, thermal diffusion coefficient, diffusion coefficients, and mobility, and the subscript g stands for electrons, ions, or dust grains.

The net rate of flow of both charges per unit area must be balanced in the stationary state, so we get

$$\Phi_i = \Phi_e + \Phi_d. \tag{A4}$$

The dust grains are supposed to be standing still because of their heavy mass, and hence their flux comes out to be zero. Thus A4 gives

$$\Phi_i = \Phi_e. \tag{A5}$$

Using the A1, A2, and A5, and eliminating U, the net particle flux or density decay comes out to be

$$\begin{aligned}\Phi_p = &-D_{1e}n_e(\nabla T_e/T_e) - D_{2e}\nabla n_e \\ &- n_e\mu_e\left(\frac{D_{2i}\nabla n_i - D_{2e}\nabla n_e + D_{1i}n_i(\nabla T_i/T_i) - D_{1e}n_e(\nabla T_e/T_e)}{n_e\mu_e + n_i\mu_i}\right),\end{aligned} \tag{A6}$$

where $c_1 = D_{2e}/D_{2i}$, $c_2 = \mu_e/\mu_i$ and $c_3 = D_{1e}/D_{1i}$.

In this paper, axial parameters of diffusion coefficients and mobility are taken into consideration.

APPENDIX B: EFFECTIVE CRITICAL POWER

Consider a uniform waveguide propagation and a Gaussian beam propagating with a plane wave front ($R = \infty$) at $z = 0$. If at $z = 0$ the two terms on the RHS of 28 compensate each other, $\frac{d^2f}{dz^2} = 0, f = 1$(for all values of z). The critical condition for a uniform waveguide propagation is given by

$$R_{ns} = R_d, \tag{B1}$$

$$R_{ns}^2 = R_d^2, \tag{B2}$$

$$\frac{2\rho_0^2\varepsilon_a}{\varepsilon_\rho U_{00}^2} = k^2\rho_0^4, \tag{B3}$$

or

$$\frac{2\varepsilon_a}{\varepsilon_\rho k^2\rho_0^2} = U_{00}^2. \tag{B4}$$

The corresponding critical power is given by

$$P_{cr} = \int_0^\infty \frac{\varepsilon_a}{8\pi}U_{00}^2(\rho,z)\frac{c}{\eta}2\pi\rho d\rho = \frac{c}{8}\varepsilon_a^{1/2}\rho_0^2U_{00}^2, \tag{B5}$$

where η is the refractive index of medium.

Substituting U_{00}^2 from B4 into B5, we obtain

$$P_{\mathrm{cr}} = \frac{c\varepsilon_{\mathrm{a}}^{1/2}}{4\varepsilon_{\rho}k^2}. \tag{B6}$$

Substituting $k^2 = \varepsilon_{\mathrm{a}}\omega_{\mathrm{d}}^2/c^2$ in B6, we obtain

$$P_{\mathrm{cr}} = \frac{c^3\varepsilon_{\mathrm{a}}^{1/2}}{4\omega_{\mathrm{d}}^2\varepsilon_{\rho}}. \tag{B7}$$

International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934

Thermodynamic analysis of vapour compression refrigeration system with mixing of Al_2O_3Nano material in fourteen ecofriendly refrigerants

R.S. Mishra

Department of Mechanical & Production Engineering, Delhi Technological University Delhi, India

Abstract

The most commonly-used method for analysis of an energy-conversion process is the first law of thermodynamics. But in the recent decades, the exergetic performance based on the second law of thermodynamics has found as useful method in the design, evaluation, optimization and improvement of vapour compression refrigeration systems. The exergetic performance analysis can not only determine magnitudes, location and causes of irreversibility in the vapour compression refrigeration systems. But also provides more meaningful assessment of power plant individual components efficiency. Using concept of exergy, a thermal model for vapour compression refrigeration system using nano refrigerants was developed and numerical computation was carried out by using Al2O3 mixed in fourteen ecofriendly refrigerants and found that thermodynamic performances in terms of first law efficiency, exergetic efficiency /second law efficiency increased significantly. © 2018 ijrei.com. All rights reserved

Keywords: Thermodynamic performances, Energy-exergy Analysis, Eco friendly refrigerants, Nano- materials.

1. Introduction

The second law analysis (i.e. exergy Computation) is widely accepted as a useful tool for obtaining overall performances of any system for finding various exergy losses occurred in its components Exergy analysis also helps in taking account the important engineering decisions regarding design parameters of a system by finding maximum exergy destruction using entropy generation principle. A conventional exergetic analysis reveals irreversibility within each component of a vapour compression refrigeration systems. Exergetic analysis provides the tool for a clear distinction between energy losses to the environment and internal irreversibility in the process because exergy analysis is a methodology for the evaluation of the performance of devices and processes, and examining the exergy at different points in a series of energy-conversion steps. With this information, efficiencies can be evaluated, and the process steps having the largest losses (i.e., the greatest margin for improvement) can be identified. For these reasons, the modern approach uses the exergy analysis in the vapour compression refrigeration systems, which provides a more realistic view of the process and a useful tool for engineering evaluation

Many researchers have carried out exergy studies of different thermal energy conversion systems describing various approach for exergy analysis and its usefulness for improving existing designs by reducing exergy destruction in a more simple and effective manner [1]. Yumrutas et. al [2] investigated of the effects of the evaporating and condensing temperatures on the pressure losses, exergy losses, second law of efficiency, and the COP of a vapour compression cycle. Dincer [3] asserts that conventional energy analysis, based on the first law of thermodynamics, evaluates energy mainly on its quantity but analysis that are based on second law considers not only the quality of energy, but also quantity of energy.. Nikolaidis and Probert [4] used exergy method for calculating ting thermodynamic performances of R22 in a two-stage compound compression cycle, with flash intercooling. Bejan [5] developed, thermodynamic model by using heat transfer irreversibility and showed that the exergetic efficiency decreases as evaporator temperature decreases. Sajadi A.R. et.al. [6] experimentally investigated turbulent convective heat transfer coefficient and pressure

Corresponding author: R. S. Mishra
Email Address: hod.mechanical.rsm@dtu.ac.in

drop of TiO_2 dispersed in water nano fluid in the circular tube and also compared experimental results with correlation of Nusselt number and not concluded that how much (%) convective heat transfer coefficient increased. Huang Dan, et.al. [7], observed the effect of hybrid nano fluid mixture Al_2O_3 in the plate heat exchanger and found that convective heat transfer coefficient is increased as compared without nano fluid. Zeinali S Heris et.al [8] experimentally investigated the convective heat transfer increment in the laminar flow forced convection heat transfer due to increase in the thermal conductivity of nano fluid due to nano particle fluctuations present in that fluid The above investigators have not studied in detail regarding performances of the vapour compression refrigeration systems in which nano refrigerants was circulated in the whole system. Therefore present investigation take care the overall system thermal performances in terms of first law efficiency (COP) and second law efficiency/ exergetic efficiency and various efficiencies of compressor used in the system.

2. Result and Discussion

2.1 Variation of Evaporator and condenser temperature on thermal performances by mixing TiO_2 nano particles

Table-1 shows the performance of vapour compression refrigeration system using fourteen ecofriendly (alternate) refrigerants mixed with Al_2O_3. It was found that refrigerants R134a R404a and R507a give similar trends in their first law efficiency (COP), second law efficiency, exergetic efficiency and exergy destruction ratio, compressor efficiency, isentropic efficiency and volumetric efficiency with slightly variations. Similarly R1234yf and R125 and R600a gives similar trends with slightly variation, however lower than R134a. The best performances were found using R290 hydrocarbon and R407c which is higher than R134a. The worst performances were found by using R410a, R236fa, and super worst first and second law performances were observed by using Al_2O_3 mixed in R245fa, R123 refrigerants. Table-2 shows the variation of thermal performance with varying evaporator temperature. It is found that first law efficiency (COP) increases with increasing evaporator temperature, In case of ,second law efficiency and exergetic efficiency, it is found that exergetic efficiency is increases with increasing evaporator temperature up to a certain temperature where exergetic efficiency, second law efficiency becomes optimum(i.e. maximum) and then started decreases with increasing evaporator temperature, the optimum value of evaporator temperature is found to be 274K (i.e. 1^{o}C) while in case of optimum second law efficiency, the evaporator temperature to be 271K (i.e -2 ^{o}C). Similarly exergy destruction ratio which is a ratio of total exergy destruction Exergy (Losses in the system was sum of exergy destruction in components to the exergy of fuel or exergy of product) was evaluated based on exergy of fuel (i.e. total electrical power required for running compressor) and also EDR was evaluated based on exergy of product (i.e. effective utilization of cooling load) was decreasing while increasing evaporator temperature. The Isentropic compressor efficiency and volumetric efficiency is also increases with increasing evaporator temperature and it has optimum value at 274K (i.e. 1^{o}C). Similarly Table-3 shows the variation of thermal performance with varying condenser temperature. It is found that first law efficiency (COP) decreases with decreasing condenser temperature, In case of second law efficiency and exergetic efficiency, it is found that exergetic efficiency is decreases with increasing condenser temperature. Similarly isentropic efficiency and volumetric efficiency decreases by increasing condenser temperature while exergy destruction ratio is also increases with increasing condenser temperature

Table-1: Variation of COP with evaporator temperature of Vapour compression usingAl_2O_3 nano materials in ecofriendly for condenser temp=48 ^{0}C & evaporator temperature=-5^{0}C

Refrg.	COP	EDR	Rational EDR	Exergetic Efficiency	Pressure Ratio (r_P)	Compressor Efficiency	Isentropic Efficiency	Volumetric Efficiency	Second Law Efficiency
R134a	4.36	1.659	0.6239	0.3761	6.032	0.8378	0.6339	0.5938	0.4860
R404a	4.316	1.694	0.6288	0.3712	5.003	0.8793	0.7321	0.6274	0.4831
R1234yf	4.004	1.976	0.6640	0.3360	5.475	0.8520	0.6992	0.6052	0.4479
R1234ze	3.836	2.150	0.6826	0.3174	6.129	0.8283	0.6183	0.5860	0.4294
R227ea	3.147	3.162	0.7597	0.2403	6.328	0.8157	0.5813	0.5758	0.3527
R407c	4.736	1.391	0.5818	0.4182	5.737	0.8514	0.6730	0.6067	0.5301
R507a	4.328	1.685	0.6275	0.3725	4.939	0.8818	0.7355	0.6293	0.4844
R125	4.044	1.935	0.6593	0.3407	4.97	0.8788	0.7339	0.6269	0.4527
R290	4.826	1.335	0.5718	0.4282	4.73	0.8835	0.7456	0.6308	0.5403
R600a	4.009	1.963	0.6632	0.3368	5.759	0.8411	0.6705	0.5964	0.4088
R152a	5.164	1.143	0.5334	0.4666	5.968	0.8465	0.6434	0.6008	0.5786
R32	6.952	0.5008	0.3337	0.6663	5.086	0.9008	0.7273	0.6447	0.7783
R410a	2.14	6.836	0.8727	0.1276	5.043	0.8924	0.7299	0.6379	0.2396
R236fa	1.812	9.998	0.9091	0.09093	7.426	0.7761	0.2031	0.5438	0.2029

Table- 2: Variation of thermal performances (First law efficiency (COP) and Exergetic efficiency) with variation of evaporator temperature using Al_2O_3 mixed R134a for condenser temp=48 ^0C.

Evaporator Temperature (K)	COP	EDR	Rational EDR	Exergetic Efficiency	Pressure Ratio (r_p)	Isentropic compressor Efficiency	Volumetric Efficiency	ED_Ratio	Second law Efficiency
263	2.182	5.356	0.8427	0.1573	7.319	0.7966	0.2567	0.7096	0.2904
268	4.36	1.659	0.6239	0.3761	6.032	0.8378	0.6399	0.512	0.488
270	4.87	1.492	0.5987	0.4013	5.596	0.8519	0.688	0.495	0.505
271	5.095	1.451	0.5920	0.4080	5.392	0.8585	0.7061	0.4924	0.5076
272	5.309	1.428	0.5881	0.4119	5.198	0.8648	0.7308	0.4925	0.5075
273	5.517	1.418	0.5864	0.4136	5.012	0.8709	0.7316	0.4948	0.5052
274	5.724	1.417	0.5862	0.4138	4.834	0.8767	0.7408	0.4986	0.5014
275	5.932	1.424	0.5875	0.4125	4.664	0.8823	0.7485	0.5038	0.4962
276	6.146	1.438	0.5898	0.4102	4.501	0.8876	0.7550	0.5101	0.4899
277	6.366	1.458	0.5932	0.4068	4.345	0.8927	0.7608	0.5174	0.4826
278	6.595	1.484	0.5975	0.4025	4.196	0.8976	0.7659	0.5256	0.4744
283	7.932	1.722	0.6326	0.3674	3.538	0.9194	0.7859	0.5796	0.4204

Table- 3: Variation of thermal performance (First law efficiency (COP) and Exergetic efficiency) with variation of Condenser temperature using Al_2O_3 mixed R134a for evaporator temp=268 K

Condenser Temperature (K)	COP	EDR	Rational EDR	Exergetic Efficiency	Pressure Ratio (r_p)	Isentropic Compressor Efficiency	Volumetric Efficiency	ED_Ratio	Second law Efficiency
328	2.373	5.507	0.8463	0.1537	7.182	0.8017	0.319	0.7344	0.2656
323	3.88	2.102	0.6776	0.3224	6.375	0.8279	0.5777	0.5657	0.4343
318	5.017	1.224	0.5503	0.4497	5.584	0.8520	0.6891	0.4383	0.5617
313	6.096	0.7531	0.4297	0.5704	4.893	0.8742	0.7379	0.3176	0.6824
308	7.358	0.405	0.2882	0.7118	4.269	0.8944	0.7634	0.1763	0.8237

3. Conclusion

Following conclusions were drawn while using Al_2O_3 in the ecofriendly refrigerants.

- The first law efficiency (COP) increases with increasing evaporator temperature
- The exergetic efficiency is increases with increasing evaporator temperature up to a certain temperature where exergetic efficiency, second law efficiency becomes optimum (i.e. maximum) and then started decreases with increasing evaporator temperature.
- The optimum value of evaporator temperature is found to be 274K (i.e. 1^{o}C) while in case of optimum second law efficiency, the evaporator temperature to be 271K (i.e. -2 ^{o}C).
- Exergy destruction ratio (EDR) is decreasing while increasing evaporator temperature.
- Exergy destruction ratio(EDR) is also increases with increasing condenser temperature
- The best performances (COP and exergetic efficiency) were found using R290 hydrocarbon and R407c which is higher than R134a.
- Refrigerants R134a R404a and R507a give similar trends in their first law efficiency (COP), second law efficiency, exergetic efficiency and exergy destruction ratio, compressor efficiency, isentropic efficiency and volumetric efficiency with slightly variations.
- HFO-1234yf, R125 and R600a gives similar trends with slightly variation, however lower than R134a.
- Nano mixed R1234ze gives lowe performance as compared with nano mixed HFO-1234yf
- The worst performances were found by using R410a, R236fa
- Super worst first and second law performances were observed by using Al_2O_3 mixed in R245fa, R123 refrigerants.
- The Isentropic compressor efficiency and volumetric efficiency is also increases with increasing evaporator temperature and it has optimum value at 274K (i.e. 1^{o}C).
- first law efficiency (COP) decreases with decreasing condenser temperature
- Similarly isentropic efficiency and volumetric efficiency decreases by increasing condenser temperature.

References

[1] R. S. Mishra, Appropriate Vapour Compression Refrigeration Technologies for Sustainable Development International Journal of Advance Research & Innovations, 2014, 551-556

[2] D. Yumrutas et. al Exergy analysis of vapour compression refrigeration systems, Exergy, an Int J, 2 (2002) 266-272

[3] Dincer I. Refrigeration systems and applications. London: Wiley; 2003. pp. 26–27.

[4] Nikolaidis C, Probert D. Exergy-method analysis of a two stage vapor compression refrigeration plants performance. Appl Energy. 1998;60:241–56

[5] Bejan A, Theory of heat transfer irreversible refrigeration plants, Int J Heat mass transfer, 32(9), (1989), 1631-39.

[6] Sajadi A.R. [2011].Investigation of turbulent convective heat transfer and pressure drop of TiO2/water nano fluid in circular tube, International communications in Heat and mass transfer, Elsevier,Vol-38, page-1474-1478.

[7] Huang Dan , et.a. {2016] , Effect of hybrid nano fluid mixture in the plate heat exchangers, Experimental thermal and fluid science, Elsevier, Vol-72, pp 190-196.

[8] Zeinali S Heris et.al [2007] "Experimental investigation of convective heat transfer coefficient using Al_2O_3 /water nano fluid in circular tube" , J of Heat & fluid flow, Elsevier, Vol-28, page-203-210.

International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934

Thermodynamic analysis of vapour compression refrigeration systems using alternative refrigerants

R.S. Mishra

Department of Mechanical & Production Engineering, Delhi Technological University Delhi, India

Abstract

Two methods for predicting thermal performances of various vapour compression refrigeration systems have been considered in this paper. The numerical computation was carried out for fifteen multiple and compressor systems refrigeration systems for different loads on various evaporators using (i) first method known as energy – exergy analysis for finding exergy destruction ratio and exergetic efficiency and first law efficiency in terms of COP (coefficient of performance) and second method is known as irreversibility prediction using entropy generation principle considered the exergy defects (losses) in the various components and computing total exergy destruction ratio and rational efficiency and second law efficiency. It was observed that by altering the load on the evaporators, the thermal performances effects significantly. © 2018 ijrei.com. All rights reserved

Keywords: Thermodynamic Analysis, Energy-Exergy Analysis, Entropy Generation principle, Irreversibility Analysis

1. Introduction

The most commonly-used method for analysis of an energy-conversion process is the first law of thermodynamics. But in the recent decades, the exergetic performance based on the second law of thermodynamics has found as useful method in the design, evaluation, optimization and improvement of vapour compression refrigeration systems.

The exergetic performance analysis can not only determine magnitudes, location and causes of irreversibility in the vapour compression refrigeration systems, but also provides more meaningful assessment of power plant individual components efficiency.

A conventional exergetic analysis reveals irreversibility within each component of a vapour compression refrigeration systems. Exergetic analysis provides the tool for a clear distinction between energy losses to the environment and internal irreversibility in the process because exergy analysis is a methodology for the evaluation of the performance of devices and processes, and examining the exergy at different points in a system components using energy-conversion concept. With this concept, various efficiencies can be evaluated in the VCRS, and the process steps having the largest losses (i.e., the greatest margin for improvement) can be identified. For these reasons, the modern entropy generation principles approach uses the exergy analysis in the vapour compression refrigeration systems, which provides a more realistic view of the process and a useful tool for engineering evaluation.

2. Energy Exergy Analysis of Vapour Compression Refrigeration Systems

The second law analysis (i.e. exergy Computation) is widely accepted as a useful tool for obtaining overall performances of any system for finding various exergy losses occurred in its components Exergy analysis also helps in taking account the important engineering decisions regarding design parameters of a system by finding maximum exergy destruction using entropy generation principle.

Many researchers have carried out exergy studies of different thermal energy conversion systems describing various approach for exergy analysis and its usefulness for improving existing designs by reducing exergy destruction in a more simple and effective manner [1-3]. Yumrutas et. al [4,7] investigated of the effects of the evaporating and

Corresponding author: R. S. Mishra
Email Address: hod.mechanical.rsm@dtu.ac.in

condensing temperatures on the pressure losses, exergy losses, second law of efficiency, and the COP of a vapour compression cycle. Dincer [5] asserts that conventional energy analysis, based on the first law of thermodynamics, evaluates energy mainly on its quantity but analysis that are based on second law considers not only the quality of energy, but also quantity of energy. Kumar et al. [6] also computed the exergetic analysis of a VCR system using R11 and R12 as refrigerants. Nikolaidis and Probert [7] used exergy method for calculating ting thermodynamic performances of R22 in a two-stage compound compression cycle, with flash intercooling. Bejan [8] developed, thermodynamic model by using heat transfer irreversibility and showed that the exergetic efficiency decreases as evaporator temperature decreases.

3. Results and Discussion

The percentage exergy destruction in the evaporator varies with the type of refrigerants. It becomes high when R236fa is used and lower when R32 is used .By using R245fa it becomes 20% while using R123 it has 20.43% .In the case comparing R134a, R1234yf and R1234ze , it becomes 16.15% by using R134a and when R1234yf is used it becomes less around 15.8% as 16.16% by using R134a.and 18.3% by using R1234ze. The maximum exergy destruction in the system components it is found maximum (i.e. higher) in the condenser and low in the throttle valves. By comparing alternative refrigerants, the exergy destruction is low by using R227ea around 35.4% and maximum by using R32 In the case of using HFO refrigerants for replacing HFC-134a and other Refrigerants, it is around similar trends with slightly less by using R1234yf, around 41.22% as compared to R134a is 41.94%. In case of compressors it is almost similar trends, however, it is higher 35.49% when R1234yf is used and 34.8% when R1234ze is used as compared to 34.5% when R134a is used. It becomes 34.8% when R227ea is used and 34.07% when R245fa is used as compared to R134a is used it becomes 35.5%. Still it is also varying by using different alternative refrigerants. It gives higher exergy destruction by using R227ea and lowest by using R23. By comparing exergy destruction by using HFO refrigerants for replacing R134a, it gives 7.624% by using R1234yf as higher as comparing with R134a but 7.457% by using R1234ze as compared to R134a around 7.402%.The first law efficiency in terms of COP (coefficient of performance)by using alternative refrigerants , the maximum COP is found 3.766 by using R123 while by using R245fa it becomes slightly less around 3.727. Similarly exergetic efficiency is higher by using M-32 and slightly less by using HFO refrigerants 28.59% by using R1234yf and 28.82% by using R1234ze as compared to R134a is 29.97%. The total power required to run whole system in terms of exergy of fuel is high by using M-32 is 61.52 “kW’ and low by using R123 is 55.76 “kW’ for 210 kW of total load of evaporator.

Vapour compressor refrigeration system with multiple evaporators at different temperatures with single compressor, individual expansion valves using following data:

Q_EVA_1=70”kW” Q_EVA_2=105”kW” Q_EVA_3=35 kW, T_EVA_1=268 “K”,T_EVA_2=273 “K” ,T_EVA_3=28 3“K”

We considered vapour compressor refrigeration system with multiple evaporators at different temperatures with compound compression, individual expansion valves (System-1) for energy-exergy- analysis. To validate computational results from developed thermal model of system-6, the following input values have been taken. The computed results of system-1 for 100% compressors efficiency were compared and shown in Table-1(a) respectively. For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects)in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves are shown in Table-1(b) to Table-1 (d) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances.

Table-1(a) Validation of results of vapour compressor refrigeration system with Multiple evaporators at different temperatures with single compressor for 100% compressor efficiency

Parameter	*Program*	*Ref [13]*
COP	*5.232*	*4.80*
Total Work (KW)	*40.14*	*43.73*

Table-1(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1=0.80) , $T_{_EVA_1}$=268"K", $T_{_EVA_2}$=273"K", $T_{_EVA_3}$=280"K", Effect of superheating= 5°C , T_R=T_EVA+5, (i.e. T_R_1=273"K", T_R _2=278"K", T_R_3=285"K"), Q_EVA_1=70 "kW', Q_EVA_2=105"kW', Q_EVA_3=35"kW',T_Cond=303"K", T_subcooled=303"K"

Refrigerants	First law Efficiency (System COP)	System EDR	Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	Secnd Law Efficiency
R12	4.185	1.823	0.3204	50.18	16.08	0.4033
R134a	4.140	1.861	0.3155	50.71	16.0	0.3974
R1234yf	4.061	1.946	0.3041	51.71	15.73	0.3843
R1234ze	4.164	1.917	0.3092	50.43	15.59	0.3913
R227ea	3.966	2.152	0.2842	52.95	15.05	0.3620
R236fa	4.14	2.093	0.2921	50.73	14.82	0.3731
R245fa	4.297	1.911	0.3120	48.87	15.25	0.3964

Table-1.(c)Exergy Destruction of various components based on exergy of fuel of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1=0.80) , $T_{_EVA_1}$=268"K", $T_{_EVA_2}$=273"K", $T_{_EVA_3}$=280"K", Effect of superheating= 5°C , T_R=T_EVA+5, (i.e. T_R_1=273"K", T_R _2=278"K", T_R_3=285"K"), Q_EVA_1=70 "kW', Q_EVA_2=105"kW', Q_EVA_3=35"kW',T_Cond=303"K", T_subcooled=303"K"

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	Total % Losses	Second Law effectiveness
R12	11.20	4.495	24.49	18.21	64.57	0.3543
R134a	11.68	4.834	23.9	18.31	65.05	0.3495
R1234yf	11.21	5.114	24.03	18.82	66.05	0.3303
R1234ze	12.47	4.969	23.10	18.74	65.72	o.3228
R227ea	14.22	6.22	21.69	19.02	68.27	0.3173
R236fa	14.62	4.93	22.57	19.01	67.67	0.3233
R245fa	13.24	3.767	23.93	18.69	65.65	0.3435

Table-1.(d) Exergy Destruction of various components based on Total exergy destruction of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1=0.80) , $T_{_EVA_1}$=268"K", $T_{_EVA_2}$=273"K", $T_{_EVA_3}$=280"K", Effect of superheating= 5°C , T_R=T_EVA+5, (i.e. T_R_1=273"K", T_R _2=278"K", T_R_3=285"K"), Q_EVA_1=70 "kW', Q_EVA_2=105"kW', Q_EVA_3=35"kW',T_Cond=303"K", T_subcooled=303"K"

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	Second law Effectiveness	% Irreversibility= Total Exergy Losses to the Exergy of Fuel
R12	19.18	7.696	41.93	31.17	0.3543	64.57
R134a	19.89	8.239	40.7	31.18	0.3495	65.05
R1234yf	18.95	8.642	40.61	31.8	0.3303	66.05
R1234ze	21.04	8.381	38.96	31.62	o.3228	65.72
R227ea	23.26	10.17	35.46	31.11	0.3173	68.27
R236fa	23.92	8.066	36.92	31.09	0.3233	67.67
R245fa	22.21	6.317	40.13	31.34	0.3435	65.65

System-2 : vapour compressor refrigeration system with Multiple evaporators at different temperatures with multiple compressors, individual expansion valves ($Q_{_EVA_1}$=105"kW" $Q_{_EVA_2}$=70"kW" $Q_{_EVA_3}$=35"kW"). We considered vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves (System-2) for energy-exergy- analysis . To validate computational results from developed thermal model of system-6, the following input values have been taken. The computed results of system-6 for 100% compressors efficiency were compared and shown in Table-2(a) respectively. For finding thermal performances in terms of first law efficiency (COP), EDR(Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects)in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves are shown in Table-2(b) to Table-2 (c) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances

Table-2(a) Thermal Performances (First law efficiency and Second law efficiency, etc.) of of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_EVA_1=263"K", T_EVA_2=278"K", T_EVA_3=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K", Q_EVA_1=35 "kW', Q_EVA_2=70"kW', Q_EVA_3=105"kW',T_subcooled=303"K"

Refrigerants	First law Efficiency (System COP)	EDR	% Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	Secnd Law Efficiency
R12	5.697	2.195	30.58	36.86	11.27	0.7569
R134a	5.654	2.213	30.35	37.14	11.27	0.7524
R1234yf	5.560	2.253	29.85	37.77	11.27	0.740
R1234ze	5.624	2.199	30.49	36.97	11.27	0.7559
R-32	5.445	2.320	29.23	38.56	11.27	0.7247
R227ea	5.461	2.309	29.32	38.45	11.27	0.7268
R236fa	5.664	2.213	30.45	37.08	11.27	0.7538
R245fa	5.832	2.136	31.34	35.97	11.27	0.7769
R123	5.877	2.118	31.55	35.73	11.27	0.7821

T*able-2.(b)Exergy Destruction of various components based on exergy of fuel of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_EVA_1=263"K", T_EVA_2=278"K", T_EVA_3=283"K", Q_EVA_1=35 "kW', Q_EVA_2=70"kW', Q_EVA_3=105"kW',T_subcooled=303"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K"*

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	Total % Losses	Rational Efficiency
R12	10.86	6.944	30.67	18.56	67.03	0.3297
R134a	10.74	7.614	30.07	18.63	67.06	0.3294
R1234yf	8.731	8.397	31.0	18.97	67.1	0.3290
R1234ze	10.40	7.931	29.56	19.01	66.9	0.3310
R-32	10.29	6.992	33.73	17.29	68.30	0.3120
R227ea	10.07	10.15	28.22	19.02	67.46	0.3254
R236fa	11.87	8.047	29.25	19.04	67.14	0.3286
R245fa	11.14	6.137	30.53	18.99	66.79	0.3321
R123	11.01	5.501	31.43	18.77	66.72	0.3328

Table-2.(c) Exergy Destruction of various components based on Total exergy destruction of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_EVA_1=263"K", T_EVA_2=278"K", T_EVA_3=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K", Q_EVA_1=35 "kW', Q_EVA_2=70"kW', Q_EVA_3=105"kW',T_subcooled=303"K"

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	%Irreversibility	Second Law effectiveness
R12	16.20	10.36	45.75	27.68	67.03	0.3297
R134a	16.02	11.35	44.84	27.78	67.06	0.3294
R1234yf	13.01	12.51	46.2	28.27	67.1	0.3290
R1234ze	15.54	11.86	44.19	28.41	66.9	0.3310
R-32	15.06	10.24	49.39	25.31	68.30	0.3120
R227ea	14.93	15.04	41.83	28.2	67.46	0.3254
R236fa	16.09	11.99	43.57	28.36	67.14	0.3286
R245fa	16.68	9.188	45.71	28.42	66.79	0.3321
R123	16.5	8.245	47.11	28.14	66.72	0.3328

System-3: Vapour compressor refrigeration system with multiple evaporators at different temperatures with multiple compression, individual expansion valves

(*Q_EVA_1=70"kW" Q_EVA_2=105"kW" Q_EVA_3=35"kW" T_EVA_1=268"K", T_EVA_2=273"K", T_EVA_3=278"K")*

We considered vapour compressor refrigeration system with multiple evaporators at different temperatures with compound compression, individual expansion valves (System-3) for energy-exergy- analysis. To validate computational results from developed thermal model of system-6, the following input values have been taken. The computed results of system-3 for 100% compressors efficiency were compared and shown in Table-3(a) respectively. For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects)in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour

compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves are shown in Table-3(b) to Table-3 (d) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances

Table-3(a) Validation of Results of Vapour compressor refrigeration system with Multiple evaporators at different temperatures with multiple compression, individual expansion valves for 100% compressor efficiency (Q_{EVA_1}=70"kW" Q_{EVA_2}=105"kW" Q_{EVA_3}=35"kW"T_{EVA_1}=268"K", T_{EVA_2}=273"K", T_{EVA_3}=278"K)

Parameter	Program	Ref [13]
COP	6.128	5.56
Total Work (KW)	34.27	38.4

Table-3(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of vapour compressor refrigeration system with Multiple evaporators at different temperatures with multiple compression, individual expansion valves Q_EVA_1=70"kW" Q_EVA_2=105"kW" Q_EVA_3=35"kW"T_{EVA_1}=268"K", T_{EVA_2}=273"K", T_{EVA_3}=278"K using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_{EVA_1}=268"K", T_{EVA_2}=273"K", T_{EVA_3}=278"K", T_R_1=273"K", T_R _2=278"K", T_R_3=283"K", Q_EVA_1=70 "kW', Q_EVA_2=105"kW', Q_EVA_3=35"kW',T_subcooled=303"K"

Refrigerants	First law Efficiency (System COP)	EDR	% Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	Second Law Efficiency
R12	4.902	1.798	36.98	42.89	15.82	
R134a	4.877	1.824	36.74	43.06	15.82	
R1234yf	4.815	1.866	36.27	43.61	15.82	
R1234ze	4.908	1.816	36.97	42.79	15.82	
R-32	4.629	1.937	34.87	45.37	15.82	0.7247
R227ea	4.744	1.935	35.74	44.26	15.82	
R236fa	4.895	1.837	36.88	42.9	15.82	
R245fa	5.023	1.749	37.84	41.81	15.82	
R123	5.046	1.722	0.3801	41.62	15.82	0.7821

Table-3.(c)Exergy Destruction of various components based on exergy of fuel of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_{EVA_1}=268"K", T_{EVA_2}=273"K", T_{EVA_3}=278"K", T_R_1=273"K", T_R _2=278"K", T_R_3=283"K", Q_EVA_1=70 "kW', Q_EVA_2=105"kW', Q_EVA_3=35"kW',T_subcooled=303"K"

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	% Total Exergy Destruction	Second Law effectiveness
R12	14.95	5.868	27.07	18.51	66.40	0.3360
R134a	15.47	6.386	26.57	18.59	67.01	0.3299
R1234yf	14.46	6.892	27.36	18.97	67.68	0.3232
R1234ze	15.51	6.599	26.04	19.0	67.15	0.3285
R-32	13.75	6.013	30.66	17.11	67.53	0.3247
R227ea	16.73	8.431	24.97	19.02	69.15	0.3085
R236fa	16.33	6.624	25.76	19.04	67.75	0.3225
R245fa	16.38	5.013	26.8	18.98	66.17	0.3383
R123	14.58	4.536	27.58	18.74	65.44	0.3456

Table-3.(d) Exergy Destruction of various components based on Total exergy destruction of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_{EVA_1}=263"K", T_{EVA_2}=278"K", T_{EVA_3}=283"K", T_R_1=268"K", T_R _2=283"K", T_R_3=288"K", Q_{EVA_1}=35 "kW', Q_{EVA_2}=70"kW', Q_{EVA_3}=105"kW',T_subcooled=303"K"

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	% Irreversibility	Second Law effectiveness
R12	22.52	8.837	40.77	27.83	66.40	0.3360
R134a	23.08	9.529	39.65	27.75	67.01	0.3299
R1234yf	21.37	10.18	40.42	28.03	67.68	0.3232
R1234ze	23.10	9.828	38.78	28.30	67.15	0.3285
R-32	20.36	8.903	45.4	25.34	67.53	0.3247
R227ea	24.2	12.19	36.11	27.51	69.15	0.3085
R236fa	24.10	9.776	38.02	28.10	67.75	0.3225
R245fa	23.24	7.577	40.5	28.68	66.17	0.3383
R123	22.28	6.931	42.15	28.64	65.44	0.3456

System-4 : Vapour compressor refrigeration system with Multiple evaporators at different temperatures with multiple compression with individual expansion valves **(**Q_EVA_1=70"kW" Q_EVA_2=105"kW" Q_EVA_3=35 "kW", T_EVA_1=268 "K", T_EVA_2=273"K", T_EVA_3=278 "K") We considered vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves (System-4) for energy-exergy- analysis . To validate computational results from developed thermal model of system-6, the following input values have been taken. The computed results of system-4 for 100% compressors efficiency were compared and shown in Table-4(a) .

respectively. For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects)in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves are shown in Table-4(b). It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances.

Table-4(a) Validation of Results of vapour compression refrigeration system with multiple evaporators at different temperatures with multiple compression with individual expansion valves for 100% compressor efficiency

Parameter	Program	Ref [13]
COP	5.938	5.56
Total Work (KW)	35.37	38.4

Table-4(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of ofvapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_EVA_1=268"K", T_EVA_2=273"K", T_EVA_3=278"K", T_R_1=273"K", T_R _2=278"K", T_R_3=283"K", Q_EVA_1=70 "kW', Q_EVA_2=105"kW', Q_EVA_3=35"kW',T_subcooled=303"K"

Refrigerants	First law Efficiency (System COP)	EDR	% Exergetic Efficiency	Exergy of Fuel (KW)	Exergy of Product (KW)	% Exergy Losses	Rational Efficiency	Second Law effectiveness
R12	4.723	2.195	35.58	44.47	15.82	64.42	0.3558	0.5266
R134a	4.676	2.213	35.23	44.91	15.82	64.77	0.3523	0.5235
R1234yf	4.582	2.253	34.51	45.83	15.82	65.49	0.3451	0.5123
R1234ze	4.691	2.199	35.34	44.74	15.82	64.66	0.3534	0.5261
R227ea	4.48	1.963	33.75	46.87	15.82	66.25	0.3375	0.5016
R236fa	4.668	1.844	35.16	44.99	15.82	64.84	0.3516	0.5225
R245fa	4.835	1.746	36.42	43.43	15.82	63.58	0.3642	0.5412

System-5: Vapour compressor refrigeration system with Multiple evaporators at different temperatures with single compressor, multiple expansion valves and back pressure valves for Q_EVA_1=105"kW" Q_EVA_2=70"kW" Q_EVA_3=35"kW"

Table-5(a) Validation of Results of vapour compressor refrigeration system with multiple evaporators at different temperatures with single compressor, multiple expansion valves and back pressure valves for 100% compressor efficiency

Parameter	Program	Ref [13]
COP	4.530	4.53
Total Work (KW)	46.36	46.7

Table-5(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of ofvapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_EVA_1=263"K", T_EVA_2=278"K", T_EVA_3=283"K", T_R_1=268"K", T_R _2=283"K", T_R_3=288"K",

Refrigerants	First law Efficiency (System COP)	EDR	% Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	Secnd Law Efficiency
R12	3.624	1.709	30.56	57.95	17.74	0.3784
R134a	3.579	1.747	29.97	58.68	17.59	0.3715
R1234yf	3.509	1.842	28.59	59.85	17.91	0.356
R1234ze	3.603	1.846	28.82	58.28	16.8	0.360
R-32	3.414	1.328	35.86	61.52	22.06	0.4298
R227ea	3.419	2.088	26.18	61.41	16.08	0.3295
R236fa	3.582	2.074	26.67	58.62	15.64	0.3374
R245fa	3.727	1.871	28.90	56.35	16.28	0.3629

R123	3.766	1.934	28.43	55.76	15.85	0.3587

Table-5.(c)Exergy Destruction of various components based on exergy of fuel of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_EVA_1=263"K", T_EVA_2=278"K", T_EVA_3=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	Total % Losses	Second Law efficiency
R12	8.06	3.638	22.58	17.94	0.6216	0.3784
R134a	8.462	3.875	21.96	18.06	0.6285	0.3715
R1234yf	8.317	4.015	21.71	18.62	0.6440	0.356
R1234ze	9.74	3.968	20.99	18.51	0.640	0.360
R-32	0.3281	3.796	27.23	16.47	0.5792	0.4208
R227ea	11.37	4.915	19.35	19.03	0.6705	0.3295
R236fa	12.25	3.916	20.35	18.33	0.6626	0.3374
R245fa	10.84	3.004	21.81	18.42	0.6371	0.3629
R123	11.24	2.782	22.94	18.05	0.6413	0.3587

Table-5.(d) Exergy Destruction of various components based on Total exergy destruction of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_EVA_1=263"K", T_EVA_2=278"K", T_EVA_3=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp
R12	15.44	6.968	43.24	34.36
R134a	16.16	7.402	41.94	34.49
R1234yf	15.79	7.624	41.22	35.36
R1234ze	18.30	7.457	39.45	34.79
R-32	12.23	7.972	56.75	34.59
R227ea	10.58	7.078	35.4	34.81
R236fa	12.12	8.991	36.78	34.04
R245fa	20.04	7.078	40.33	34.07
R123	20.43	5.057	41.70	32.81

System-6: Vapour compressor refrigeration system with multiple evaporators at different temperatures with individual compressors, individual expansion valves

We considered vapour compressor refrigeration system with multiple evaporators at different temperatures with compound compression, individual expansion valves (System-6) for energy-exergy- analysis. To validate computational results from developed thermal model of system-6, the following input values have been taken. The computed results of system-6 for 100% compressors efficiency were compared and shown in Table-6(a) respectively. For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects)in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves are shown in Table-6(b) to Table-6 (d) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances

Table-6(a): Validation of Results of vapour compressor refrigeration system with multiple evaporators at different temperatures with individual compressors, individual expansion valves for 100% compressor efficiency using following loads Q_EVA_1=70"kW" Q_EVA_2=105"kW" Q_EVA_3=35"kW"

Parameter	Program	Ref [12]
COP	5.938	5.88
Total Work (KW)	35.37	35.72

Table-6(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of ofvapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_EVA_1=263"K", T_EVA_2=278"K", T_EVA_3=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	First law Efficiency (System COP)	EDR	% Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	Second Law efficiency
R12	4.505	2.186	31.36	46.62	14.62	0.5995
R134a	4.399	2.264	30.72	47.74	14.62	0.5854

R1234yf	4.204	2.418	29.26	49.96	14.62	0.5594
R1234ze	4.379	2.281	30.48	47.95	14.62	0.5828
R-32	4.336	2.298	30.18	48.43	14.62	0.5770
R227ea	4.018	2.575	27.97	52.26	14.62	0.5348
R236fa	4.329	2.319	30.13	48.51	14.62	0.5761
R245fa	4.612	2.114	32.10	45.53	14.62	0.6138

Table-6.(c) Exergy Destruction of various components based on exergy of fuel of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) $T_{_EVA_}1$=263"K", $T_{_EVA_}2$=278"K", $T_{_EVA_}3$=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	Total % Losses	Second Law efficiency
R12	8.81	13.73	26.23	19.29	58.64	31.36
R134a	8.572	15.2	26.18	19.35	69.38	30.72
R1234yf	6.854	17.24	27.06	19.59	70.74	29.26
R1234ze	8.178	15.87	25.84	19.63	69.52	30.48
R-32	8.479	13.89	28.60	18.38	69.82	30.18
R227ea	7.622	20.2	24.55	19.66	72.03	27.97
R236fa	8.465	16.21	25.54	19.65	69.87	30.13
R245fa	8.996	12.54	26.74	19.59	67.90	32.10

Table-6.(d) Exergy Destruction of various components based on Total exergy destruction of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) $T_{_EVA_}1$=263"K", $T_{_EVA_}2$=278"K", $T_{_EVA_}3$=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp
R12	12.85	20.02	38.99	26.14
R134a	12.27	21.93	37.78	27.92
R1234yf	9.685	24.36	38.25	27.69
R1234ze	11.76	22.83	37.17	28.23
R-32	12.23	20.03	41.27	26.51
R227ea	10.58	28.05	34.08	27.29
R236fa	12.12	23.2	36.55	28.13
R245fa	13.26	18.48	39.4	28.81

System-7: Vapour compressor refrigeration system with Multiple evaporators at different temperatures with individual compressors, individual expansion valves for different evaporator loads.

($Q_{_EVA_1}$=70"kW" $Q_{_EVA_2}$=105"kW" $Q_{_EVA_3}$=35"kW" $T_{_EVA_}1$=263"K", $T_{_EVA_}2$=278"K", $T_{_EVA_}3$=283"K).

Thermal analysis of vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves (System-7) is carried out . To validate computational results from developed thermal model of system-7, the following input values have been taken. The computed results of system-7 for 100% compressors efficiency were compared and shown in Table-7(a) respectively. For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects)in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves are shown in Table-7(b) to Table-7 (d) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances

Table-7(a) : Validation of Results of vapour compressor refrigeration system with Multiple evaporators at different temperatures with individual compressors, individual expansion valves($Q_{_EVA_1}$=70"kW" $Q_{_EVA_2}$=105"kW" $Q_{_EVA_3}$=35"kW" for 100% compressor efficiency

Parameter	Program	Ref [12]
COP	5.683	5.96
Total Work (KW)	36.95	35.24

Table-7(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of vapour compression refrigeration system using alternative refrigerants for actual conditions: Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80 $T_{_EVA}_1$=263"K", $T_{_EVA}_2$=278"K", $T_{_EVA}_3$=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	First law Efficiency (System COP)	EDR	Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	Secnd Law Efficiency
R12	4.505	2.186	0.3136	46.62	14.62	0.5995
R134a	4.399	2.264	0.3062	47.74	14.62	0.5854
R1234yf	4.204	2.418	0.2926	49.96	14.62	0..5594
R1234ze	4.379	2.281	0.3048	47.95	14.62	0.5828
R-32	4.336	2.298	0.3018	48.43	14.62	0.5770
R227ea	4.018	2.575	0.2797	52.26	14.62	0.5348
R236fa	4.329	2.319	0.3013	48.51	14.62	0.5761
R245fa	4.612	2.114	0.3210	45.53	14.62	0.6138
R123	4.714	2.045	0.3281	44.55	14.62	0.6273

Table-7.(c) Exergy Destruction of various components based on exergy of fuel of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80), $T_{_EVA}_1$=263"K", $T_{_EVA}_2$=278"K", $T_{_EVA}_3$=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	Total % Losses	Second Law efficiency
R12	8.81	13.73	26.73	19.29	68.56	0.3144
R134a	8.572	15.20	26.18	19.35	69.31	0.3069
R1234yf	6.854	17.24	27.06	19.59	70.75	0.2925
R1234ze	8.1781	15.87	25.84	19.63	69.51	0.3049
R-32	8.479	13.89	28.60	18.38	69.35	0.3065
R227ea	7.622	20.20	24.55	19.66	72.03	0.2797
R236fa	8.465	16.21	25.54	19.65	69.87	0.3013
R245fa	8.996	12.54	26.74	19.59	67.87	0.3213
R123	9.039	11.13	27.52	19.42	67.11	0.3289

Table-7.(d) Exergy Destruction of various components based on Total exergy destruction of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80)

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp
R12	12.85	20.02	38.99	28.14
R134a	12.37	21.93	37.78	27.92
R1234yf	9.689	24.36	38.25	27.69
R1234ze	11.76	22.83	37.17	28.23
R-32	12.23	20.03	41.24	26.51
R227ea	10.58	28.05	34.08	27.29
R236fa	12.12	23.20	36.55	28.13
R245fa	13.26	18.48	39.4	28.87
R123	13.47	16.59	41.0	28.94

System-8(a): Vapour compressor refrigeration system with multiple evaporators at different temperatures with individual compressors, multiple expansion valves

We considered vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves (System-8). To validate computational results from developed thermal model of system-8, the following input values have been taken. The computed results of system-8 for 100% compressors efficiency were compared and shown in Table-8(a) respectively. For finding thermal performances in terms of first law efficiency (COP), EDR(Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects)in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves are shown in Table-8(b) to Table-8 (d) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances.

Table-8(a): Validation of Results of VCRS for 100% compressor efficiency (Q_EVA_1=70"kW" Q_EVA_2=105"kW" Q_EVA_3=35"kW")

Parameter	*Program*	*Ref [13]*
COP	*6.485*	*6.25*
Total Work (KW)	*32.38*	*33.6*

Table-8(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of vapour compression refrigeration system using alternative refrigerants for actual conditions(Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80), $T_{_EVA}_1$=263"K", $T_{_EVA}_2$=278"K", $T_{_EVA}_3$=283"K",T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	First law Efficiency (System COP)	EDR	% Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	Secnd Law Efficiency
R12	5.188	1.81	36.11	40.48	14.62	0.6904
R134a	5.113	1.825	36.01	40.59	14.62	0.6885
R1234yf	5.127	1.848	35.69	40.96	14.62	0.6824
R1234ze	5.213	1.810	36.26	40.29	14.62	0.6937
R-32	4.894	1.958	34.06	42.91	14.62	0.6515
R227ea	5.072	1.898	35.30	41.40	14.62	0.675
R236fa	5.207	1.824	36.24	40.33	14.62	0.693
R245fa	5.318	1.758	37.02	40.48	14.62	0.7078
R123	5.331	1.742	37.61	39.49	14.62	0.7095

Table-8 (c) Exergy Destruction of various components based on exergy of fuel of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) $T_{_EVA}_1$=263"K", $T_{_EVA}_2$=278"K", $T_{_EVA}_3$=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	Total % Losses	Second Law efficiency
R12	13.66	4.847	28.35	18.5	65.35	0.3465
R134a	14.02	5.219	27.88	18.59	65.70	0.3430
R1234yf	12.69	5.475	28.30	18.97	65.96	0.3404
R1234ze	13.95	5.337	27.39	19.0	65.67	0.3437
R-32	12.85	5.116	31.80	17.12	66.68	0.3332
R227ea	14.83	6.765	26.4	19.02	67.01	0.3299
R236fa	14.68	5.273	27.13	19.04	66.12	0.3388
R245fa	14.04	5.116	28.11	18.97	65.09	0.3491
R123	13.41	3.634	28.88	18.73	64.66	0.3534

Table-8.(d) Exergy Destruction of various components based on Total exergy destruction of vapour compression refrigeration system using alternative refrigerants(for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) $T_{_EVA}_1$=263"K", $T_{_EVA}_2$=278"K", $T_{_EVA}_3$=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp
R12	20.90	7.416	43.37	28.31
R134a	21.33	7.941	42..34	28.30
R1234yf	19.21	8.301	43.71	28.76
R1234ze	21.24	8.127	41.70	28.93
R-32	18.96	7.671	47.69	25.67
R227ea	22.12	10.09	39.40	28.38
R236fa	22.2	7.975	41.03	28.80
R245fa	21.57	6.096	43.19	29.15
R123	20.74	5.057	41.70	28.31

System-9: Vapour compressor refrigeration system with multiple evaporators at different temperatures with compound compression, individual expansion valves and flash intercoolers

We considered vapour compressor refrigeration system with multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers (System-9) for thermal (energy-exergy analysis. To validate computational results from developed thermal model of system-9, the following input values have been taken. The computed results of system-9 for 100% compressors efficiency were compared and shown in Table-9(a) respectively.

For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects)in terms of irreversibility of system occurred in the

components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers are shown in Table-9(b) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances.

Table-9(a): Validation of Results of Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves and flash intercoolers for 100% compressor efficiency (Q_{EVA_1}=70"kW" Q_{EVA_2}=105"kW" Q_{EVA_3}=35"kW

Parameter	Program	Ref [10]
COP	5.683	5.960
Total Work (KW)	36.95	35.24

Table 9-.(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of ofvapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_{EVA_1}=263"K", T_{EVA_2}=278"K", T_{EVA_3}=283"K", T_R_1=268"K", T_R _2=283"K", T_R_3=288"K",

Refrigerants	First law Efficiency (System COP)	EDR	Exergetic Efficiency	Exergy_ Fuel (KW)	Exergy_ Product (KW)	% Exergetic Efficiency_II	Second Law Efficiency	%Total Exergy Defect
R12	4.723	1.811	0.3558	44.47	15.82	0.3598	0.5286	64.42
R134a	4.676	1.839	0.3523	44.91	15.82	0.3523	0.5235	64.77
R1234yf	4.582	1.897	0.3451	45.83	15.82	0.3451	0.5129	65.49
R1234ze	4.691	1.83	0.3534	44.76	15.82	0.3534	0.5251	64.66
R227ea	4.481	1.963	0.3375	46.87	15.82	0.3375	0.5016	66.25
R236fa	4.668	1.844	0.3516	44.99	15.82	0.3516	0.5225	64.89
R245fa	4.835	1.746	0.3642	43.43	15.82	0.3642	0.5412	63.58

***System-10**:* Vapour compressor refrigeration system with multiple evaporators at different temperatures with compound compression, individual expansion valves and flash intercoolers

We considered vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers (System-10).For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects) in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers are shown in Table-10(a) and Table-10 (b) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances.

Table 10-.(a) Thermal Performances (First law efficiency and Second law efficiency, etc.) of vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves and flash intercoolers using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_{EVA_1}=268"K", T_{EVA_2}=273"K", T_{EVA_3}=278"K", T_R_1=268"K", T_R _2=283"K", T_R_3=288"K", Q_EVA_1=35"kW,Q_EVA_2=70"kW" Q_EVA_3=105"kW"

Refrigerants	First law Efficiency (System COP)	EDR	Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	%Total Exergy Defect	% Exergetic Efficiency_II
R12	5.26	1.892	0.3458	39.93	13.81	0.6625	0.5888
R134a	5.091	1.988	0.3347	41.25	13.81	0.6653	0.5699
R1234yf	5.0	2.042	0.3287	42.0	13.81	0.6713	0.5597
R1234ze	5.112	1.975	0.3361	41.08	13.81	0.6639	0.5722
R227ea	4.902	2.103	0.3223	42.84	13.81	0.6777	0.5488
R236fa	5.093	1.986	0.3348	41.23	13.81	0.6652	0.5701
R245fa	5.134	1.983	0.3375	40.9	13.81	0.6542	0.5747

Table 10-.(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves and flash intercoolers using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_EVA_1=268"K", T_EVA_2=273"K", T_EVA_3=278"K", T_R_1=268"K", T_R _2=283"K", T_R_3=288"K", Q_EVA_1=35"kW,Q_EVA_2=70"kW" Q_EVA_3=105"kW"

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	% loss sub cooler	%Total Exergy Defect	% Exergetic Efficiency_II
R12	9.960	7.448	27.91	18.875	2.039	0.6625	0.5888
R134a	9.841	8.17	27.35	18.94	2.208	0.6653	0.5699
R1234yf	8.063	8.99	28.21	19.23	2.631	0.6713	0.5597
R1234ze	9.484	8.49	26.91	19.27	2.23	0.6639	0.5722
R227ea	9.22	10.79	25.67	19.29	2.806	0.6777	0.5488
R236fa	9.881	8.617	26.62	19.3	2.095	0.6652	0.5701
R245fa	10.18	6.622	27.8	19.25	1.553	0.6542	0.5747

System-11*:* We considered vapour compressor refrigeration system with multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers (System-11) for thermodynamic analysis. For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects) in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers are shown in Table-11(a) and Table-11 (b) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances.

Table 11-.(a) Thermal Performances (First law efficiency and Second law efficiency, etc.) of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_EVA_1=263"K", T_EVA_2=278"K", T_EVA_3=283"K", T_R_1=268"K", T_R _2=283"K", T_R_3=288"K", Q_EVA_1=105"kW" Q_EVA_2=35"kW ,Q_EVA_3=70"kW"

XX

Refrigerants	First law Efficiency (System COP)	EDR	Exergetic Efficiency	Exergy_ Fuel (KW)	Exergy_Product (KW)	%Total Exergy Defect	Rational Efficiency
R12	4.713	1.813	0.3555	44.56	15.84	64.45	0.3448
R134a	4.666	1.841	0.3520	45.01	15.84	64.8	0.3448
R1234yf	4.57	1.90	0.3448	45.95	15.84	65.52	0.3448
R1234ze	4.68	1.832	0.3531	44.87	15.84	64.69	0.3448
R227ea	4.467	1.967	0.3371	47.01	15.84	66.29	0.3448
R236fa	4.656	1.847	0.3513	45.11	15.84	64.87	0.3448
R245fa	4.824	1.747	0.3640	43.53	15.84	63.60	0.3448

Table 11-.(b) Exergy Destruction of various components based on exergy of product of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80), T_EVA_1=263"K", T_EVA_2=278"K", T_EVA_3=283"K", T_R_1=268"K", T_R _2=283"K", T_R_3=288"K", Q_EVA_1=105"kW" Q_EVA_2=35"kW ,Q_EVA_3=70"kW"

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	% loss sub cooler	%Total Exergy Defect	Rational Efficiency_II	% Exergetic Efficiency_II
R12	9.312	8.156	25.99	19.06	1.899	64.45	0.3555	0.5275
R134a	4.733	8.962	25.44	19.13	2.054	64.8	0.3520	0.5235
R1234yf	7.563	9.921	26.19	19.41	2.444	65.52	0.3448	0.5129
R1234ze	8.816	9.351	25.01	19.44	2.073	64.69	0.3531	0.5251
R227ea	8.571	11.88	23.78	19.46	2.599	66.29	0.3371	0.5016
R236fa	9.21	9.545	24.71	19.47	1.945	64.87	0.3513	0.5225
R245fa	9.494	7.367	25.87	19.2	1.445	63.60	0.364	0.5412

System-12

We considered vapour compressor refrigeration system with multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers (System-12). To validate computational results from developed thermal model of system-12, the following input values have been taken. The computed results of system-13 for 100% compressors efficiency were compared and shown in Table-12 respectively. For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy

Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects) in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers are shown in Table-12(b) and 12 (c) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances.

Table-12(a): Validation of Results of Vapour compression refrigeration system with Multiple evaporators at different temperatures with compound compression, individual expansion valves and flash intercoolers for 100% compressor efficiency using following loads (Q_{EVA_1}=35"kW" Q_{EVA_2}=70"kW" Q_{EVA_3}=105"kW") at following temperatures (T_{EVA_1}=263"K", T_{EVA_2}=278"K", T_{EVA_3}=283"K)

Parameter	Program	Ref [12]
COP	6.44	6.50
Total Work (KW)	32.61	32.77

Table-12(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of ofvapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_{EVA_1}=263"K", T_{EVA_2}=278"K", T_{EVA_3}=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	First law Efficiency (System COP)	EDR	Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	Rational efficiency	Second Law efficiency
R12	5.134	1.963	0.3375	40.9	13.81	0.3375	0.5747
R134a	5.091	1.988	0.3347	41.25	13.81	0.3347	0.5699
R1234yf	5.0	2.042	0.3287	42.0	13.81	0.3287	0.5597
R1234ze	5.112	1.975	0.3361	41.08	13.81	0.3361	0.5722
R-32	4.89	2.111	0.3215	42.95	13.81	0.3215	0.5473
R227ea	4.902	2.103	0.3223	42.84	13.81	0.3223	0.5488
R236fa	5.093	1.986	0.3346	41.23	13.81	0.3348	0.5701
R245fa	5.26	1.892	0.3458	39.93	13.81	0.3458	0.5888
R123	5.299	1.87	0.3484	39.63	13.81	0.3484	0.5932

Table-12.(c) Exergy Destruction of various components based on exergy of fuel of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_{EVA_1}=263"K", T_{EVA_2}=278"K", T_{EVA_3}=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	% Loss_ Sub_cooler	% Loss_ (F1+F2)	% Total Loss/ Irreversibility
R12	10.34	7.058	27.91	18.87	2.039	0.02573	66.25
R134a	10.26	7.744	27.35	18.94	2.208	0.0240	66.53
R1234yf	8.505	8.545	28.21	19.23	2.632	0.00107	67.13
R1234ze	9.905	8.066	26.91	19.27	2.231	0.00393	66.39
R-32	9.892	7.104	30.63	17.76	2.366	0.1042	67.85
R227ea	9.733	10.28	25.67	19.29	2.806	0.0059	67.77
R236fa	10.30	8.199	26.62	19.30	2.095	0.000753	66.52
R245fa	10.51	6.29	27.8	19.25	1.553	0.009836	65.42
R123	10.38	5.634	28.63	19.06	1.429	0.02436	65.16

We considered vapour compressor refrigeration system with multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers (System-13). To validate computational results from developed thermal model of system-13, the following input values have been taken. The computed results of system-13 for 100% compressors efficiency were compared and shown in Table-13 respectively. For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects) in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers are shown in Table-13(b) and 13 (c) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances.

Table-13(a):Validation of Results of Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers for 100% compressor efficiency using following loads Q_{EVA_1}=105"kW" Q_{EVA_2}=70"kW" Q_{EVA_3}=35"kW"

Parameter	Program	Ref [10]
COP	5.794	4.90
Total Work (KW)	36.25	42.64

Table-13(b): Thermal Performances (First law efficiency and Second law efficiency, etc.) of ofvapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_{EVA}_1=263"K", T_{EVA}_2=278"K", T_{EVA}_3=283"K", T_R_1=268"K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	First law Efficiency (System COP)	EDR	Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	Rational efficiency	Second Law efficiency
R12	4.589	1.143	0.4860	45.76	22.24	0.4445	0.6107
R134a	4.576	1.159	0.4846	45.89	22.24	0.4382	0.6089
R1234yf	4.537	1.188	0.4805	46.29	22.24	0.4293	0.6038
R-32	4.396	1.227	0.4656	47.77	22.24	0.4289	0.5851
R227ea	4.492	1.231	0.4757	46.75	22.24	0.4145	0.5978
R236fa	4.616	1.159	0.4877	45.60	22.24	0.4347	0.6129
R245fa	4.70	1.096	0.4977	44.68	22.24	0.4544	0.6255
R123	4.709	1.081	0.4986	44.6	22.24	0.4608	0.6266

Table-13.(c) Exergy Destruction of various components based on exergy of fuel of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) T_{EVA}_1=263"K", T_{EVA}_2=278"K", T_{EVA}_3=283"K", T_R_1= 268 "K", T_R_2=283"K", T_R_3=288"K",

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	% Loss_ Subcooler	% Loss_ (F1+F2)	% Total Loss /Irreversibility
R12	4.146	4.748	25,27	19.41	1.876	0.08344	57.54
R134a	4.599	5.092	24.91	19.46	2.039	0.06972	56.17
R1234yf	3.686	5.247	26.03	19.68	2.438	0.00255	57.03
R-32	3.66	5.193	26.97	18.84	2.223	0.4131	57.01
R227ea	5.872	6.463	23.90	19.72	2.596	0.0001	58.55
R236fa	5.294	3.837	24.5	19.74	1.926	0.001834	56.53
R245fa	4.272	5.066	25.3	19.70	1.419	0.03014	54.56
R123	3.574	3.535	25.87	19.55	1.302	0.07858	53.91

System-14: The thermal performances in terms of first law efficiency using energy-exergy analysis is presented in Table-14(a) to Table-14 (d) respectively .Table-14(b)to Table-14(d) represented irreversibility computation using entropy generation principles. The various losses occurred in terms of exergy destruction in components (Exergy defects) in the system components shown in Table-14(b),to Table-14(d) respectively. The considered vapour compressor refrigeration system with multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers (System-14) is used for thermodynamic analysis. To validate computational results from developed thermal model of system-14, the following input values have been taken. The computed results of system-14 for 100% compressors efficiency were compared and shown in Table-14(a) respectively. For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects) in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers are shown in Table-14(b) to table-14 (d) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances

Table-14(a) : *) Thermal Performances (First law efficiency and Second law efficiency, etc.) of vapour compressor refrigeration system with Multiple evaporators at same temperature with single compressor, individual expansion valves using alternative refrigerants (Validation of Results of VCRS for 100% compressor efficiency) for given data: T_{EVA}_1=268"K", T_{EVA}_2=268"K", T_{EVA}_3=268"K", T_R_1=273"K", T_R_2=273"K", T_R_3=273"K", Q_EVA_1=35 "kW', Q_EVA_2=105"kW', Q_EVA_3=70"kW', T_cond=313"K"*

Parameter	*Program*	*Ref [13]*
COP	*5.232*	*5.07*
Total Work (KW)	*40.14*	*41.4*

Table-14(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of vapour compressor refrigeration system with Multiple evaporators at same temperature with single compressor, individual expansion valves *using alternative refrigerants (for Compressor efficiency_1=0.80, $T_{_EVA_}1$=268"K", $T_{_EVA_}2$=268"K", $T_{_EVA_}3$=268"K", T_R_1=273"K", T_R _2=273"K", T_R_3=273"K", Q_EVA_1=35 "kW', Q_EVA_2=105"kW', Q_EVA_3=70"kW', T_cond=313"K"T_subcooled=303"K"*

Refrigerants	First law Efficiency (System COP)	EDR	Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	Secnd Law Efficiency
R12	4.186	1.585	0.3833	50.17	19.23	0.4685
R134a	4.137	1.591	0.3789	50.76	19.23	0.4632
R1234yf	4.039	1.645	0.3698	52.0	19.23	0.4521
R1234ze	4.144	1.586	0.3795	50.68	19.23	0.4638
R-32	3.959	1.706	0.3626	53.04	19.23	0.4432
R227ea	3.927	1.715	0.3596	53.47	19.23	0.4396
R236fa	4.112	1.609	0.3766	51.07	19.23	0.4603
R245fa	4.281	1.518	0.3920	49.05	19.23	0.4792
R123	4.328	1.493	0.3964	48.52	19.23	0.4845

Table-14.(c)Exergy Destruction of various components based on exergy of fuel of vapour compressor refrigeration system with Multiple evaporators at same temperature with single compressor, individual expansion valves *using alternative refrigerants (for Compressor efficiency_1=0.80, $T_{_EVA_}1$=268"K", $T_{_EVA_}2$=268"K", $T_{_EVA_}3$=268"K", T_R_1=273"K", T_R _2=273"K", T_R_3=273"K", Q_EVA_1=35 "kW', Q_EVA_2=105"kW', Q_EVA_3=70"kW', T_cond=313"K"T_subcooled=303"K"*

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	Total % Losses	Rational Efficiency(%)
R12	8.521	9.039	24.0	18.42	59.98	40.02
R134a	8.394	9.961	23.41	18.52	60.28	39.72
R1234yf	6.966	11.17	23.72	18.96	60.82	39.18
R1234ze	8.0	10.5	22.71	18.98	60.19	39.81
R-32	8.061	8.620	28.31	16.86	61.85	38.15
R227ea	7.78	13.47	21.4	19.02	61.67	38.33
R236fa	8.372	10.83	22.34	19.04	60.59	39.41
R245fa	8.651	8.357	23.54	18.95	59.5	40.5
R123	8.617	7.451	24.42	18.68	61.85	38.15

Table-14.(d) Exergy Destruction of various components based on exergy of product of vapour compressor refrigeration system with Multiple evaporators at same temperature with single compressor, individual expansion valves *using alternative refrigerants (for Compressor efficiency_1=0.80, $T_{_EVA_}1$=268"K", $T_{_EVA_}2$=268"K", $T_{_EVA_}3$=268"K", T_R_1=273"K", T_R _2=273"K", T_R_3=273"K", Q_EVA_1=35 "kW', Q_EVA_2=105"kW', Q_EVA_3=70"kW', T_cond=313"K"T_subcooled=303"K"*

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	Total % Irreversibility	Secnd Law Efficiency
R12	14.21	15.07	40.01	30.17	59.98	0.4685
R134a	13.92	16.52	38.83	30.73	60.28	0.4632
R1234yf	11.45	18.37	38.99	31.18	60.82	0.4521
R1234ze	13.29	17.45	37.72	31.74	60.19	0.4638
R-32	13.03	13.95	45.77	27.26	61.85	0.4432
R227ea	12.61	21.84	34.70	30.84	61.67	0.4396
R236fa	13.82	17.88	36.87	31.43	60.59	0.4603
R245fa	14.54	14.04	39.57	31.85	59.5	0.4792
R123	14.56	12.9	41.27	31.57	58.52	0.4845

***System-15*:** Vapour compressor refrigeration system with Multiple evaporators at different temperatures with single compressor, individual expansion valves and back pressures valves. The considered vapour compressor refrigeration system with Multiple evaporators at different temperatures with compound compression, multiple expansion valves and flash intercoolers (System-15) is used for thermodynamic analysis. To validate computational results from developed thermal model of system-14, the following input values have been taken. The computed results of system-14 for 100% compressors efficiency were compared and shown in Table-15(a) respectively. For finding thermal performances in terms of first law efficiency (COP), EDR (Exergy Destruction Ratio based on exergy of input) exergetic efficiency using energy-exergy analysis. For finding various exergetic losses (Exergy defects) in terms of irreversibility of system occurred in the components and total exergy destruction in system along with rational efficiency and second law efficiency of the Vapour compressor refrigeration system with Multiple evaporators at different temperatures with

compound compression, multiple expansion valves and flash intercoolers are shown in Table-15(b) to table-15 (d) respectively. It was found that refrigerant R123 gives highest COP and R-245fa give slightly less performances

Table-15(a) Validation of Results of vapour compressor refrigeration system with Multiple evaporators at different temperatures with single compressor, individual expansion valves and back pressures valves using alternative refrigerants (for Compressor efficiency_1=1.0 $T_{_EVA}_1$=263"K", $T_{_EVA}_2$=273"K", $T_{_EVA}_3$=283"K", T_R_1=268"K", T_R _2=278"K", T_R_3=288"K", Q_EVA_1=35 "kW', Q_EVA_2=70"kW', Effect of super heating =5°C, Q_EVA_3=105"kW',T_subcooled=303"K"

Parameter	Program	Ref [10]
COP	4.162	4.38
Total Work (KW)	33.64	31.9

Table-15(b) Thermal Performances (First law efficiency and Second law efficiency, etc.) of vapour compressor refrigeration system with Multiple evaporators at different temperatures with single compressor, individual expansion valves and back pressures valves using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) $T_{_EVA}_1$=263"K", $T_{_EVA}_2$=278"K", $T_{_EVA}_3$=283"K", T_R_1=268"K", T_R _2=283"K", T_R_3=288"K", Q_EVA_1=35 "kW', Q_EVA_2=70"kW', Q_EVA_3=105"kW',T_subcooled=303"K"

Refrigerants	First law Efficiency (System COP)	EDR	Exergetic Efficiency	Exergy_Fuel (KW)	Exergy_Product (KW)	Secnd Law Efficiency	Rational efficiency
R12	3.33	2.62	0.2811	42.05	11.82	0.5991	0.3480
R134a	3.243	2.786	0.2738	43.16	11.82	0.5193	0.3390
R1234yf	3.10	3.117	0.2582	45.16	11.66	0.4925	0.3203
R1234ze	3.239	2.97	0.2632	43.22	11.38	0.5075	0.3279
R227ea	2.948	3.692	0.2351	47.5	11.16	0.457	0.2938
R236fa	3.197	3.295	0.2441	43.8	11.69	0.4838	0.3074
R245fa	3.423	2.741	0.2669	40.9	11.92	0.5241	0.3349

Table-15.(c) Exergy Destruction of various components based on Total exergy destruction of vapour compression refrigeration system using alternative refrigerants (for Compressor efficiency_1= Compressor efficiency_2= Compressor efficiency_3=0.80) $T_{_EVA}_1$=263"K", $T_{_EVA}_2$=278"K", $T_{_EVA}_3$=283"K", T_R_1=268"K", T_R _2=283"K", T_R_3=288"K", Q_EVA_1=35 "kW', Q_EVA_2=70"kW', Q_EVA_3=105"kW',T_subcooled=303"K"

Refrigerants	% loss Eva	% loss valve	% loss Condenser	% loss comp	% Irreversibility= Total Exergy Losss/ Exergyof Fuel	Rational Efficiency
R12	2.936	42.06	30.65	24.36	65.20	0.3480
R134a	2.892	44.87	28.75	23.68	66.10	0.3390
R1234yf	1.431	48.48	26.91	23.18	67.97	0.3203
R1234ze	3.421	46.08	26.8	23.70	67.21	o.3279
R227ea	3.15	52.89	22.22	21.94	70.62	0.2938
R236fa	5.455	45.86	25.25	23.45	69.26	0.3074
R245fa	5.868	39.14	29.97	25.20	66.51	0.3349

4. Conclusion

The following conclusions were drawn

(i) Use of HFO refrigerants for Replacing HFC -134a is more significant due to low GWP as compared to R134a of 1430

(ii) Thermal performances using HFO refrigerants is slightly less than by using R134a although R123 refrigerant gives higher first law efficiency(COP)

(iii) By altering evaporator load effects thermal performances considerably.

(iv) The maximum exergy destruction in the system components it is found maximum (i.e. higher) in the condenser and low in the throttle valves.

(v) The percentage exergy destruction in the evaporator varies with the type of refrigerants. It becomes high when R236fa is used and lower when R32 is used

References

[1] MishraR.S.[2014],Appropriate Vapour CompressionRefrigeration Technologies for Sustainable Development International Journal of Advance Research & Innovations, 2014, 551-556.

[2] Chopra K., Sahni V., Mishra R. S.[2014], Thermodynamic Analysis of Multiple evaporators vapour compression Refrigeration Systems with R-410a, R290, R1234yf, R502, R404a, R152a and R134a, International Journal of Air conditioning and Refrigeration, 22(1), 2014, 14.

[3] Chopra K., Sahni V., Mishra R. S. [2014], Methods for improving first and second law efficiencies of vapour compression refrigeration systems using flash intercooler with ecofriendly refrigerants, 1, 2014, 50- 64.

[4] D. Yumrutas et. Al[2002] Exergy analysis of vapour compression refrigeration systems, Exergy, an Int J, 2 (2002) 266-272.

[5] Dincer I.[2003] Refrigeration systems and applications. London: Wiley; 2003. pp. 26–27.

[6] Kumar S, Prevost M, Bugarel R. [1989]Exergy analysis of a compression refrigeration system. Heat Recovery Syst CHP. 1989;9(2):151–157.

[7] Nikolaidis C, Probert D[1998]. Exergy-method analysis of a two stage vapor compression refrigeration plants performance. Appl Energy. 1998;60:241–256.
[8] Bejan A,[1989] Theory of heat transfer irreversible refrigeration plants, Int J Heat mass transfer, 32(9), (1989), 1631-1639.
[9] Mishra R. S.[2013], Irriversibility Analysis of MultiEvaporators Vapour Compression Refrigeration Systems Using New and Refrigerants: R134a, R290, R600, R600a, R1234yf,R502, R404a and R152a and R12, R502" International Journal Advance Research & Innovations, 1, 2013, 180-193.
[10] Khurmi R.S. et.al [2014] A text book of Refrigeration and air conditioning, Eurasia publication House (P) Limited, , New Delhi.

To Determine the Effect of Dimpled Poppet Valve on the Flow in Engine Cylinder by Port Flow Simulation

Anirudh Singh[1], Apoorv Dubey[1] and Raj Kumar Singh[1]

[1]*Mechanical Engineering Department, Delhi Technological University, Delhi 42, India*

Abstract. Intake system is one of the crucial sub-systems in engine, the geometrical configuration of the inlet ports and the valves and their opening schedule creates organized motions in the cylinder, known as swirl (about the cylinder axis) and tumble (orthogonal to the cylinder axis) which help in proper mixing of fuel and air. Surface texturing of intake valve can improve the performance of engine drastically by generating more swirl, turbulent kinetic energy and reducing the coefficient of skin friction in the valve proximity area. Making dimples on intake valve is one of the many ways in which surface texturing can be done and its effectiveness can be easily measured by conducting a steady-state port flow simulation on ANSYS IC fluent solver using standard k-omega model. Port flow analysis was carried on dimpled intake valve at a lift of 5mm for 5 dimple ratio ranging from 0.1 to 0.3 in steps of 0.05. The dimple depth-over diameter ratio is a very crucial parameter in determining the optimum geometry of dimple on which best results are obtained. The results showed that the maximum mass flow rate is obtained at dimple ratio of 0.2 with an increase of 0.005% over the simple intake valve and a 38.4% decrease in the area weighted average skin friction coefficient in valve proximity area was seen in case of dimpled valve.

1 Introduction

The 4 stroke internal combustion engine consists of four cycles namely intake, compression, power stroke and exhaust. Out of these four processes, intake and compression strokes are the most important as these strokes have great effect on determining the pattern of air flow coming inside cylinder during intake stroke and the condition needed for the proper vaporization and mixing of fuel during fuel injection in compression stroke[1]. For maximizing the air passing into the cylinder, intake valve design has to be optimized. Hence optimum design of intake valve is crucial in engine design. Xiaofeng Yang, Zhaohui Chen, and Tang-Wei Kuo described that the development and optimization of intake system is to find the best trade-off between a) minimum flow restriction (discharge coefficient or volumetric efficiency) and b) an optimal in-cylinder charge motion (swirl or tumble ratios) [2]. Chenbo Ma, Jianjun Sun, Yan Wang , Bo Yu , Qiuping Yu and Qiaoan Tu stated surface texturing as a viable technique to significantly enhance the load-carrying capacity, friction coefficient, and wear resistance of vital components and to maximize the effect of surface texturing, many theoretical and experimental studies have been carried out that concentrate on the geometrical parameters of textures[3]. Etsion I. stated that to fully benefit from surface texturing, a proper optimization of the geometrical parameters, including the dimple depth-over diameter ratio (DOD) and area density, must be performed [4]. Therefore, the determination of the optimum dimple DOD ratio is very important. This concept of surface texturing was used to introduce dimples on the poppet valve and then measuring its effectiveness by conducting the Port flow simulation on ANSYS IC Fluent solver.

2 Flow in cylinder

All fluid flow is either laminar or turbulent, whether a flow is laminar or turbulent is usually determined by the Reynolds number. The flow in the cylinder is divided into several distinct phases. Firstly, the geometrical configuration of the intake system and the opening schedule of the inlet valves creates organized motions in the cylinder, known as swirl (about the cylinder axis) and tumble (orthogonal to the cylinder axis) and second, the jet itself is turbulent, and in addition much of the directed (non-turbulent) energy in the jet is converted to turbulence, resulting in a very high turbulence level during the inlet stroke[5]. The flow through the inlet valve forms a hollow conical jet, which is roughly equal to the angle of the valve seat.

An air flow bench is a device used for testing the internal aerodynamic qualities of an engine component and is used primarily for testing the intake and exhaust ports of cylinder heads of internal combustion engines. A

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

flow bench consists of an air pump, a metering element, temperature and pressure measuring instruments such as manometers, and various controls. The test piece is connected in series with the pump and measuring element and air is pumped through the whole system. As the volume flow rate through the metering element is known which is equal to the flow through the test piece, the mass flow rate can be calculated using the known pressure and temperature data to calculate air densities, and then multiplied by the volume flow rate. Most flow testing is done between 10 and 28 inches of water pressure (2.5 to 7 kilopascals).

Nur Adila Mohamad Shafie et al.[6] showed that the minimum and maximum deviation as small as 1.6% and 5.1% respectively were obtained when they compared the results obtained by port flow simulation and those obtained from engine flow bench test(experimental test) which shows that port flow simulation can accurately predict the breathing parameter of engine.

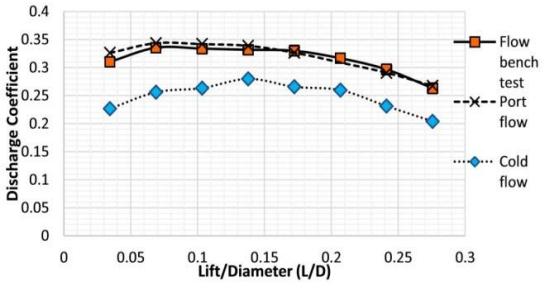


Figure 1. Deviation in results obtained from Port flow simulation and engine flow bench (experimental test) [6]

3 Methodology

A simplified design of port has been used for conducting this simulation, the design of intake plenum, outplenum, cylinder and port is kept same for all the cases in order to compare the effect of introducing dimples on the poppet valve over the simple valve. In case of the dimpled valves, two sets of 20 equidistant dimples were introduced on the valve. The print diameter of the dimples is kept at a constant of 1.2mm and the depth of the dimple is varied, governed by the dimple ratio h/D =c (here D=print diameter, h=depth, c=constant ranging from 0.1 to 0.3 in steps of 0.05). The port flow analysis in all the cases is carried out at an intake valve lift of 5mm and the exhaust port is deactivated by defining the exhaust valve lift to be 0mm[7].

The placement of dimples on poppet valve can be seen in fig. 2 for more better understanding of the geometry on which analysis is being carried upon, their configuration as well as their shape have a great effect on various parameters like drag, skin friction coefficient, standton number and reynolds number of the flow[8].

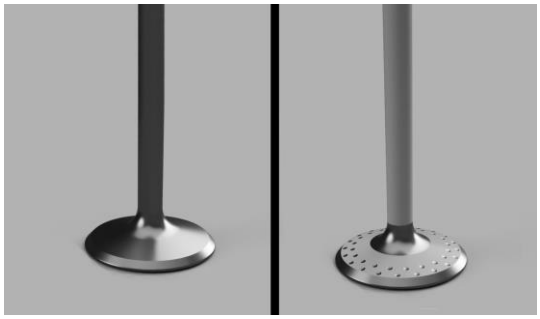


Figure 2. Simple valve and dimpled valve

3.1 Decomposition and meshing

The assembled model is then defined in the IC Input Manager in the design modeler and decomposed into the intake, exhaust ports and the cylinder. Four post planes (swirl planes) were added at a distance of 15mm, 30mm, 45mm and 60mm from the reference point for analyzing the swirl and velocity distribution due to introduction of dimples on poppet valve. Values of various parameters inserted in the input manager are shown in the Fig. 3.

The decomposed geometry is then meshed using a hybrid mesh topology. The maximum mesh size is reduced to 2.5mm, the number of inflation layers is increased to 12 and the chamber growth rate is reduced to 1.12. The size of the elements in the valve proximity area is kept at 0.2mm to capture the properties of flow precisely in that area. The inflation layers are essential for capturing the features of the boundary layers perfectly as prismatic elements provide good alignment with the flow near wall boundaries and thus special emphasis is laid on them. For regions other than the boundary layers, tetrahedral elements are predominant. The values of various mesh parameters are shown in Table 1.

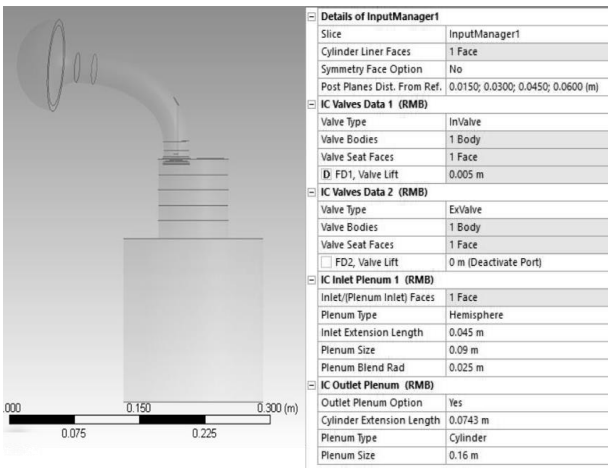


Figure 3. Decomposed geometry and input parameters

Table 1. IC Port Mesh Parameters

S.no	Mesh Parameters	Input value
1	Mesh type	Hybrid
2	Reference size(mm)	0.993
3	Maximum mesh size(mm)	2.5
4	Curvature normal angle(degree)	18
5	Number of cell across gap	5
6	Number of inflation layer	12
7	Chamber size(mm)	1.2
8	Chamber growth rate	1.12
9	Valve proximity face size(mm)	0.2

3.2 Solver Setup

For the purpose of simulation, standard k-omega model is used and the boundary conditions applied on various surfaces is shown in Table 2. An area weighted average surface monitor was set on the ice-int-chamber2-outplenum and a convergence monitor was applied on the plane with convergence criteria of $1*10^{-5}$ and the numbers of iterations to be ignored was set to 600. Different zones on which boundary condition is applied can be seen in Fig. 4.

Table 2. Boundary conditions

S.no	Zones	Type	Temperature	Gauge Pressure(Pa)
1	ice-slipwall-outplenum	wall	300 K	X
2	ice-slipwall-inplenum	wall	300K	X
3	ice-outlet	pressure outlet	300K	5000
4	ice-inlet-inplenum1	pressure inlet	300K	0

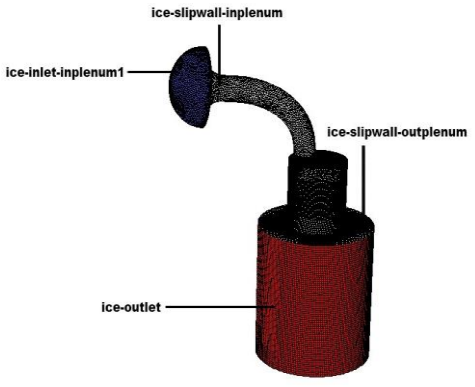


Figure 4. Different Zones on which boundary condition is applied

Once the solution had converged, obtained results were analyzed for all the case.

4 Result And Discussion

To compare the results obtained from port flow simulation conducted with simple intake poppet valve to that obtained from dimpled intake poppet valve with various dimple ratios, some key parameters were focused upon. In total four major parameters were studied to determine the effect of introduction of dimples on the in-cylinder flow. These parameters are as follows:

4.1 Mass Flow Rate

Mass flow rate is one of the many parameters that is given major focus upon as it determines the ability of intake port by which the engine can move the charge of air into and out of the cylinder. The results obtained after port flow simulation for dimple ratio from 0.1-0.3 showed that we get maximum mass flow rate of 34.663g/s at the dimple ratio of 0.2. Fig. 5 shows the variation of mass flow rate (g/s) to the dimple ratio (h/D). The curve shows that the mass flow rate increases for dimple ratio from 0.1 to maximum at 0.2 and the suddenly decreases for dimple ratio from 0.2 to 0.3. Thus it can be concluded that there is an optimum valve of dimple ratio that will increase our mass flow rate for the same port design. The mass flow rate obtained for the simple valve was 34.661g/s. Thus there is a small increase in mass flow rate of 0.005%.The value of mass flow rate obtained for other dimple ratios can be seen in Table 3.

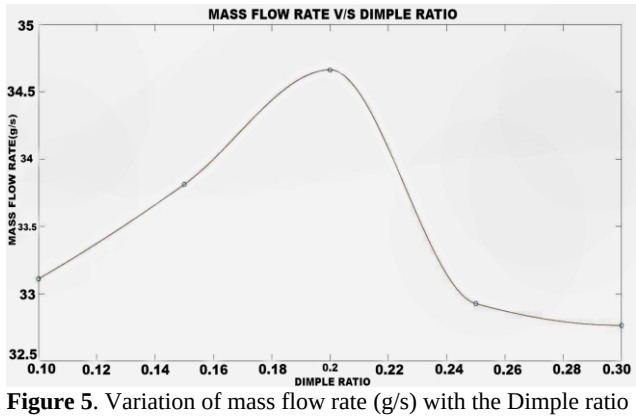


Figure 5. Variation of mass flow rate (g/s) with the Dimple ratio (h/D)

4.2 Turbulent Kinetic Energy

Mohd Aizad Sazrul Sabrudin, Mohd Farid Muhamad Said and Zulkarnain Abdul Latiff tried to study the effect of asymmetric intake valve lift configuration towards in-cylinder air flow behaviour by conducting a steady state port flow simulation and described that Turbulent kinetic energy (TKE) measurement on swirl monitoring planes is one of the methods that can be used to visualize the flow mixture behaviour inside the combustion chamber [9].

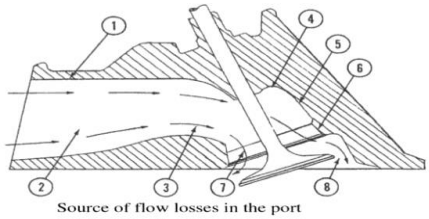
Instead of focussing on maximum or minimum turbulent kinetic energy (m^2/s^2) obtained on the swirl plane, focus on the area weighted average turbulent kinetic energy on the first swirl plane (SWIRL PLANE-1) was laid. Table 3 shows the value of area weighted average turbulent kinetic energy obtained on the first swirl plane monitor formed during the geometry decomposition process.

4.3 Swirl

There are a number of reasons for inducing swirl. High turbulence levels at ignition produce higher effective flame speeds, and more reliable combustion at very lean ratios, or with exhaust gas recirculation. At normal air/fuel ratios, the higher speed will allow the flame front to reach the end gas before the chemical reaction resulting in auto ignition has time to take place, permitting higher compression ratios without knock, the idea of swirl or tumble is to encapsulate some momentum of the inlet valve jet in the organized motion (swirl or tumble), which is less dissipative than the turbulence. The value of area weighted average swirl obtained on all four swirl planes for all the cases is shown in Table 3.

4.4 Coefficient of Skin Friction in Valve Proximity Area

According to Superflow Technologies Flowbench Operators' Manual and study conducted by Abdul Rahim Ismail et al.[10,11], source of flow losses in the port are wall friction, bend at valve guide, contraction at push-rod, expansion behind valve guide, expansion in 25 and 30 degrees, bend at the exit of the valve and expansion exiting valve. These sources of flow losses can be seen in Fig. 6.



Source of flow losses in the port

Source of flow loss	% of loss
1. Wall friction	4
2. Contraction at push-rod	2
3. Bend at valve guide	11
4. Expansion behind valve guide	4
5. Expansion 25 degrees	12
6. Expansion 30 degrees	19
7. Bend to exit valve	17
8. Expansion exiting valve	31
	100

Figure 6. Sources of flow losses in the port [10,11]

The area weighted average of skin friction coefficient in the valve proximity area was analyzed and it can be seen from Table 3. that for all values of dimple ratio(0.1-0.3), the area weighted average of skin friction coefficient in valve proximity area was less than that in case of simple intake poppet valve. Also, it can be seen from Fig.7 that as the dimple ratio is increased from 0.1 to 0.3 there is an increase in the coefficient of skin friction in the valve proximity area.

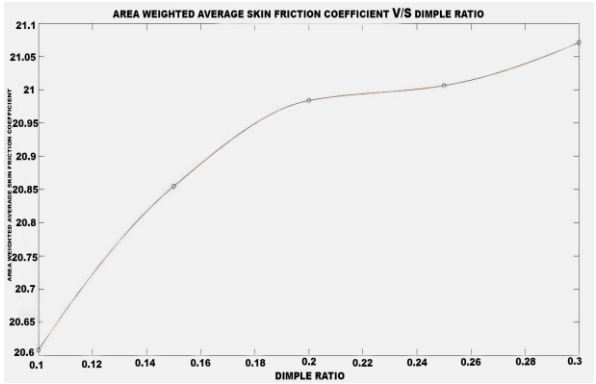


Figure 7. Variation of area-weighted average skin friction coefficient with dimple ratio

The lowest value of area weighted average skin friction coefficient(20.608) is obtained with a dimple ratio of 0.1, while the value of area weighted average skin friction coefficient obtained in case of simple valve is 33.507.Thus the skin friction coefficient of valve is decreased by 38.4% due to introduction of dimples.

The velocity distribution on all the swirl planes as well as cutplane for all the cases can be seen in Fig. 8. Major focus was given upon these four parameters to evaluate the effectiveness of the dimples created on the valve.

Table 3. Values of various parameters obtained for all the cases

Parameter	Simple Valve	Dimpled Valve				
		0.1	0.15	0.2	0.25	0.3
Mass flow rate(g/s)	34.661	33.111	33.812	34.663	32.927	32.765
Turbulent kinetic energy(m^2/s^2)	48.41	53.25	52.33	51.56	58.75	57.57
Skin friction coefficient	33.507	20.608	20.854	20.983	21.006	21.071
Swirl on Swirl plane-1	-0.02	0.007	0.022	0.023	0.015	0.036
Swirl on Swirl plane-2	-0.05	0.013	0.041	0.029	0.044	0.035
Swirl on Swirl plane-3	-0.04	0.0137	0.07	0.036	0.019	0.078
Swirl on Swirl plane-4	-0.01	0.046	0.018	0.055	0.011	0.12

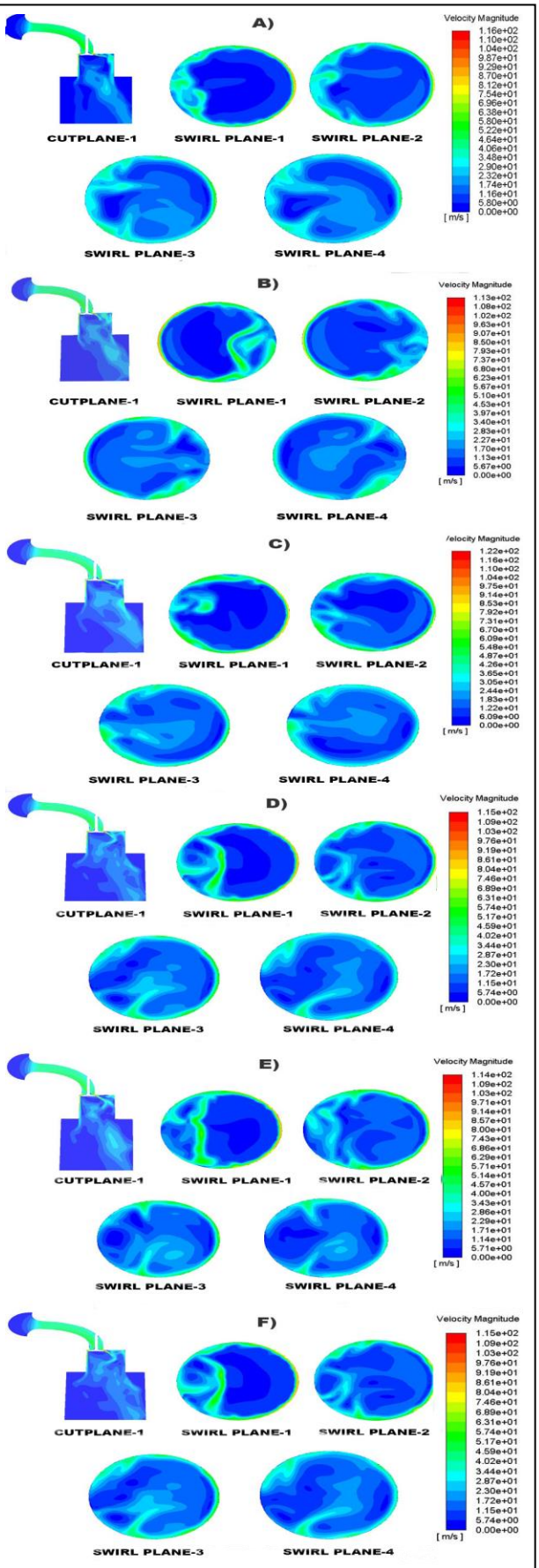


Figure 8. Velocity distribution on Swirl planes and Cutplane: A) simple valve, B) dimple ratio-0.1, C) dimple ratio-0.15, D) dimple ratio-0.2, E) dimple ratio-0.25, F) dimple ratio-0.3

5 Conclusion

In this work, the effect of introduction of dimples on the intake valve on the various performance characteristics of the engine was studied. The port flow simulation was carried with gauge outlet pressure of 5000Pa and the intake valve was given a lift of 5mm. Port flow analysis was carried on dimpled intake valve at a lift of 5mm for 5 dimple ratio ranging from 0.1 to 0.3 in steps of 0.05. The main conclusions are as follows:

1. The results obtained after port flow simulation for dimple ratio from 0.1-0.3 showed that we get maximum mass flow rate of 34.663g/s at the dimple ratio of 0.2. The mass flow rate increases for dimple ratio from 0.1 to maximum at 0.2 and the suddenly decreases for dimple ratio from 0.2 to 0.3.
2. The lowest value of area weighted average skin friction coefficient(20.608) is obtained with a dimple ratio of 0.1, while the value of area weighted average skin friction coefficient obtained in case of simple valve is 33.507.Thus the skin friction coefficient of valve is decreased by 38.4% due to introduction of dimples.
3. The value of area weighted Turbulent Kinetic energy (m^2/s^2) on SWIRL PLANE-1 was found to be greater in case of dimpled valve for every dimple ratio over the normal valve.
4. Higher value of swirl obtained on each swirl plane in case of dimpled valve shows the ability of the engine to mix air and fuel properly and effectively, that will enable a controlled as well as high speed flame front which will produce more power for the same quantity of the provided fuel.

References

[1] Heywood JB Internal Combustion Engine Fundamentals. Mcgraw-Hill; 1989
[2] Yang X, Chen Z, Kuo T Pitfalls for accurate steady-state port flow simulations. Journal of Engineering for Gas Turbines and Power 2013;135(6):061601–061601
[3] Ma C, Sun J, Wang Y, et al. On the optimum dimple depth-over-diameter ratio for textured surfaces. Adv. Mech Eng 2017; 9: 1–8.
[4] Etsion I, Kligerman Y and Halperin G. Analytical and experimental investigation of laser-textured mechanical seal faces. Tribol T 1999; 42: 511–516.
[5] FLOW IN THE CYLINDER by John L. Lumley, Cornell University (excerpt from his book: Engines – an introduction. Cambridge University Press, 1999)
[6] Nur Adila Mohamad Shafie, Mohd Farid Muhamad Said, Azhar Abdul Aziz, Zulkarnain Abdul Latiff, Ahmad Kamal Mat Yamin and Noreffendy Tamaldin MATEC Web Conf., 90 (2017) 01067, DOI: https://doi.org/10.1051/matecconf/20179001067

[7] Winroth, P.M., Ford, C.L. & Alfredsson, P.H. Exp Fluids (2018) 59: 24. DOI: https://doi.org/10.1007/s00348-017-2478-8
[8] Leontiev A I, Kiselev N A, Vinogradov Y A, Strongin M M, Zditovets A G and Burtsev S A 2017 Experimental investigation of heat transfer and drag on surfaces coated with dimples of different shape Int. J. Therm. Sci.118152–67
[9] Mohd Aizad Sazrul Sabrudin, Mohd Farid Muhamad Said and Zulkarnain Abdul Latiff. Effects of Asymmetric Intake Valve Lift Configuration Towards In-Cylinder Air Flow Behaviour. ARPN Journal of Engineering and Applied Sciences; VOL. 12, NO. 7, APRIL 2017
[10] SuperFlow Technologies Group, 2004. SF-1020 Flowbench Operators' Manual, SuperFlow Corporation, Colorado Springs, USA.
[11] Abdul Rahim Ismail, Rosli Abu Bakar and Semin, An Investigation of Valve Lift Effect on Air Flow and Coefficient of Discharge of Four Stroke Engines Based on Experiment,American Journal of Applied Sciences 5 (8): 963-971, 2008, ISSN 1546-9239.

Journal of Materials Science: Materials in Electronics
https://doi.org/10.1007/s10854-018-0217-4

UV radiation emitting Gd^{3+} activated Sr_2SiO_4 host system prepared by sol–gel procedure: structural, electron paramagnetic resonance, and luminescence studies

Vijay Singh[1] · N. Singh[1] · M. S. Pathak[1] · Sumandeep Kaur[2] · M. Jayasimhadri[2] · S. Watanabe[3] · T. K. Gundu Rao[3]

Received: 6 September 2018 / Accepted: 11 October 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract
Trivalent gadolinium (Gd^{3+}) activated strontium silicate (Sr_2SiO_4:xGd^{3+}; $0.01 \le x \le 0.11$) systems have been synthesized via sol–gel technique. The structural, morphological, luminescent, and magnetic properties have been investigated using X-ray diffraction patterns, scanning electron microscopy (SEM), photoluminescence (PL), and electron paramagnetic resonance (EPR) technique, respectively. The synthesized strontium silicate samples exhibit a mixed phase of α′ and β-Sr_2SiO_4. The SEM images show irregular morphologies with a random distribution of particles. The PL spectra confirm the ultraviolet emission from the synthesized samples and the optimum concentration of Gd^{3+} ions is observed at 0.05 mol. With low dopant concentrations, Sr_2SiO_4:Gd^{3+} sample displays EPR lines, mainly in the low field region with effective g-values 5.8, 4.5, 3.1, and 2.7. As dopant concentration increases, a broad line appears with $g_{eff} \sim 2.1$ along with the low field lines. A relatively strong crystal field is experienced by Gd^{3+} ion due to distortions in the immediate environment of the ion, which lead to low field EPR lines. These studies allow the use of gadolinium-activated strontium silicate in phototherapy lamps.

1 Introduction

Lanthanide activated inorganic materials have been extensively investigated due to their unique electronic, optical, and magnetic properties [1–3]. Among various inorganic hosts, silicates are known as excellent host materials for many technological applications due to their admirable chemical, physical, thermal, and high irradiation power stability [4–8]. Alkaline earth metal orthosilicates (M_2SiO_4; M = Ca, Sr, Ba) have attracted much attention in the recent years [9–14]. Alkaline earth silicates are resistant to many chemicals and air moisture [15]. Strontium orthosilicate (Sr_2SiO_4), in particular, has attracted attention due to its special structural features, excellent physical, and chemical stability [16]. The Eu^{2+} activated Sr_2SiO_4 has been used commercially in white LEDs in place of YAG:Ce [17, 18]. The luminescence properties of Sr_2SiO_4 activated with different rare earth (RE) ions such as Eu^{3+}, Sm^{3+}, Eu^{2+}, Ce^{3+}, Tb^{3+}, Pr^{3+}, and Dy^{3+} have been extensively investigated due to their applications in white LEDs [12, 15, 19–22]. However, a lack of information exists regarding the exploration of the luminescence properties of the Gd^{3+} activated Sr_2SiO_4 system.

Among the lanthanide activated inorganic materials, trivalent gadolinium (Gd^{3+}) based systems have gained significant importance because of their simple electronic energy level scheme that can be used as a spectroscopic probe for host-dopant energy transfer dynamics, site swapping, point group symmetry, intricate structure-photoluminescence correlation, etc. [23–25]. In addition, Gd^{3+} ion contains seven unpaired electrons which make it an important ion as a water relaxation agent in MRI applications. Currently, Gd^{3+} ions are attracting more attention as neutron capture therapy agents in targeted radiotherapy for cancer treatment [26, 27]. In addition, a Gd ion can act as an auxiliary ion to serve as a trap center in several hosts for energy storage in long afterglow materials, and ionizing radiation detectors [28].

Ultraviolet (UV) radiations have been segregated into three different types: long wavelength UV rays (UV-A: 380–320 nm), medium wavelength UV rays

✉ Vijay Singh
vijayjiin2006@yahoo.com

[1] Department of Chemical Engineering, Konkuk University, Seoul 05029, Republic of Korea

[2] Department of Applied Physics, Delhi Technological University, Delhi 110 042, India

[3] Institute of Physics, University of Sao Paulo, Sao Paulo, SP 05508-090, Brazil

Published online: 16 October 2018

(UV-B: 320–280 nm), and short wavelength UV rays (UV-C: 280–190 nm) [29]. It is well known that UV radiations are environmental hazards to the skin. However, there are beneficial effects of controlled UV radiations such as in the production of vitamin D, water purifiers, photochemical reaction, and in phototherapy [30, 31]. Phototherapy helps in controlling skin diseases, jaundice in newborn babies, and mental illness. Therefore, the phototherapy lamps that use UV-A and UV-B radiations are increasingly demanding in the medical field for the treatment of various skin disorders such as vitiligo vulgaris, psoriasis, periodontitis, etc [32, 33]. The applications of UV-B radiations in controlling skin diseases have prompted researchers to explore optically active materials which emit in the narrow spectral region from 310 to 313 nm [34, 35].

As such, there have been no reports pertaining to the photophysical properties of Sr_2SiO_4 doped with varying concentrations of Gd^{3+} ion. It has been noted that the paramagnetic Gd^{3+} ion is very sensitive to electron paramagnetic resonance (EPR) spectroscopy. Gd^{3+} has been of great interest because of its ubiquitous nature (popularly known as the U-spectrum). The Gd^{3+} ions give the EPR signals even at room temperature and these signals are attributed to their own environment in crystalline and polycrystalline materials. Considering the importance of Gd^{3+} ions activated compound, in this report, we investigated the structural, morphological, luminescent, and magnetic properties of Gd^{3+} activated Sr_2SiO_4 system using X-ray diffraction (XRD), scanning electron microscope (SEM), photoluminescence (PL), and EPR spectroscopic techniques, respectively.

2 Experimental

A series of Sr_2SiO_4:xGd^{3+}; $0.01 \le x \le 0.11$ (SS1–SS6) samples were prepared by a sol–gel method using high-purity $Sr(NO_3)_2$, SiO_2, citric acid, and $Gd(NO_3)_3{\cdot}6H_2O$ as starting materials. The details of sample composition, sample code, and starting materials were given in Table 1. In the typical synthesis, certain stoichiometric amounts of strontium nitrate, fumed silica, gadolinium nitrate, and citric acid (citric acid/metal ion = 2:1, molar ratio) were dissolved in 10 ml de-ionized water in a 50 ml porcelain crucible. Here the citric acid serves as a complexing/chelating agent while stirring for about 1 h in order to obtain a complete homogeneous solution of all reactants. The obtained homogeneous solution was kept in an oven at 110 °C until the dried gel was formed. Afterward, the dried-gel was fired at 400 °C for 2 h. Finally, the resulting brown-color mass was crushed using a mortar and pestle and then annealed at 1100 °C for 3 h to obtain the final product.

The XRD patterns of the samples were recorded using RIGAKU Miniflex-II diffractometer. Cu-Kα radiation (λ = 1.5406 Å) was used as an X-ray source. Morphology details were obtained using an SEM (S-3400, Hitachi, Japan). Photoluminescence measurements were carried out using a Shimadzu RF-5301PC, spectrofluorophotometer equipped with Xenon flash lamp. The EPR spectra of the samples were recorded using a JEOL FE1X ESR Spectrometer, operating in the X-band frequencies, with a field modulation of 100 kHz.

3 Results and discussion

3.1 X-ray diffraction studies

The XRD patterns of the sol–gel synthesized Gd^{3+} activated Sr_2SiO_4 (SS1–SS6) compound were shown in Fig. 1. Since Sr_2SiO_4 exists in two different crystallographic phases as α′ and β-Sr_2SiO_4, the experimental data has been compared with the standard data of orthorhombic α′-Sr_2SiO_4 and monoclinic β-Sr_2SiO_4 having JCPDS file no. 39-1256 and 38-0271, respectively. As it can be seen from Fig. 1, the diffraction patterns of Sr_2SiO_4:Gd^{3+} (SS1–SS6) samples are quite similar to the orthorhombic α′-Sr_2SiO_4 structure with a space group Pnma. However, peaks of monoclinic β-Sr_2SiO_4 phase were observed in Sr_2SiO_4:Gd^{3+} samples. In the literature, it has been reported that both the phases can co-exist since the transition temperature is low and can easily undergo short range rearrangement without cleaving any bond between the coordinated atoms [36–39]. Based on the diffraction pattern, it was observed that there are presences of few

Table 1 Detailed information of sample code, and starting materials

Samples	Sample code	Starting materials			
Sr_2SiO_4:$Gd_{0.01}$	SS1	Sr = 1.6928 g	Si = 0.2400 g	C.A = 3.0736 g	Gd = 0.0180 g
Sr_2SiO_4:$Gd_{0.03}$	SS2	Sr = 1.6928 g	Si = 0.2400 g	C.A = 3.0736 g	Gd = 0.0541 g
Sr_2SiO_4:$Gd_{0.05}$	SS3	Sr = 1.6928 g	Si = 0.2400 g	C.A = 3.0736 g	Gd = 0.0902 g
Sr_2SiO_4:$Gd_{0.07}$	SS4	Sr = 1.6928 g	Si = 0.2400 g	C.A = 3.0736 g	Gd = 0.1263 g
Sr_2SiO_4:$Gd_{0.09}$	SS5	Sr = 1.6928 g	Si = 0.2400 g	C.A = 3.0736 g	Gd = 0.1624 g
Sr_2SiO_4:$Gd_{0.11}$	SS6	Sr = 1.6928 g	Si = 0.2400 g	C.A = 3.0736 g	Gd = 0.1985 g

Sr = $Sr(NO_3)_2$, Si = SiO_2, C.A = citric acid, Gd = $Gd(NO_3)_3{\cdot}6H_2O$

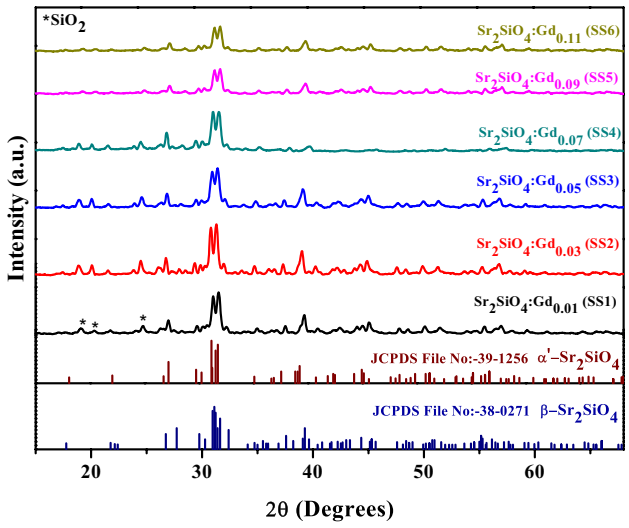


Fig. 1 X-ray diffraction patterns of Sr_2SiO_4:Gd^{3+} (SS1–SS6) compound

weak unreacted SiO_2 peaks which were marked as asterisk (*). It is known that α′-Sr_2SiO_4 and β-Sr_2SiO_4 have similar crystal structures (shown in Fig. 2). In Sr_2SiO_4 structure, each Sr^{2+} ion occupies two kinds of inequivalent symmetry sites, S1 and S2, which occur in the lattice in the same amount. The coordination numbers of the Sr^{2+} sites in both the phases are nearly same. Sr^{2+} ion is coordinated with ten oxygens in the S1 site, whereas in S2 site Sr^{2+} ion is coordinated with nine oxygens [40, 41]. The ionic radius of Gd^{3+} is very similar to Sr^{2+} and larger than Si^{4+}, therefore, dopant ions (Gd^{3+}) can substitute Sr^{2+} in two different sites (S1 and S2).

3.2 SEM studies

Figure 3 shows the SEM images of Sr_2SiO_4:Gd^{3+} (SS3) compound at different resolutions. The micrographs show irregular morphology and highly dense particles (shown in Fig. 3a, b). Furthermore, the particle size is in a wide distribution due to agglomerated nature of the particles. It was hard to estimate the particles sizes from the SEM images because of the agglomeration of particles. Figure 3c shows that the particles contain some pores and these might be formed due to the liberation of gas molecules during the sintering process. The high magnified image shows that agglomerates are composed of some small size spherical particles in the adhesion of bigger particles (shown in Fig. 3d).

3.3 Photoluminescence studies

Figure 4a shows the photoluminescence excitation (PLE) spectra of the Sr_2SiO_4:Gd^{3+} (SS1–SS6) systems monitored keeping the emission wavelength at 313 nm. The PLE spectra exhibit excitation bands correspond to the f-f intra-configurational transitions at 243, 245, 251, 273, 275, and 279 nm are attributed to transitions from $^8S_{7/2}$ to $^6D_{5/2,\ 3/2}$,$^6D_{7/2}$, $^6D_{9/2}$, $^6I_{15/2,\ 13/2}$, $^6I_{9/2,\ 17/2}$, and $^6I_{7/2}$, respectively [42, 43]. The transition corresponding to $^8S_{7/2} \rightarrow {}^6I_{15/2,\ 13/2}$ is an intense narrow band as compared to the other transitions. The broad excitation band with a maximum at around 245 nm might be due to the combination of the host lattice absorption and weak ^6D_J transition of Gd^{3+} ions. Qi et al. [21] studied green long-after-glow luminescence of Tb^{3+} in Sr_2SiO_4. They observed two overlapping broad bands with maximum at 285 nm and 228 nm due to 4f–4f5d transition and the host

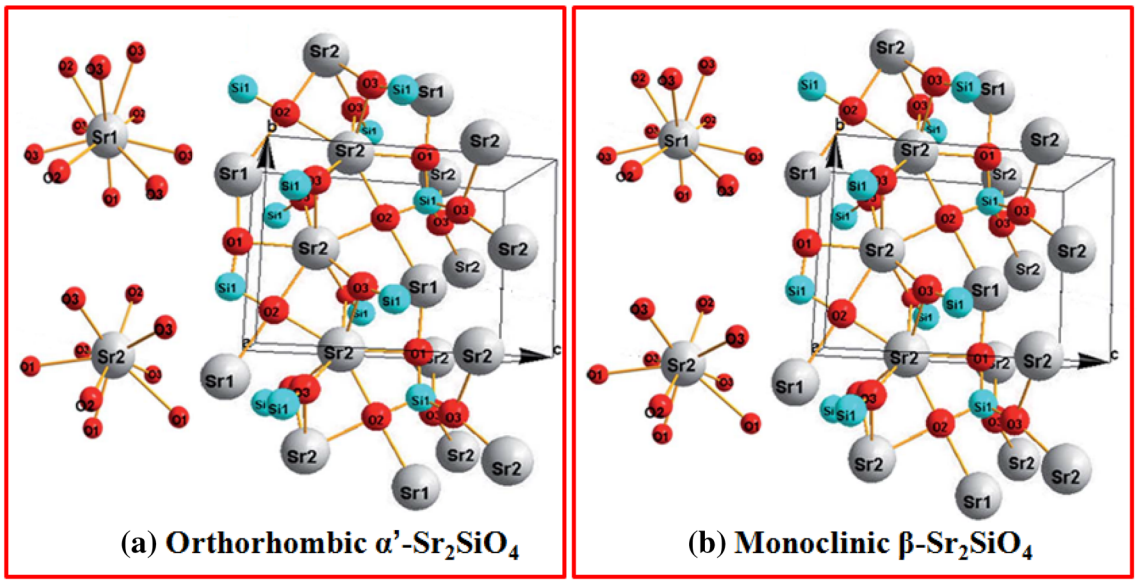


Fig. 2 Schematic diagram of **a** orthorhombic, and **b** monoclinic Sr_2SiO_4 unit cells

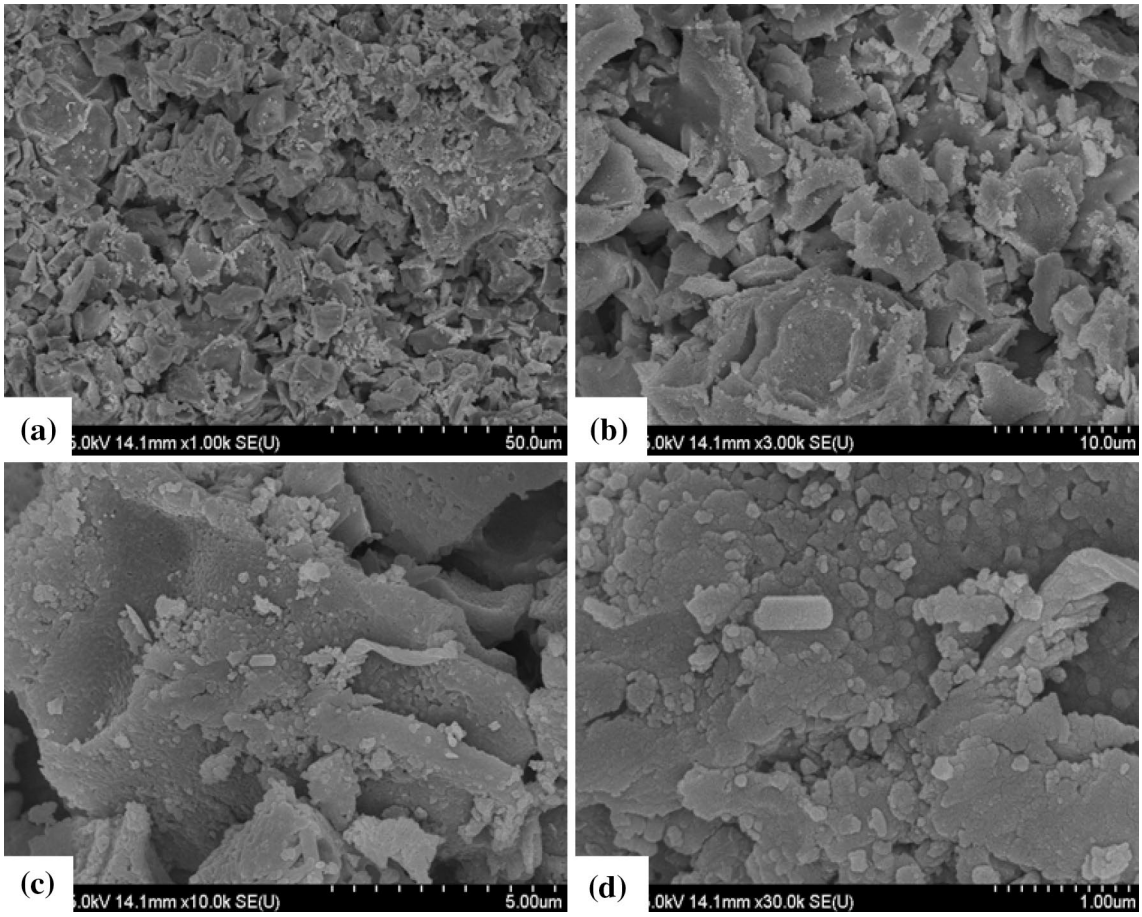


Fig. 3 SEM images of Sr_2SiO_4:Gd^{3+} (SS3) compound

absorption, respectively. As the observed excitation bands match with the commercial UV-based LEDs as well as CFL emission, the system is quite suitable to use in solid state lighting applications [38]. Among all excitation bands, the most intense excitation band has been observed at 273 nm, indicating maximum absorption of incident radiation at this particular wavelength. Therefore, emission spectra were recorded under 273 nm excitation. The emission spectra of the synthesized systems exhibit two emission bands at 307 and 313 nm attributed to $^6P_{5/2} \rightarrow {}^8S_{7/2}$ and $^6P_{7/2} \rightarrow {}^8S_{7/2}$ transitions, respectively as shown in Fig. 4b [30, 44, 45].

Figure 5 shows the schematic energy level diagram for Gd^{3+} ion in Sr_2SiO_4 systems. From this figure, it is observed that the ^6P_J and ^6I_J energy levels of Gd^{3+} ions are closely packed. Firstly Gd^{3+} can be excited from the ^8S state to ^6D and ^6I states. Afterward, the ions decay to the level ^6P_J state through non-radiative energy transfer process and finally relax to the ^8S state, which gives the NB-UVB emission at 313 nm. Figure 6 shows emission with the variation of Gd^{3+} doping concentration. Emission of Gd^{3+} ions increased with the increase of Gd^{3+} concentration up to 0.05 mol, then decreased due to concentration quenching. When the concentration of Gd^{3+} ions increases, the distance between the Gd^{3+} ions decreases. This leads to non-radiative energy transfers between Gd^{3+} ions. The non-radiative energy transfer of the luminescence of phosphor system is based on the resonance transfer by exchange interaction or electric multipole–multipole interaction [46, 47]. The critical distance for energy transfer (R_c) in Sr_2SiO_4:Gd^{3+} can be calculated by Blasse formula [48]:

$$R_c \approx 2\left[\frac{3V}{4\pi\chi_c N}\right]^{1/3}$$

where V is the volume of unit cell, χ_c is the critical concentration of dopant and N is the number of available sites for the dopant in the unit cell. For the Sr_2SiO_4:Gd^{3+} system, $\chi_c = 0.05$, $V = 391.20$Å^3 [49], and $N = 8$. The R_c of Gd^{3+} in Sr_2SiO_4:Gd^{3+} system was found to be 12.31 Å. The critical distance is higher than 5 Å for exchange interaction. Therefore, multipolar interaction is dominant for the energy transfer between the nearest–neighbour ions. The PL emission studies show that the band at 313 nm which is a prerequisite for curing skin diseases in phototherapy [32, 33].

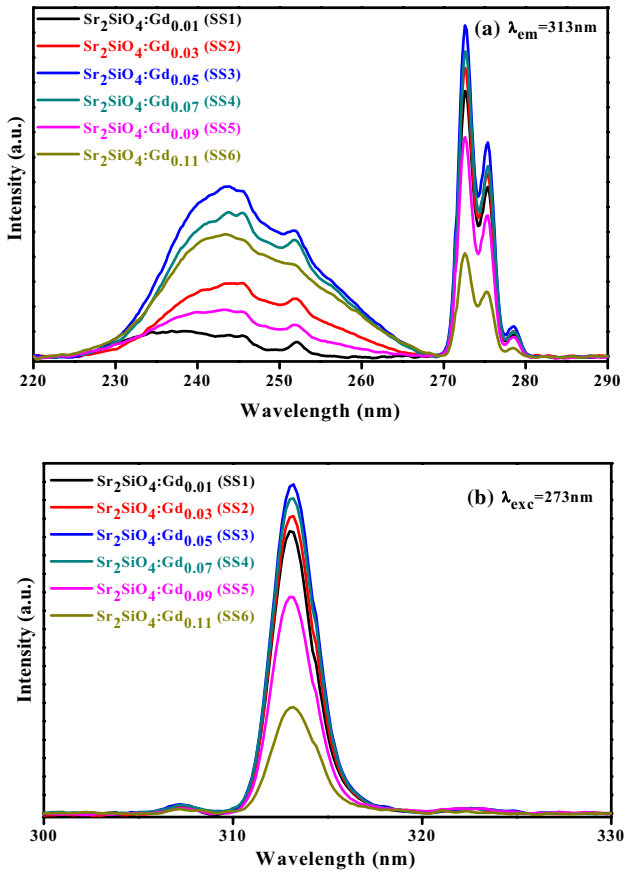


Fig. 4 **a** Photoluminescence excitation (PLE) spectra of $Sr_2SiO_4:Gd^{3+}$ (SS1–SS6) compound (λ_{em} = 313 nm) and **b** emission spectra of $Sr_2SiO_4:Gd^{3+}$ (SS1–SS6) (λ_{exc} = 273 nm)

3.4 Electron paramagnetic resonance studies

The room temperature EPR spectrum of Gd^{3+} activated Sr_2SiO_4 is shown in Fig. 7a. Relatively high-intensity lines are observed in the low field region with effective *g*-values at ~ 5.8, 4.5, 3.1 and 2.7. Low-intensity lines are observed at g_{eff} ~ 2.4 and 2.1. S-state ions like Eu^{2+} and Gd^{3+} generally exhibit a simple spectrum in powder samples with an EPR line around g ~ 2.0. This spectrum arises when the ion experiences a weak crystal field and the weak crystal field results from an ion located at undistorted sites. Strong crystal field will be experienced by Gd^{3+} ion in case the ion is located in distorted environments. EPR lines will now appear in the low field region of the spectrum. In certain cases, Gd^{3+} ion exhibits a unique spectrum with a characteristic line shape. This spectrum is termed as U-spectrum. U-spectrum arises when the crystal field varies among the different sites of the ion. An example of such a system could be a glass system or zeolites. In a U-spectrum, apart from its unique shape, EPR lines will be seen at an effective *g*-value of 2.0, 2.8 and 5.9. In the present system, low field lines are seen and also a

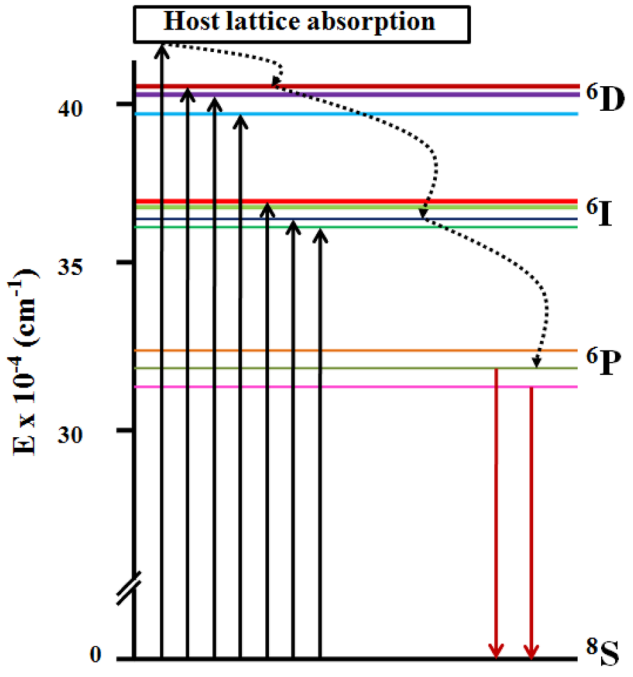


Fig. 5 Schematic energy level diagram of Gd^{3+} ion in Sr_2SiO_4 compound

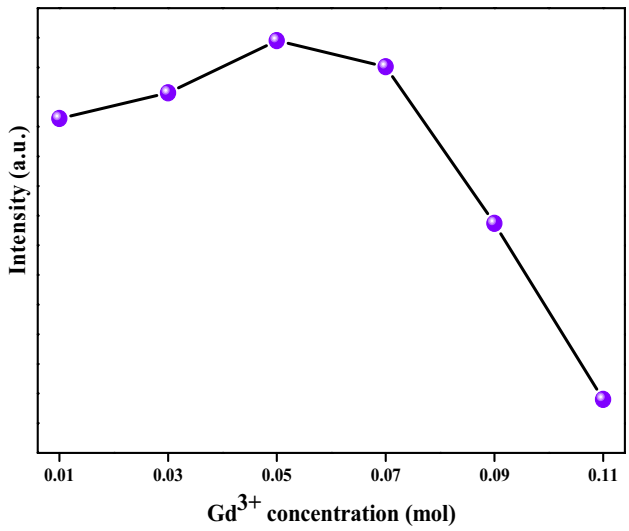


Fig. 6 The intensity variation of $^6P_{7/2} \rightarrow {}^8S_{7/2}$ (313 nm) transition with respect to Gd^{3+} ion concentration

broad line (linewidth ~ 2560 G) appears as the concentration of Gd^{3+} ion increases in the system with an effective *g*-value of 2.4. The EPR spectrum of the system with increasing activator concentration is shown in Fig. 7b.

Gd^{3+} ion is an S-state ion with a half-filled 4f shell with an electronic configuration $[Xe]4f^7$. The ion is a pure S-state with L = 0 and is characterized by a completely spherical charge cloud. Crystal field resulting from the neighboring

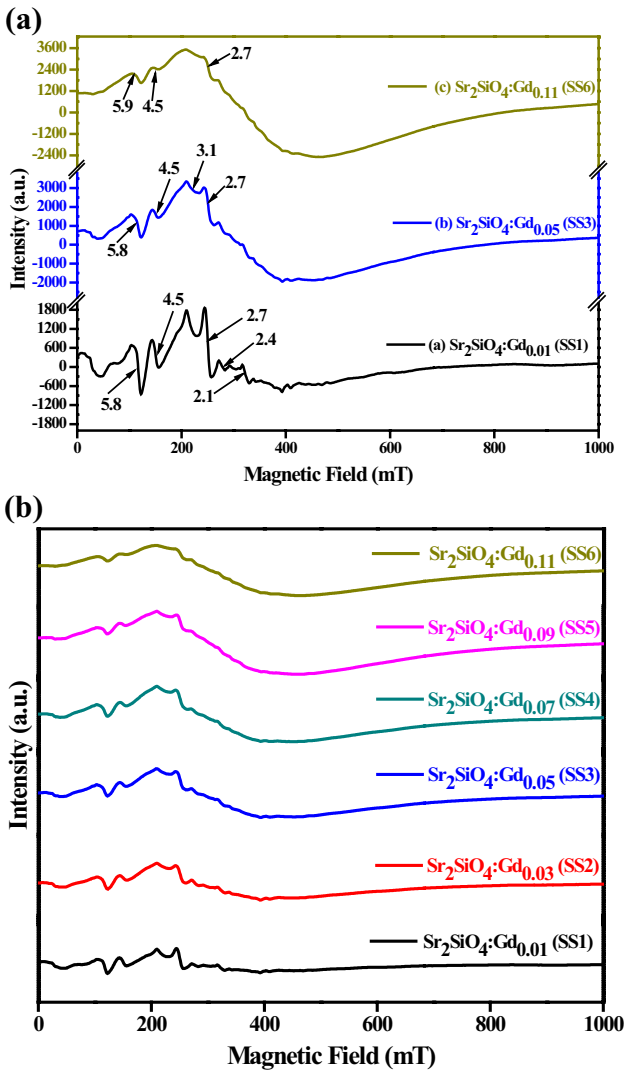


Fig. 7 **a** Effective *g* values marked for the samples (SS1, SS3, SS6) and **b** EPR spectra of Sr_2SiO_4:Gd^{3+} (SS1–SS6) compounds

ions has no effect on the ion due to spherical charge distribution. However, crystal field effect does arise resulting in splitting due to higher order perturbations involving spin–orbit coupling. There will be an admixture of higher state like $^6P_{7/2}$ and $^6D_{7/2}$ with $L \neq 0$ with the pure S-state. The ion interacts with the crystal field through the mediation of spin–orbit coupling. This admixture results in ground-state splitting and results in crystal field effects in the EPR spectrum. As the environment around Gd^{3+} changes, there will be changes in the magnitude of crystal field seen by the ion and the observed EPR spectrum changes.

Brodbeck and Iton [50] have discussed the different ESR spectrum of Gd^{3+} ion that may be observed in powder samples. They considered the spectra based on the magnitude of crystal field interaction (H_{CF}) relative to the Zeeman splitting hν (ν is the microwave frequency). Three regions were identified to classify the EPR spectra viz., weak, intermediate and strong. These regions depend on the ratio $H_{CF}/h\nu$. Weak crystal field region is defined by $H_{CF}/h\nu \leq 1/4$. Here the spectrum is observed mainly in the region of $g \sim 2.0$. Intermediate region is considered in two parts viz., (1) lower intermediate and (2) higher intermediate. In the lower intermediate region, $1/4 \leq H_{CF}/h\nu \leq 1$ and the EPR lines are expected over a wide range of g-values with values varying in the region $2.0 < g < \infty$. Higher intermediate crystal field region corresponds to the case where lines are seen with $g > 2.0$. For strong crystal field case $H_{CF}/h\nu \geq 4$. Here, EPR lines arise from transitions controlled by resonances within the Kramer's levels.

Sr_2SiO_4 exists in two structurally different forms. The β-Sr_2SiO_4 crystallizes in monoclinic symmetry with space group $P2_1/n$. On the other hand, α′-Sr_2SiO_4, has orthorhombic symmetry with space group Pnma [51]. There are no major differences in the crystal structure of the two forms of Sr_2SiO_4. There exist two non-equivalent sites for the Sr^{2+} ion in the lattice. In the case of site S1, the Sr^{2+} ion is coordinated by ten oxygens while nine oxygens coordinate the ion for the S2 site. The two sites exist in equal amounts in the lattice. The difference in the two phases is a small rotation of SiO_4 tetrahedra in α′-Sr_2SiO_4.

Divalent cation (Sr^{2+}) and tetravalent cation (Si^{4+}) are present in Sr_2SiO_4. Their presence leads to mixed occupancy wherein Sr^{2+} ions are partially replaced by Si^{4+} ions. The existence of mixed occupancy comes from antisite cation disorder. This disorder is a point defect in crystals. Theoretical calculations predict the disorder to be present in a crystal lattice [52]. Their presence has been confirmed by experimental investigations [53, 54]. It has also been possible to directly observe the defects [55]. Due to antisite disorder, cation vacancies may form in Sr_2SiO_4. It has been shown that vacancies will form with ease in crystals with cation disorder [56].

Several studies have shown that α′-β phase transformation of Sr_2SiO_4 occurs through a short-range rearrangement without breaking any bonds [36–39]. Sr^{2+} ion has nine- and tenfold coordination in Sr_2SiO_4 and its ionic radius is 1.31 Å and 1.36 Å respectively in this coordination [57]. Si^{4+} has fourfold coordination and its ionic radius is 0.26 Å in a fourfold coordination. Gd^{3+} ionic radius is 1.11 Å in a ninefold coordination. As Gd^{3+} ionic radius is much larger than Si^{4+}, the activator Gd^{3+} ion substitutes Sr^{2+} ion in Sr_2SiO_4. As there are two Sr^{2+} sites viz., S1 and S2, it is very likely the activator ion is present at both the sites of Sr^{2+}.

Jiansheng and Yanli [58] have studied the possible changes in the thermophysical properties of $LaMgAl_{11}O_{18}$ (LMA) doped with Sm^{2+} and Gd^{3+} ions. Changes were expected in the LMA lattice, specifically in the environment of La^{3+}, as the effective ionic radius of Gd^{3+} ion is smaller than La^{3+}. The decrease of cell parameters has been

observed with increasing Gd^{3+} concentration and also there was a contraction of the unit volume. They came to the conclusion that this effect was the result of the substitution of a La^{3+} ion by the smaller Gd^{3+} ion. Jiansheng and Yanli's [58] Raman investigation of the same system showed that there was a shortening of chemical bonds as a result of the strengthening of bonds with increasing Gd^{3+} concentration. Based on these results, Jiansheng and Yanli [58] surmised that the LaO_2 network was distorted due to the presence of a Gd^{3+} ion in the lattice which led to local distortions. In the present system also the activator Gd^{3+} ion has a smaller ionic radius compared to Sr^{2+} ion. It is, therefore, speculated that there will be local distortions in the proximity of Gd^{3+} ion which substitutes Sr^{2+} ion. Local distortions give rise to relatively stronger crystal field experienced by Gd^{3+} ion. According to Brodbeck and Iton [50], this would lead to ESR lines in the low field region of the spectrum. The present EPR results are in accordance with this expectation. It was mentioned earlier that vacancies will be present in the Sr_2SiO_4 lattice. A vacancy located near a Gd^{3+} ion may alter the symmetry at the ion and result in distortions. This also can lead to a stronger crystal field experienced by Gd^{3+} ion.

Crystal structure studies have shown that there are two distinct sites for the Sr^{2+} ion in Sr_2SiO_4. The observed distinct EPR lines seen with low activator concentration could arise from Gd^{3+} ion located at one of these sites. Figure 7b shows the EPR spectra observed in Sr_2SiO_4 with increasing concentration of the Gd^{3+} ion. A broad line with a linewidth of 2378 G and a *g*-value equal to 2.1 gradually increases in intensity as Gd^{3+} concentration increases. It is speculated that this broad line arises from Gd^{3+} ion located in the second Sr^{2+} site in Sr_2SiO_4. The broadness of the EPR line suggests that Gd^{3+} ion at these sites are likely to be quite close to each other resulting in dipolar interaction between the ions.

4 Conclusions

Gd^{3+} activated strontium silicate (Sr_2SiO_4:xGd^{3+}; $0.01 \leq x \leq 0.11$) systems with increasing Gd^{3+} concentration have been synthesized by sol–gel technique. The obtained system shows the mixed phase of the Sr_2SiO_4 structure with a combination of monoclinic and orthorhombic structure. The SEM images show the irregular morphology of particles with some agglomeration. Under the excitation of 273 nm, the PL spectrum shows dominant emission bands from $^6P_{7/2} \rightarrow {}^8S_{7/2}$ (313 nm) transitions of Gd^{3+} ions. The optimum concentration of Gd^{3+} ions is found to be 0.05 mol. Photoluminescence results suggest that the Gd^{3+} activated strontium silicate system could be useful for phototherapy lamps. The observed EPR spectrum in the Gd^{3+} activated Sr_2SiO_4 system is due to the Gd^{3+} ions located at two Sr^{2+} sites in the lattice. The relatively smaller effective ionic radius of Gd^{3+} ion induces distortions and leads to the appearance of EPR lines in the low field region of the spectrum.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT) (2018M2B2A9065656).

References

1. M. Zahedi, S.A. Hassanzadeh-Tabrizi, A. Saffar-Teluri, Sol–gel synthesis and luminescence properties of Ba_2SiO_4:Sm^{3+} nanostructured phosphors. Ceram. Int. **44**, 10169–10174 (2018)
2. P.S. Mbule, O.M. Ntwaeaborwa, B.M. Mothudi, M.S. Dhlamini, Structural and optical characterization of nanoparticulate manganese doped zinc silicate phosphors prepared by sol–gel and combustion methods. J. Lumin. **179**, 74–82 (2016)
3. N.A. Zaitseva, T.A. Onufrieva, J.A. Barykina, T.I. Krasnenko, E.V. Zabolotskaya, R.F. Samigullina, Magnetic properties and oxidation states of manganese ions in doped phosphor Zn_2SiO_4:Mn. Mater. Chem. Phys. **209**, 107–111 (2018)
4. F. Wang, W. Wang, Y. Jin, Photoluminescence properties of Ce^{3+}/Mn^{2+} doped calcium zirconium silicate phosphors with energy transfer for white LEDs. Ceram. Int. **42**, 16626–16632 (2016)
5. I. Mikalauskaite, E. Raudonyte-Svirbutaviciene, A. Linkeviciute, M. Urbonas, A. Katelnikovas, Luminescence and luminescence quenching of $Sr_3Lu_2(Si_3O_9)_2$:Ce^{3+} phosphors. J. Lumin. **184**, 185–190 (2017)
6. Y. Zhu, Y. Liang, S. Liu, H. Li, J. Chen, W. Lei, Design of hierarchical composite silicate for full-color and high thermal stability phosphors. Chem. Eng. J. **345**, 327–336 (2018)
7. Y. Guo, S.H. Park, B.C. Choi, J.H. Jeong, J.H. Kim, Effect of La^{3+} ion doping on the performance of Eu^{2+} ions in novel $Sr_3CeNa(PO_4)_2SiO_4$ phosphors. J. Alloys Compd. **724**, 763–773 (2017)
8. J. Xu, L. Cao, Y. Feng, Y. Huang, Y. Wang, L. Qin, H.J. Seo, Improved scintillation luminescence and thermal stability of $In_2Si_2O_7$ ceramic phosphor. J. Lumin. **183**, 166–172 (2017)
9. S.K. Raut, N.S. Dhoble, S.J. Dhoble, Optical properties of Eu, Dy, Mn activated M_2SiO_4, (M2=Ca, Sr, Zn) orthosilicate phosphors. J. Lumin. **134**, 325–332 (2013)
10. Q. Li, S.E. Brown, L.J. Broadbelt, J.-G. Zheng, N.Q. Wum, Synthesis and characterization of MCM-41-supported Ba_2SiO_4 base catalyst. Microporus Mesoporus Mater. **59**, 105–111 (2003)
11. J. Wen, Y.-Y. Yeung, L. Ning, C.-K. Duan, Y. Huang, J. Zhang, M. Yin, Effects of vacancies on valence stabilities of europium ions in β-Ca_2SiO_4:Eu phosphors. J. Lumin. **178**, 121–127 (2016)
12. K. Szczodrowski, J. Barzowska, N. Górecka, M. Grinberg, The role of compensation defects in Eu^{3+} stabilization under reductive atmosphere in Sr_2SiO_4 matrix. J. Alloys Compd. **748**, 44–50 (2018)
13. A. Vyas, C.P. Joshi, P.D. Sahare, S.V. Moharil, NIR emission in Ba_2SiO_4:Eu^{2+}, Nd^{3+} phosphors with near UV/violet excitation. J. Alloys Compd. **743**, 789–794 (2018)
14. J.M. Porras-Vázquez, L. dos Santos-Gómez, P.R. Slater, D. Marrero-López, E.R. Losilla, Crystallochemistry and electrical properties of Al-doped Sr_2SiO_4 electrolytes. Ceram. Int. **42**, 16317–16324 (2016)
15. S.K. Gupta, N. Pathak, S.K. Thulasidas, V. Natarajan, Local site symmetry of Sm^{3+} in sol-gel derived α′-Sr_2SiO_4: probed by emission and fluorescence lifetime spectroscopy. J. Lumin. **169**, 669–673 (2016)

16. C. Zhao, D. Zhu, W. Gao, T. Han, L. Peng, M. Tu, Effect of doping Gd^{3+} on crystal structure and luminescent properties of Sr_2SiO_4:Eu^{2+} phosphor. J. Rare Earths **33**, 693–699 (2015)
17. J.K. Park, M.A. Lim, C.H. Kim, H.D. Park, J.T. Park, S.Y. Choi, White light-emitting diodes of GaN-based Sr_2SiO_4:Eu and the luminescent properties. Appl. Phys. Lett. **82**, 683–685 (2003)
18. Y. Li, Q. Yu, L. Huang, J. Wang, Q. Su, Near ultraviolet and visible-to-near-infrared spectral converting properties and energy transfer mechanism of Sr_2SiO_4:Ce^{3+}, Pr^{3+} phosphor. Opt. Mater. Exp. **4**, 227–233 (2014)
19. C. Hsu, R. Jagannathan, C. Lu, Luminescent enhancement with tunable emission in Sr_2SiO_4:Eu^{2+} phosphors for white LEDs. Mater. Sci. Eng. B **167**, 137–141 (2010)
20. M.A. Tshabalala, F.B. Dejene, S.S. Pitale, H.C. Swart, O.M. Ntwaeaborwa, Generation of white-light from Dy^{3+} doped Sr_2SiO_4 phosphor. Phys. B **439**, 126–129 (2014)
21. W. Qi, Q. Jian-Bei, S. Zhi-Guo, Z. Da-Cheng, X. Xu-Hui, Green long-after-glow luminescence of Tb^{3+} in Sr_2SiO_4. Chin. Phys. B **23**, 064211–064215 (2014)
22. M. Pardha Saradhi, N. Lakshminarasimhan, S. Boudin, K. Vijay Kumar Gupta, U.V. Varadaraju, B. Raveau, Enhanced luminescence of Sr_2SiO_4:Dy^{3+} by sensitization (Ce^{3+}/Eu^{2+}) and fabrication of white light-emitting-diodes. Mater. Lett. **117**, 302–304 (2014)
23. A.O. Chauhan, N.S. Bajaj, S.K. Omanwar, Synthesis and photoluminescence study of narrow-band UVB-emitting $LiSr_4(BO_3)_3$:Gd^{3+}, Pr^{3+} phosphor. Bull. Mater. Sci. **40**, 1–6 (2017)
24. A.O. Chauhan, A.B. Gawande, S.K. Omanwar, Narrow band UVB emitting phosphor $LaPO_4$:Gd^{3+} for phototherapy lamp. Optik **127**, 6647–6652 (2016)
25. P.P. Mokoena, M. Gohain, B.C.B. Bezuidenhoudt, H.C. Swart, O.M. Ntwaeaborwa, Luminescent properties and particle morphology of $Ca_3(PO_4)_2$:Gd^{3+}, Pr^{3+} phosphor powder prepared by microwave assisted synthesis. J. Lumin. **155**, 288–292 (2014)
26. T. Jiang, Y. Liu, S. Liu, N. Liu, W. Qin, Upconversion emission enhancement of Gd^{3+} ions induced by surface Plasmon field in Au@$NaYF_4$ nanostructures co-doped with Gd^{3+}-Yb^{3+}-Tm^{3+} ions. J. Colloid Interface Sci. **377**, 81–87 (2012)
27. N.J.J. Johnson, W. Oakden, G.J. Stanisz, R.S. Prosser, F.C.J.M. van Veggel, Size-tunable, ultra small $NaGdF_4$ nanoparticles: insights into their T_1 MRI contrast enhancement. Chem. Mater. **23**, 3714–3722 (2011)
28. F. Nakamura, T. Kato, G. Okada, N. Kawano, N. Kawaguchi, Scintillation, dosimeter and optical properties of MgF_2 transparent ceramics doped with Gd^{3+}. Mater. Res. Bull. **98**, 83–88 (2018)
29. World Health Organization, *Global Solar UV Index: A Practical Guide* (World Health Organization, Geneva, 2002)
30. B.V. Padlyak, A. Drzewiecki, T.B. Padlyak, V.T. Adamiv, I.M. Teslyuk, Resonant excited UV luminescence of the Gd^{3+} centres in borate glasses, co-doped with Gd and Ag. Opt. Mater. **79**, 302–309 (2018)
31. P.P. Mokoena, I.M. Nagpure, V. Kumar, R.E. Kroon, E.J. Olivier, J.H. Neethling, H.C. Swart, O.M. Ntwaeaborw, Enhanced UVB emission and analysis of chemical states of $Ca_5(PO_4)_3OH$:Gd^{3+}, Pr^{3+} phosphor prepared by co-precipitation. J. Phys. Chem. Solids **75**, 998–1003 (2014)
32. S. Dogra, A.J. Kanwar, Narrow band UVB phototherapy in dermatology. Indian J. Dermatol. Venereol. Leprol. **70**, 205–209 (2004)
33. W. Köster, A. Wiskemann, Phototherapy with UV-B in vitiligo. Z. Hautkr. **65**, 1022–1024 (1990)
34. V. Singh, G. Sivaramaiah, J.L. Rao, R.S. Kumaran, P.K. Singh, T.-S. Kim, L.K. Kim, Luminescence and EPR studies of ultraviolet light emitting $La_2Zr_2O_7$:Gd^{3+} phosphor powder. J. Mater. Sci. Mater. Electron. **26**, 5195–5201 (2015)
35. V. Singh, G. Sivaramaiah, J.L. Rao, S.H. Kim, Investigation of new UV-emitting, Gd-activated $Y_4Zr_3O_{12}$ phosphors prepared via combustion method. J. Lumin. **157**, 82–87 (2015)
36. J.H. Lee, Y.J. Kim, Photoluminescent properties of Sr_2SiO_4:Eu^{2+} phosphors prepared by solid-state reaction method. Mater. Sci. Eng. B Solid-State Mater. Adv. Technol. **146**, 99–102 (2008)
37. J. Barzowska, A. Chru, K. Przegi, K. Szczodrowski, Dosimetric features of strontium orthosilicate (Sr_2SiO_4) doped with Eu^{2+}. Radiat. Phys. Chem. **104**, 31–35 (2014)
38. J.K. Han, M.E. Hannah, A. Piquette, G.A. Hirata, J.B. Talbot, K.C. Mishra, J. Mckittrick, Structure dependent luminescence characterization of green-yellow emitting Sr_2SiO_4:Eu^{2+} phosphors for near UV LEDs. J. Lumin. **132**, 106–109 (2012)
39. J.H. Lee, Y.J. Kim, A correlation between a phase transition and luminescent properties of Sr_2SiO_4:Eu^{2+} prepared by a flux method, J. Ceram. Proc. Res. **10**, 81–84 (2009)
40. C. Zhao, D.-C. Zhu, Y. Pu, L.-L. Peng, M.-J. Tu, The effect of doping La^{3+} on structure and luminescent properties of Sr_2SiO_4:Eu^{2+} phosphors. Ceram. Int. **41**, 13341–13347 (2015)
41. L.-C. Ju, C. Cai, Q.-Q. Zhu, J.-Y. Tang, L.-Y. Hao, X. Xu, Color tunable Sr_2SiO_4:Eu^{2+} phosphors through the modification of crystal structure. J. Mater. Sci. Mater. Electron. **24**, 4516–4521 (2013)
42. W.T. Carnall, P.R. Fields, K. Rajnak, Spectral intensities of the trivalent lanthanides and actinides in solution. II. Pm^{3+}, Sm^{3+}, Eu^{3+}, Gd^{3+}, Tb^{3+}, Dy^{3+}, and Ho^{3+}. J. Chem. Phys. **49**, 4412–4423 (1968)
43. S. Tamboli, B. Rajeswari, S.J. Dhoble, Investigation of UV-emitting Gd^{3+}-doped $LiCaBO_3$ phosphor. Luminescence **31**, 551–556 (2016)
44. V. Singh, G. Sivaramaiah, N. Singh, J.L. Rao, P.K. Singh, M.S. Pathak, D.A. Hakeem, EPR and PL studies on UVB-emitting gadolinium-doped $SrAl_{12}O_{19}$ phosphors. Optik **158**, 1227–1233 (2018)
45. A. Kumar, D.K. Rai, S.B. Rai, Luminescence of Gd^{3+} ions doped in oxyfluoroborate glass. Solid State Commun. **117**, 387–392 (2001)
46. S. Kaur, A.K. Vishwakarma, N. Deopa, A. Prasad, M. Jayasimhadri, A.S. Rao, Spectroscopic studies of Dy^{3+} doped borate glasses for cool white light generation. Mater. Res. Bull. **104**, 77–82 (2018)
47. Z. Xia, R.-S. Liu, K.-W. Huang, V. Drozd, $Ca_2Al_3O_6F$:Eu^{2+}:a green-emitting oxyfluoride phosphor for white light-emitting diodes. J. Mater. Chem. **22**, 15183–15189 (2012)
48. G. Blasse, Energy transfer between inequivalent Eu^{2+} ions. Philips Res. Rep. **24**, 131–136 (1969)
49. C. Guo, Y. Xu, F. Lv, X. Ding, Luminescent properties of Sr_2SiO_4:Eu^{2+} nanorods for near-UV white LED. J. Alloys Compd. **497**, L21–L24 (2010)
50. C.M. Brodbeck, L.E. Iton, The EPR spectra of Gd^{3+} and Eu^{2+} in glassy systems. J. Chem. Phys. **83**, 4285–4299 (1985)
51. M. Catti, G. Gazzoni, G. Ivaldi, G. Zanini, The $\beta \leftrightarrow \alpha'$ phase transition of Sr_2SiO_4. I. Order–disorder in the structure of the α' form at 383 K, Acta. Crystallogr. B **39**, 674–679 (1983)
52. M.M. Kuklja, Defects in yttrium aluminium perovskite and garnet crystals: atomistic study. J. Phys. Condens. Matter **12**, 2953–2967 (2000)
53. A.P. Patel, M.R. Levy, R.W. Grimes, R.M. Gaume, R.S. Frigelson, K.J. McClellan, C.R. Stanek, Mechanisms of nonstoichiometry in $Y_3Al_5O_{12}$. Appl. Phys. Lett. **93**, 191902–191902 (2008)
54. J. Dong, K. Lu, Noncubic symmetry in garnet structures studied using extended X-ray-absorption fine-structure spectra. Phys. Rev. B **43**, 8808–8821 (1991)
55. D. Truong, M.K. Devaraju, T. Tomai, I. Honma, Direct observation of antisite defects in $LiCoPO_4$ cathode materials by annular dark- and bright-field electron microscopy. ACS Appl. Mater. Interfaces **5**, 9926–9932 (2013)

56. N. Yuan, X. Liu, F. Meng, D. Zhou, First-principles study of La_2CoMnO_6: a promising cathode material for intermediate-temperature solid oxide fuel cells due to intrinsic Co-Mn cation disorder. J. Meng. Ionics **21**, 1675–1681 (2015)
57. R.D. Shannon, Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides, Acta Crystallogr. A32 (1976) 751
58. L. Jiansheng, L. Yanli, Effects of Sm^{3+} and Gd^{3+} co-doping on thermo-physical properties of $LaMgAl_{11}O_{19}$. Trans. Ind. Ceram. Soc. **73**, 270–276 (2014)

Video-based Person Re-identification Using Spatial-Temporal Attention Networks

Shivansh Rao
Delhi Technological University
shivansh_bt2k15@dtu.ac.in

Tanzila Rahman
University of Manitoba
rahmant4@cs.umanitoba.ca

Mrigank Rochan
University of Manitoba
mrochan@cs.umanitoba.ca

Yang Wang
University of Manitoba
ywang@cs.umanitoba.ca

arXiv:1810.11261v1 [cs.CV] 26 Oct 2018

ABSTRACT

We consider the problem of video-based person re-identification. The goal is to identify a person from videos captured under different cameras. In this paper, we propose an efficient spatial-temporal attention based model for person re-identification from videos. Our method generates an attention score for each frame based on frame-level features. The attention scores of all frames in a video are used to produce a weighted feature vector for the input video. Unlike most existing deep learning methods that use global representation, our approach focuses on attention scores. Extensive experiments on two benchmark datasets demonstrate that our method achieves the state-of-the-art performance. Code is available at *https://github.com/rshivansh/Spatial-Temporal-attention*.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision**; **Object identification**;

KEYWORDS

Person Re-Identification, Attention Mechanism

1 INTRODUCTION

In this paper, our goal is to solve the problem of video-based person re-identification. Given a video containing a person, the goal is to identify the same person from other videos possibly captured under different cameras. Person re-identification is useful in a wide range of applications, e.g video surveillance, police investigation, etc. A common strategy for person re-identification is to formulate it as a metric learning problem. Given the query video and a candidate video, the goal is to develop algorithms to compute the distance between these two videos. If the distance is small, it means the two videos likely contain the same person. See Figure 1 for an illustration. Previous work in person re-identification falls into two broad categories: image-based re-identification and video- based re-identification. Earlier work [10, 17, 19, 20, 22, 26, 28, 31] in this area focuses on the former, where the inputs to these systems are pairs of images and the goal is to identify whether they are images of the same person. Recently, video-based person re-identification is receiving increasing attention [8, 12, 16, 21, 24, 32, 33]. Compared with static images, video-based person re-identification is a more natural setting for practical applications such as video surveillance.

Person re-identification (either image or video based) is a challenging problem since the images/videos are often captured under different camera views. This can cause large variations in illumination, body pose, viewpoint, etc. Compared with static images, the temporal information in videos can potentially provide additional information that can help disambiguate the identity of a person. Previous work in this area has explored ways of exploiting this temporal information. A common strategy [16, 24, 32] is to use temporal pooling to combine frame-level features to represent the entire video sequence. Then this video-level feature vector can be used for re-identification.

Previous work [24, 32] has made the observation that not all frames in a video are informative. For example, if the person is occluded in a frame, ideally we would like the feature representation of the video to ignore this frame and focus on other useful frames. A natural way of solving this problem is to use the attention models [1, 18, 23] that have been popular in visual recognition recently. In [24, 32] , RNN is used to model the temporal information of the frames and generate the attention score for each frame for person re-identification.

In this paper, we propose a new attention model for video-based re-identification. Compared with previous works [24, 32], our model has several novelties. First, instead of using RNN, we directly produce the attention score of each frame based on the image feature of this frame. Our experimental results show that this simpler method outperforms RNN-based attention method. Since the attention score of each frame is calculated based on the frame, the computation of attention scores over all frames can be easily made parallel and take full advantage of the GPU hardware. Second we propose a new spatio-temporal attention model that allows useful information to be extracted from each frame without succumbing to occlusions and misalignments.

Our contributions include:

1. A new attention mechanism for video-based person re- identification. Unlike previous work [32] that uses RNN to generate the attentions, our model directly generates attentions based on frame-based features. As a consequence, the computation of the attentions is much simpler and can be easily parallelized. In contrast, RNN has to process frames in a sequential order, so the computation cannot be made parallel. Despite of its simplicity, our model outperforms the more sophisticated RNN-based attention mechanism in [32] .

2. Our newly designed network learns multiple spatial attention along with temporal attention. In order to represent the overall spatial attention of the frame, we need multiple hops of attention [11] that focus on different parts of the frame. In addition, we also study

Unpublished working draft. Not for distribution.

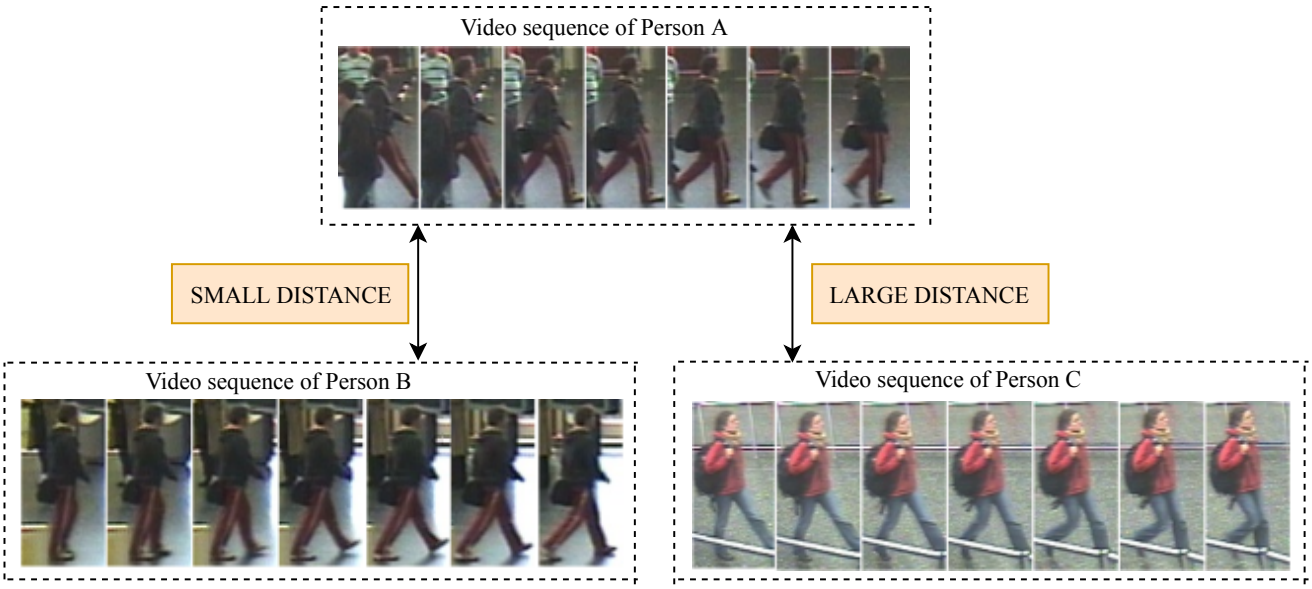


Figure 1: Illustration of the video-based person re-identification problem. In this case, our goal is to identify person A from two video sequences in the second row. If two videos contain the same person, we would like the distance between them to be small. Otherwise, we would like the distance to be large. Some frames in a video sequence may be affected by occlusions and are not informative about the person's identity. In this paper, we use an attention model to focus on informative frames for re-identification.

the effect of incrementing the value of number of hops of spatial attention.

2 RELATED WORK

There has been extensive work on person re- identification from static images. Early work in this area uses hand-crafted feature representations [3, 9, 14, 15, 30]. Most of these methods involve extracting feature representations that are invariant to viewpoint changes, then learning a distance metric to measure the similarity of two images.

Deep learning approaches, in particularly deep convolutional neural networks (CNNs), have achieved tremendous successes in various visual recognition tasks [7]. In many areas of computer vision, CNNs have replaced hand-engineering feature representations with features learned end-to-end from data. Recently, CNNs have been used for image-based person re-identification [10, 17, 19, 20, 22, 26, 28, 31]. These methods use deep network architecture such as Siamese network [4] to map images to feature vectors. These feature vectors can then be used for re- identification. Although the performance of image-based person re-identification has increased significantly, this is not a very realistic setting for practical applications. To address the limitation of image-based re-identification, a lot of recent work has began to explore video-based re-identification [8, 12, 16, 21, 24, 25, 32, 33] since it is closer to real-world application settings. Compared with static images, videos contain temporal information that is potentially distinctive for differentiating a person's identity. Some prior work has explored ways of incorporating temporal information in deep convolutional neural network for re-identification. For example, McLaughlin et al . [16] use CNN on each frame in a video and incorporate a recurrent layer on the CNN features. Temporal pooling is then used to combine frame-level features into a single video-level feature vector for re-identification.

Our work is also related to a line of research on incorporating attention mechanism in deep neural networks. The attention mechanism allows the neural networks to focus on part of the input and ignore the irrelevant information. It has been successfully used in many applications, including machine translation [1], image captioning [23], visual question answering [18], etc. In video-based re-identification, the attention mechanism has also been explored [24, 32] . The intuition is that only a small portion of the video contains informative information for re-identification. So the attention mechanism can be used to help the model focus on the informative part of the video.

The work in [32] is the closest to ours. It uses an RNN to generate temporal attentions over frames, so that the model can focus on the most discriminative frames in a video for re-identification. In this paper, we uses temporal attentions over frames as well. But instead of using RNN-based models to generate attentions [32], we directly calculate the attention scores based on frame-based features. This makes the model much simpler and the computation of attention scores can be easily parallelized over frames. We also propose propose a new spatio-temporal attention model wherein we calculate multiple spatial attention over the frames . We demonstrate that this multiple spatial attention mechanism improves the performance of the final model.

Unpublished working. Not for distribution.

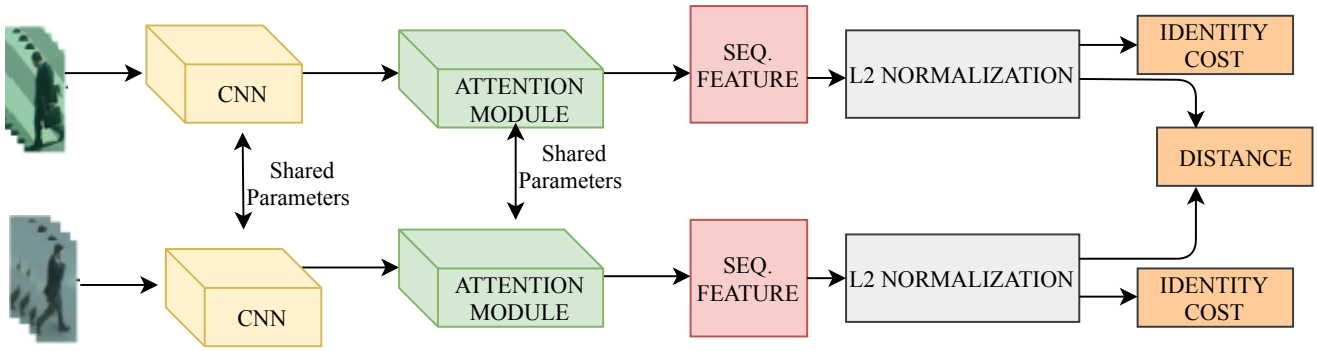


Figure 2: Overall architecture of our proposed Siamese network. It takes two input video sequences and pass to the Convolutional Neural Network (CNN) to extract features on each frame. The output from the CNN is fed to the attention module and generate an attention score for each frame. These attention scores combined with frame-level feature vectors to form a feature vector for the whole video. The video-level feature vectors are compared to decide whether the videos contain the same person.

3 OUR APPROACH

Figure 2 shows the overall architecture of our proposed approach based on the Siamese network [4]. The input to the Siamese network is a pair of video sequences corresponding to the query video and the candidate video to be compared. The output of the Siamese network is a scalar value indicating how likely these two videos contain the same person. Each video goes into one of the two branches of the Siamese network. Each branch of the Siamese network is a Convolutional neural network used to extract the features of the input video. The parameters of two branches of the Siamese network are shared. Finally, the features from the two input videos are compared to produce the final output.

When a video goes through one of the two branches of the Siamese network, we first extract per-frame features on each frame of the input video. Then we compute an attention score on each frame indicating how important this frame is for the re-identification task. The intuition is that not all frames in a video are informative. The attention scores enable our model to ignore certain frames and only pay attention to informative frames in the video. The attention scores are then used to aggregate per-frame visual features weighted by the corresponding attention score to form a feature vector for the entire video sequence. Along with calculation of temporal attention we also calculate the spatial attention for multiple hops over the input frame. Here the intuition is that in order to pay attention to multiple parts of the frame we need multiple attentions over it, which now allows useful information to be extracted without being hindered by occlusions and misalignments. We can repeat this process for several hops of attention (see Sec. 4.4), where each hop produces attention scores that focus more on the informative frames. Finally, the features of two input videos are compared to produce the output.

3.1 Frame-Level Features

Similar to [16] , we extract frame-level features using both RGB color and optical flow channels. The colors contain information about the appearance of a person, while the optical flows contain information about the movement of the person. Intuitively, both of them are useful to differentiate the identity of the person. As a preprocessing step, we convert all the input images (i.e. video frames) from RGB to YUV color space. We normalize each color channel to have a zero mean and unit variance. The Lucas-Kanade algorithm [13] is used to calculate both vertical and horizontal optical flow channels on each frame. We resize each frame to have a spatial dimension of 56×40. The optical flow field F of the frame is split into two scalar fields F_x and F_y corresponding to the x and y components of the optical flow. In the end, each frame is represented as a $56 \times 40 \times 5$ input, where the 5 channels correspond to 3 color channels (RGB) and 2 optical flow channels (x and y). We fine-tuned CNN architecture of [16] to extract frame-level features for an input video. The CNN architecture (shown in Fig. 3) consists of three stages of convolution, max-pooling, and nonlinear (tanh) activations. Each convolution filter uses 5×5 kernels with 1×1 stride and 4×4 zero padding. A fully connected layer is used in the end to produce a 128-dimensional feature vector. Given T frames in an input video, the CNN model is applied on each input frame of the video. In the end, CNN produces a 128-dimensional feature vector (i.e. $x_i \in \mathbb{R}^{128}$) to represent each frame x_i (i= 1,2,...,T) in the input video.

3.2 Temporal Attention Network

Motivated by the recent success of attention based models [1, 2, 23, 27], we propose an attention based approach for re-identifying person from videos. The intuition behind the attention based approach is inspired by the human visual processing [24]. Human brains often pay attention to different regions of different sequences when trying to re-identify persons from videos. Based on this intuition, we propose a deep Siamese architecture where each branch generates attention scores of different frames based on the frame-level CNN features. The attention score of a frame indicates the importance of this frame for the re-identification task.

As shown in Figure 2, each input video sequence (sequence of frames with optical flow) is passed to the CNN to extract frame-level

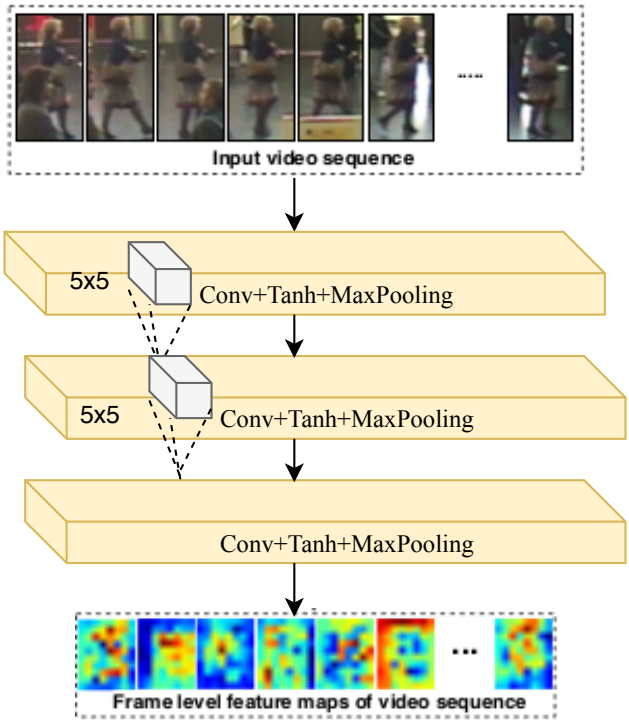


Figure 3: Our CNN architecture for extracting frame-level features. The network processes a frame (both color and optical flow channels) using a series of layers

feature maps. Using fully connected layers, CNN generates feature vector for each video frame. The sequence of feature vectors are passed to the attention network to generate attention score. More specifically, for each feature vector x_i corresponding to the i-th frame, we compute an attention score α_i^t indicating the importance of this frame. The attention score is obtained by applying a linear mapping followed by a sigmoid function. Here,we use the same parameters for the linear mapping on all frames. Let θ be the vector of parameters for the linear mapping. Now the attention score α_i^t is calculated using the following equations:

$$\alpha_i^t = \sigma(\theta^T x_i), \qquad i = 1, \cdots, T \tag{1}$$

Where $\sigma(\cdot)$ is the sigmoid function and α^t is the temporal attention matrix .We have also tried using softmax instead of sigmoid function in Eq. 1 and found that it does not perform as good as the sigmoid function. Previous work [29] has made similar observations. Once we have obtained an attention score α_i^t for each frame in the video, we can then combine the attention scores α_i^t(i = 1,2,...,T) with frame-level feature vectors to create a weighted feature vector f^t as follows:

$$f^t = \sum_{i=1}^{T} \alpha_i^t x_i \tag{2}$$

where f^t can be seen as a temporal feature vector for the entire video which takes into account the importance of each frame in the video.

3.3 Spatial Attention Network

For the task of person re-identification, due to several real time issues such as the overlooking angle of most of the surveillance equipments, the pedestrians who are captured only take a part in whole spatial images. Therefore, local spatial attention is necessary for deep networks such that useful information is extracted from each frame without getting hindered by problems such as occlusions and misalignments, and thus in this way superfluous information can be removed. In addition we employ multiple spatial attention model which can help us pay attention to multiple parts of the frame by hopping multiple times over it. In the experiment section, we will show that this multiple attention mechanism improves the performance of our model.

The basic idea of our spatial attention network is to pass the frame level features as input to the spatial attention network which has a convolution layer followed by sigmoid . Note that since our method evaluates multiple attention (see Fig.4), we will get j (where j is number of attention layers) feature vectors from the spatial attention network. The convolution filter ,$conv_j(\cdot)$, uses 5×5 kernels with 1×1 stride and 2×2 padding. $\alpha_i^{s_j}$ is the attention matrix which is multiplied by the frame level features x_i to get the weighted attention features as follows:

$$\alpha_i^{s_j} = \sigma(conv(x_i)), \quad i = 1, \cdots, T \tag{3}$$

$$f^{s_j} = \sum_{i=1}^{T} \alpha_i^{s_j} x_i, \qquad i = 1, \cdots, T \tag{4}$$

where f^{s_j} is the spatial attention feature vector of the input video. Now we add spatial attention feature vector f^{s_j} and temporal attention feature vector f^t and then sum it up for all attention layers to get the final feature vector as follows:

$$F_j = f^{s_j} + f^t \tag{5}$$

$$F = \sum_j F_j \tag{6}$$

We find out empirically the value of j to be 3 (See section 4.4 for more details).

3.4 Model Learning

Our model is a form of the Siamese network (Figure 2). It has two identical branches with shared parameters. The detail architecture of each branch is shown in Figure 4. Let F_1 and F_2 be the feature vectors of two input videos from the Siamese network. We now calculate Euclidean distance between the feature vectors in a manner similar to [16, 24] and apply the squared hinge loss ($Loss_{hinge}$) as follows:

$$\mathbb{L}_{hinge} = \begin{cases} 1/2 \left\| F_1 - F_2 \right\|^2, & X_1 = X_2 \\ 1/2 \lceil max(0, m - \left\| F_1 - F_2 \right\|) \rceil^2, & X_1 \neq X_2 \end{cases} \tag{7}$$

where the hyper-parameter m represents the margin of separating two classes in $\mathbb{L}_{hinge}$. Here we use X_1 and X_2 to represent the identities of the persons from two input videos.The idea is that if the two videos contain the same person (i.e. $X_1 = X_2$) the distance between the feature vectors should be small. Otherwise, the distance should be large if the persons are different (i.e. $X_1 \neq X_2$). Similar to [16], we also use another loss (i.e. identity loss $Loss_{id}$) to each

Unpublished working draft. Not for distribution.

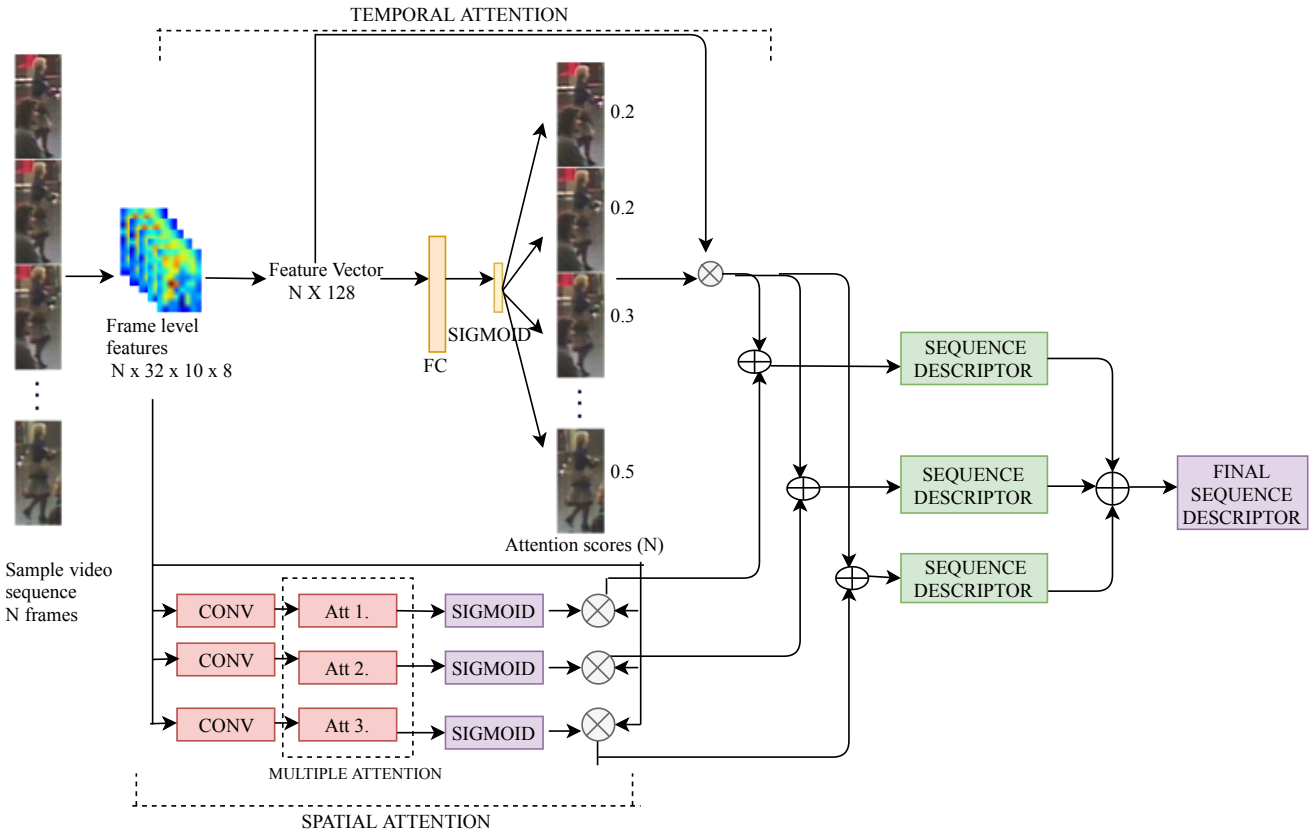


Figure 4: Illustration of our proposed attention network architecture. Attention module consists of 2 branches, in the temporal branch (upper) we generate N attention scores by applying linear mapping on the feature vectors followed by a sigmoid function. In the lower branch we calculate multiple spatial attention over the frame level features followed by a sigmoid function, spatial attention features are then added with temporal attention features to form final attention scores for the entire video.

branch of the Siamese network to predict the person's identity. We use a linear classifier to predict one of the person's identity from the feature vector extracted through each branch of the Siamese network. We then apply a Softmax loss over the prediction for each Siamese branch. The final loss is the combination of two identity losses ($Loss_{id1}$ and $Loss_{id2}$) from each Siamese branch and the hinge loss as follows:

$$\mathbb{L}_{final} = \mathbb{L}_{id1} + \mathbb{L}_{hinge} + \mathbb{L}_{id2} \tag{8}$$

We use stochastic gradient decent to optimize the loss function define in Eq. 8. After training, we remove all loss functions including the identity and hinge losses. During testing, we only use the feature vectors to compute the distance between two input videos for re-identification.

4 EXPERIMENTS

In this section, we firstly introduce the datasets used in our experiments (Sec. 4.1). We then describe the experimental setup and some implementation details (Sec. 4.2). We present the results of experiment in Sec 4.3 and Sec 4.4.

4.1 Datasets

We conduct experiments on two benchmark datasets: iLIDS-VID [21] and PRID-2011 [6].

iLIDS-VID Dataset: This dataset consists of video sequences of 300 persons where each person is captured by a pair of non-overlapping cameras. The length of each video sequence varies from 23 to 192 frames with an average of 73 frames. The dataset is quite challenging due to lot of occlusions, illumination changes, background clutters and so on.

PRID-2011 Dataset: This dataset contains video sequences of 749 persons. For the first 200 persons (or identities), there are two video sequences captured by two different cameras. The remaining persons appear in only one camera. Each sequence contains between 5 to 675 frames, with an average of 100 frames. In terms of complexity this dataset is relatively simple than iLIDS-VID.

Some sample frames of these three datasets are shown in Figure 5. Table 1 shows the summary of these two benchmark datasets.

Figure 5: Sample images of two benchmark datasets used in our experiments. The first two rows show sample images from the iLIDS-VID dataset captured by two different cameras. The next two rows show sample images from the PRID-2011 dataset.

Table 1: Summary of basic information of the two datasets used in our experiments.

Dataset	iLIDS-VID	PRID-2011
Total no. of id.	300	749
No. id in multiple cameras	300	200
No. track-lets	600	400
No. of boxes	44k	40k
Image resolution	64x128	64x128
No. of camera	2	2
Detection procedure	hand	hand
Complexity	Challenging	Simple

4.2 Setup and Implementation Details

We follow the experiment protocol of McLaughlin et al. [16]. On each of the two datasets (iLIDS-VID and PRID- 2011), we randomly split the dataset into two equal subsets where one subset is used for training and remaining one for testing. For evaluating our proposed method, we use the Cumulative Matching Characteristics (CMC) curve which is a ranking based evaluation metric. In the ideal case, the ground-truth video sequence should have the highest rank. For each dataset, we repeat the experiment 10 times and report the average result over these 10 runs. In each run, we randomly split the dataset into training/test sets. Standard data augmentation techniques, such as cropping and mirroring, are applied to increase the amount of training data. We initialize the weights in the network using the initialization technique in [5]. For training our network, we consider equal numbers of positive and negative samples. We set the margin in the hinge loss (Eq. 7) as m = 2. The network is trained for 1100 epochs with a batch size of one. The learning rate in the stochastic gradient descent is set to be $1e^{-4}$. Due to the variable-length of video sequences in both datasets, we use sub-sequences of 16 consecutive frames (T = 16) during training. Sometimes, this length is greater than the real sequence length. In that case, we consider the whole set of images (frames) as the sub-sequence. A full epoch consists of a pair of positive and negative sample. During testing, we consider a video sequence captured by the first camera as the probe sequence and a video sequence captured by the second camera as a gallery sequence. We use at most 128 frames in a testing video sequence. Again, if the length is greater than the real sequence, we consider the whole set of images as the video sequence. Similar strategies have been used in previous work [16].

4.3 Results

We present the results on the two benchmark datasets and compare with other state-of-the-art methods in Table 2 and Table 3. From

Unpublished
Not for distri

the CMC rank, we see that our method outperforms all other state-of-the-art methods by nearly 4% and 11% on rank-1 accuracy on the iLIDS-VID and PRID-2011 dataset,respectively. Figure 6 shows some qualitative retrieval results after applying our proposed method on the challenging iLIDS-VID dataset. We also show some failure cases in Figure 7.

Table 2: Comparison of our proposed approach with other state- of-the-art methods on the iLIDS-VID dataset in terms of CMC(%) at different ranks.

Dataset	iLIDS-VID			
Method	Rank-1	Rank-5	Rank-10	Rank-20
Ours	**66**	**90**	**95**	**99**
Xu et al.[26]	62	86	94	98
Zhou et al.[35]	55.2	86.5	-	97
McLaughlin et al.[18]	58	84	91	96
Yan et al.[27]	49.3	76.8	85.3	90.1
STA[14]	44.3	71.7	83.7	91.7
VR[23]	35	57	68	78
SRID[8]	25	45	56	66
AFDA[10]	38	63	73	82
DTDL[7]	26	48	57	69

Table 3: Comparison of our proposed approach with other state- of-the-art methods on the PRID-2011 dataset in terms of CMC(%) at different ranks.

Dataset	PRID-2011			
Method	Rank-1	Rank-5	Rank-10	Rank-20
Ours	**88**	**97**	**99**	**99**
Xu et al.[26]	77	95	99	99
Zhou et al.[35]	79.4	94.4	-	99.3
McLaughlin et al.[18]	70	90	95	97
Yan et al.[27]	58.2	85.8	93.7	98.4
STA[14]	64.1	87.3	89.9	92
VR[23]	42	65	78	89
SRID[8]	35	59	70	80
AFDA[10]	43	73	85	92
DTDL[7]	41	70	78	86

4.4 Effect of Multiple Attention

We conduct empirical study on the training set of the iLIDS-VID and PRID-2011 dataset to analyze the effect of the multiple attention on the overall performance of the proposed network. We train the model on the training videos and report the performance (CMC%) on the validation set for different number of attention layers in Table 4 and Table 5 respectively.

We observe that the performance gradually improves until number of attention layer is 3. After that, the performance starts to drop. Based on this empirical result, we choose number of attention layers as 3 in our experiments.It can also be observed from Table 4 and Table 5 that when the number of attention layers is 0, i.e. when we do not incorporate the concept of spatial attention the accuracies fall by a huge margin.

Table 4: Performance for different number of attention layers on the iLIDS-VID dataset. Again, we report the performance in terms of CMC %

Dataset	iLIDS-VID			
No. of Att. layers	Rank-1	Rank-5	Rank-10	Rank-20
3	66	90	95	99
2	64	88	95	99
1	63	88	95	98
0	60	88	94	98

Table 5: Performance for different number of attention layers on the PRID-2011 dataset. Again, we report the performance in terms of CMC %

Dataset	PRID-2011			
No. of Att. layers	Rank-1	Rank-5	Rank-10	Rank-20
3	88	97	99	99
2	84	96	99	99
1	80	95	98	99
0	80	96	99	99

Unpublished working draft. Not for distribution.

5 CONCLUSIONS

In this paper, we have proposed an attention-based deep architecture for video-based re-identification. The attention module calculates frame-level attention scores, where the attention score indicates the importance of a particular frame. We perform experiments on two benchmark datasets and compare with other state-of-the-art approaches. We demonstrate that our proposed method outperforms to other state-of-the-art approaches.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *International Conference on Learning Representations.*

[2] Cıcero Nogueira dos Santos, Ming Tan, Bing Xiang, and Bowen Zhou. 2016. Attentive pooling networks. *Computing Research Repository* (2016).

[3] Douglas Gray and Hao Tao. 2008. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features. In *European Conference on Computer Vision.*

[4] Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality Reduction by Learning an Invariant Mapping. In *IEEE Conference on Computer Vision and Pattern Recognition.*

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving Deep Into Recitifers: Surpassing Human-Level Performance on ImageNet Classification. In *IEEE International Confernece on Computer Vision.*

[6] Martin Hirzer, Csaba Beleznai, Peter M. Roth, and Horst Bischof. 2011. Person Re-identification by Descriptive and Discriminative Classification. In *Scandinavian Conference on Image Analysis.*

Figure 6: Qualitative retrieval results of our proposed method on the challenging iLIDS-VID dataset. The first column represents the probe video sequence. The remaining columns correspond to retrieved video sequences sorted by their distances to the probe video sequence. Here, we use a single image to represent each retrieved video sequence. The green boxes indicate the ground-truth matches. We can see that the ground-truth matches are ranked very high in the list.

Figure 7: Examples of some failure case of our proposed method. The first row indicates the probe sequence where single image in second row represents retrieve gallery sequence of corresponding person

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*.

[8] Youjiao Li, Li Zhuo, Jiafeng Li, Jing Zhang, Xi Liang, and Qi Tian. 2017. Video-Based Person Re-identification by Deep Feature Guided Pooling. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*.

[9] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z. Li. 2015. Person Re-Identification by Local Maximal Occurence Representation and Metric Learning. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[10] Shengcai Liao and Stan Z. Li. 2015. Efficient PSD Constrained Asymmetric Metric Learning for Person Re-identification. In *IEEE International Conference on Computer Vision*.

[11] Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. 2017. A structured self-attentive sentence embedding. *In International Conference on Learning Representations*.

[12] Kan Liu, Bingpeng Ma, Wei Zhang, and Rui Huang. 2015. A Spatio-Temporal Appearance Representation for Video-based Pedestrian Re-Identification. In *IEEE International Conference on Computer Vision*.

[13] Bruce D Lucas and Takeo Kanade. 1981. An Iterative Image Registration Technique with an Application to Stereo Vision. *International joint conference on Artificial intelligence* (1981).

[14] Bingpeng Ma, Yu Su, and Frederic Jurie. 2012. Local Descriptors Encoded by Fisher Vectors for Person Re-identification. In *European Conference on Computer Vision Workshop*.

[15] Tetsu Matsukawa, Takahiro Okabe, Einoshin Suzuki, and Yoichi Sato. 2016. Hierarchical Gaussian Descriptor for Person Re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[16] Niall McLaughlin, Jesus Martinez del Rincon, and Paul Miller. 2016. Recurrent Convolutional Neural Network for Video-based Person Re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[17] Xuelin Qian, Yanwei Fu, Yu-Gang Jiang, Tao Xiang, and Xiangyang Xue. 2017. Multi-scale Deep Learning Architectures for Person Re-identification. In *IEEE International Conference on Computer Vision*.

[18] Kevin J. Shih, Saurabh Singh, and Derek Hoiem. 2016. Where To Look: Focus Regions for Visual Question Answering. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[19] Evgeniya Ustinova, Yaroslav Ganin, and Victor Lempitsky. 2017. Multiregion Bilinear Convolutional Neural Networks for Person Re-Identification. In *IEEE International Conference on Advanced Video and Signal based Surveillance*.

[20] Rahul Raman Varior, Bing Shuai, Jiwen Lu, Dong Xu, and Gang Wang. 2016. A Siamese Long Short-Term Memory Architecture for Human Re-Identification. In *European Conference on Computer Vision*.

[21] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. 2014. Person Re-Identification by Video Ranking. In *European Conference on Computer Vision*.

[22] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. 2016. Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[23] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, and Yoshua Bengio. 2015. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In *International Conference on Machine Learning*.

[24] Shuangjie Xu, Yu Cheng, Kang Gu, Yang Yang, Shiyu Chang, and Pan Zhou. 2017. Jointly Attentive Spatial-Temporal Pooling Networks for Video-based Person Re-Identification. In *IEEE International Conference on Computer Vision*.

[25] Yichao Yan, Bingbing Ni, Zhichao Song, Chao Ma, Yan Yan, and Xiaokang Yang. 2016. Person Re-Identification via Recurrent Feature Aggregation. In *European Conference on Computer Vision*.

[26] Dong Yi, Zhen Lei, and Stan Z. Li. 2014. Deep Metric Learning for Person Re-identification. In *IAPR International Conference on Pattern Recognition*.

[27] Wenpeng Yin, Hinrich Schutze, Bing Xiang, and Bowen Zhou. 2016. ABCNN: Attention-Based Convolutional Neural Networks for Modeling Sentence Pairs. *Transactions of the Association for Computational Linguistics* (2016).

[28] Li Zhang, Tao Xiang, and Shaogang Gong. 2016. Learning a Discriminative Null Space for Person Re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[29] Liming Zhao, Xi Li, Yueting Zhuang, and Jingdong Wang. 2017. Deeply-Learned Part-Aligned Representations for Person Re-Identification. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[30] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. 2014. Learning Mid-level Filters for Person Re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[31] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Salable Person Re-identification: A Benchmark. In *IEEE International Conference on Computer Vision*.

[32] Zhen Zhou, Yan Huang, Wei Wang, Liang Wang, and Tieniu Tan. 2017. See the Forest for the Trees: Joint Spatial and Temporal Recurrent Neural Networks for Video-based Person Re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[33] Xiaoke Zhu, Xiao-Yuan Jing, Fei Wu, and Hui Feng. 2016. Video-based person re-identification by simultaneously learning intra-video and inter-video distance metrics. In *International Joint Conference on Artificial Intelligence*.

Unpublished working draft. Not for distribution.