SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(80TH Edition)

AUGUST 2019

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 **Eightieth** 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 **August, 2019**. 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. 80/20 vis-à-vis Core Journals: Use of e- ShodhSindhu Resources at the University of Delhi, ***4.1.Dr. Lalita*** and Dr. Narender Kumar, Library DTU

2. A Novel Framework for Neural Architecture Search in the Hill Climbing Domain, ***8..Mudit Verma***, ***8.Pradyumna Sinha***, ***8.Karan Goyal***, ***8.Apoorva Verma*** and ***3.Seba Susan***, IT, DTU

3. Anomaly detection in group activities based on fuzzy lattices using Schrödinger equation, ***3.Rajiv Kapoor[1]***, ***6.Om Mishra[1]*** and ***3.M. M. Tripathi[2]***, [1]Electronics and [2]Electrical, DTU

4. Ant Colony Optimization and Excess Energy Calculations Based Fast Converging Energy Efficient Routing Algorithm for WSNs, ***7.1.Aarti Jain*** and Anuj Pathak, Electronics, DTU

5. Biofunctionalized Nanostructured Yttria Modified Non-Invasive Impedometric Biosensor for Efficient Detection of Oral Cancer, ***6.Suveen Kumar***, ***Shweta Panwar*** , ***7.Saurabh Kumar*** , ***7.Shine Augustine*** and ***3.Bansi D. Malhotra***, Biotechnology, DTU

6. Controlling the radiation pattern of a microstrip patch antenna using a checkerboard patterned metasurface, ***6.Nishant Shankhwar***, ***6.Ritika Ranga***, ***3.Yogita Kalra*** and ***3.Ravindra Kumar Sinha***, Applied Physicsm, DTU

7. Deformation Behaviour of Tunnels under Different Loading Conditions: State of Art, ***6.Parvesh Kumar*** and ***3.Amit Kumar Shrivastava***, Civil, DTU

8. Design and Analysis of Microstructured Optical Fiber for Supercontinuum Generation, ***6.Ankit Sharma***, ***6.Pooja Chauhan***, ***6.Ajeet Kumar***, ***3.Yogita Kalra*** and Than Singh, Applied Physics, DTU

9. Design Considerations and Capacitance Dependent Parametric Assessment of Gate Metal Engineered SiNW MOSFET for ULSI Switching Applications, Neha Gupta, Ajay Kumar and ***3.Rishu Chaujar***, Applied Physics, DTU

10. Effect of Mast Modification in Ingress Problem in Gas Turbine of Naval Ship, P S Yadav, V Saxena, H S Pali, ***3.N Kumar*** and S N Singh, Mechanical, DTU

11. Effects of La and Pb doping on structural and dielectric properties of Sr1-xAxBi2Nb2O9 (A=La, Pb and x=0.1) ferroelectric ceramic, ***6.1.Vaibhav Shrivastava***, A.K.Jha and R.G.Mendiratta, Applied Physics, DTU

12. Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets, ***3.Akshi Kumar*** and ***3.Geetanjali Garg***, CSE, DTU

13. EXERGO-ENVIRONMENTAL ANALYSIS OF AN INDIRECT FORCED CONVECTION SOLAR DRYER FOR DRYING BITTER GOURD SLICES, S. Vijayan, T.V. Arjunan and ***3.Anil Kumar***, Mechanical, DTU

14. Exploring High Performance Work Practices as Necessary Condition of HR Outcomes, ***3.Naval Garg***, B. K. Punia and Anuradha Jain, DSM, DTU

15. Formulation and application of quantum-inspired tidal firefly technique for multiple-objective mixed cost-effective emission dispatch, Kapil Deo Bodha, ***3.Vinod Kumar Yadav*** and Vivekananda Mukherjee, Electrical, DTU

16. Impact of metal silicide source electrode on polarity gate induced source in junctionless TFET, Jaya Madan, Rahul Pandey, Rajnish Sharma and ***3.Rishu Chaujar***, Applied Physics, DTU

17. INVERSE COEFFICIENT BOUNDS FOR CERTAIN CLASSES OF CLOSE TO CONVEX FUNCTIONS, ***6.PRIYANKA GOEL*** and ***3.S. SIVAPRASAD KUMAR***, Mathematics, DTU

18. Investigating Political Herd Mentality: A Community Sentiment Based Approach, ***8.Anjali Bhavan***, ***8.Rohan Mishra***, ***8.Pradyumna Prakhar Sinha***, Ramit Sawhney and Rajiv Ratn Shah,IT, DTU

19. Investigating the effect of different carrier gases on plasma-assisted multiwalled carbon nanotube growth, ***6.U. Sharma*** and ***3.S. C. Sharma***, Applied Physics, DTU

20. Lower hybrid waves excitation by a relativistic electron beam in a magnetized dusty plasma: kinetic theory, ***6.Anshu***, ***3.S. C. Sharma*** and J.Sharma, Applied Physics, DTU

21. Mechanical suitability of carbon fiber based composite for chopart prosthetic socket, F H Abdulsadah, ***3.Q Murtaza***, F Hasan and S Bhardwaj, Mechanical, DTU

22. Microstructural and wear characterisation of as-cast AS21A magnesium alloy, ***6.Sumit Joshi***, ***3.Ramesh Chandra Singh*** and ***3.Rajiv Chaudhary***, Mechanical, DTU

23. Microwave derived monoclinic Ba1-xSnxNb2O6 materials as an alternative of ITO, ***6.Ashish Joshi***, Vaibhav Shrivastava and Anurag Pritam, Applied Physics, DTU

24. Modelling and simulation of repairable mechanical systems reliability and availability, ***3.Girish Kumar***, Vipul Jain and Umang Soni, Mechanical, DTU

25. Molecular characterization, modeling, and docking analysis of late phytic acid biosynthesis pathway gene, inositol polyphosphate 6-/3-/5-kinase, a potential candidate for developin..., ***6.MANSI PUNJABI***, ***3.NAVNEETA BHARADVAJA***, ARCHANA SACHDEV and VEDA KRISHNAN, Biotechnology, DTU

26. Optimal Design of Gravity-Fed Sewer Lines Using Linear Programming, ***6.Deepak Singh***, P. S. Mahar and R. P. Singh, Civil, DTU

27. Performance and Emission Characteristics of Ternary Fuel Blends in Diesel Engines, ***3.Naveen Kumar*** and ***6.Sidharth***, Mechanical, DTU

28. PHYTIC ACID DOPED POLYANILINE NANOFIBERS: AN ADVANCED ADSORBENT FOR METHYLENE BLUE DYE, ***6.Monika Duhan*** and ***3.Raminder Kaur***, Applied Chemistry, DTU

29. Plasma bubble evolution related electron beam parameter estimation in laser wakefield acceleration, ***6.M. Yadav***, ***3.S. C. Sharma*** and D.N. Gupta, Applied Physics, DTU

30. Prediction of Compressive Strength of Cement Mortar in Normal and Aggressive Environment Using Artificial Neural Network, Kavita Verma, Shobha Ram and ***3.Alok Verma,*** Civil, DTU

31. Robust object tracking with crow search optimized multi-cue particle filter, © Gurjit Singh Walia, ***6.Ashish Kumar***, Astitwa Saxena, ***3.Kapil Sharma*** and Kuldeep Singh, IT, DTU

32. Role of Co-Solvent in the Super Critical Esterification process of Bio-Diesel Production through Karanja oil, ***6.Chandra Shekhar Singh***, ***3.Naveen Kumar***, DTU

33. Stabilization of Plane Polarized Alfven Waves by Anomalous Doppler Resonance, ***6.Rajesh Gupta***, ***3.Suresh C. Sharma*** And Ruby Gupta, Applied Physics, DTU

34. System-level Performance of Chip-based Brillouin Microwave Photonic Bandpass Filters, Yiwei Xie, ***8.1.Amol Choudhary***, Yang Liu, David Marpaung, Khu Vu, Pan Ma, Duk-Yong Choi, Steve Madden and Benjamin J. Eggleton, DTU

35. Theoretical Model for Self-trapping of Gaussian Electromagnetic Beam in Dusty Plasma, ***3.R. Sharma*** and ***3.Suresh C. Sharma***, Applied Physics, DTU

1. *Chancellor*

2. *Pro Vice Chancellor* ***2.1.*** *Ex Pro Vice Chancellor*

3. *Faculty* ***3.1.*** *Ex Faculty*

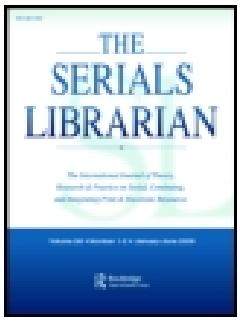
4. *Teaching-cum-Research Fellow* ***4.1.****Asst. Librarian*

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*



The Serials Librarian
From the Printed Page to the Digital Age

ISSN: 0361-526X (Print) 1541-1095 (Online) Journal homepage: https://www.tandfonline.com/loi/wser20

80/20 vis-à-vis Core Journals: Use of e-ShodhSindhu Resources at the University of Delhi

Lalita & Narender Kumar

To cite this article: Lalita & Narender Kumar (2019): 80/20 vis-à-vis Core Journals: Use of e-ShodhSindhu Resources at the University of Delhi, The Serials Librarian, DOI: 10.1080/0361526X.2019.1630347

To link to this article: https://doi.org/10.1080/0361526X.2019.1630347

Published online: 01 Aug 2019.

Submit your article to this journal

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=wser20

THE SERIALS LIBRARIAN
https://doi.org/10.1080/0361526X.2019.1630347

Routledge
Taylor & Francis Group

Check for updates

80/20 vis-à-vis Core Journals: Use of e-ShodhSindhu Resources at the University of Delhi

Lalita and Narender Kumar

ABSTRACT

The study aims to identify the core collection of the University of Delhi from electronic resources provided by e-ShodhSindhu, consortia for higher educational institutes of India on the basis of Pareto 80/20 rule. The study is based on empirical data downloaded from the e-ShodhSindhu portal for the period of 2010–2017. The usage has been statistically analyzed by different parameters such as average download per year, % of journals that cover 80% of downloads, number of titles contributed by each resource into the 80% downloads, identifying the core titles across the resources covering 80% of the downloads, identification of lesser/unused titles with the help of tables and graphs. The results suggest that the Pareto's theory needs to be relooked in the bundled subscription. In the case of the University of Delhi, 90% demands can be met by subscribing less than 20% of the titles or 10% titles can satisfy 80% of the information needs.

KEYWORDS

Core journals; electronic resources; usage; consortia; Pareto 80/20 Rule

Introduction

Today libraries are experiencing the burden of fulfilling expectations of users and meeting the challenges of financial management due to a decrease in the budget and an increase in the cost of resources. Libraries have also been facing a world of increased complexities in bringing barrier-free access to users. Also, publishers' different price models have posed other challenges. In the print environment, the libraries earlier used different techniques like the application of Velfredo Pareto's theory to identify the core collection of the library and then acquired them. One of the authors, while working with the British Council Library, Delhi, used to prepare a stock revision list for a particular subject representing the core collection in that subject. The library checked the existing stock of the subject against the list and ordered titles that were not found in the library collection.

In the print environment, the librarians are not pressurized by the publisher, as the unit cost of an item is the same for all libraries irrespective of the number of users, level of the institute, the capacity to pay, etc. There is hardly any concept of consortia or bundled subscription, big deals, etc. However, with an introduction of the World Wide Web and online distribution of journals, the pricing policies have changed drastically. In a digital environment, the publishers charge a different price from different libraries for the same unit. The pricing of digital resources depends on the number of users, level of the institute, the capacity to pay, consortia arrangements, etc. For journal subscriptions, most of the publishers offer bundled subscription to the library. In case the library wishes to adopt 'pick and choose model,' then the publishers quote a list price (generally print price) + e-access fee of journals, where the total price of selected journals might be higher than the bundled price. Hence, the libraries are forced to subscribe to bundled instead of core journals.

CONTACT Narender Kumar narenderkumar59@yahoo.com

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/wser.

Published with license by Taylor & Francis Group, LLC
© 2019 Lalita and Narender Kumar

Academic journals: Pricing

In an academic world, there are two types of journal publications, i.e., for-profit and not-for-profit. Though the intent of both is to disseminate knowledge, both have a different pricing strategy. The motive of for-profit publishers is to maximize the profit, whereas the motive of non-profit publishers is to cover the cost. However, this price difference does not reflect a difference in quality, format or editorial processes. The articles in non-profit journals receive more citations than for-profit journals.[1]

Technology-based publishing and advent of the Internet for communication have made possible of a less expensive method of scholarly communication. Unfortunately for libraries, this has not happened and the budget to subscribe journals has substantially been increasing faster than the consumer price index (CPI) as the publishers generally try to relate the cost of print and online subscription. They charge online cost equal to that of print subscription cost plus a surcharge on e-access which generally ranges from 15% to 30% of the print cost.[2] Based on this reasoning the budget to subscribe to electronic journals has increased substantially and the library needs to increase it constantly, faster than consumer price index (CPI), to sustain these subscriptions.

Publishers' logic on price increase of electronic journals

Publishers have been putting forward the following factors to defend the substantial increase in prices of electronic journals:

(1) *Infrastructure Cost*: Publishers claim that the technology-based electronic publishing does not reduce infrastructure cost. The fact that articles need to be written, reviewed, typeset, archived, and made accessible, etc., also states that infrastructure and investment is needed.[3] The publishers may administer the peer-review process, which includes activities such as finding peer reviewers, evaluating the assessments and checking manuscripts for plagiarism. They may edit the articles, which includes proofreading, typesetting, adding graphics, turning the file into standard formats, such as XML.[4] The publishers also claim that they have to invest in data storage and transmission cost and this investment grows further as files grow and new technology is introduced. If the cost of these elements increases, the cost of electronic journals will continue to increase.
(2) *New Features*: Another logic put forward is the introduction of new features, not available in print format, like search facilities, the inclusion of 10-year previous articles with the current subscription, and linked references, etc. Another new provision is that of providing alerts to the users.
(3) *Metadata and Usage Administration*: The publishers create metadata and store comprehensive usage statistics of electronic journals according to agreed industry international standards. These facilities take the responsibility from the libraries of maintaining them. To provide these facilities, the publishers need to invest.
(4) *More subscription Option*: Distribution of electronic journals through the Internet gives more options to access an article, like pay per view options, purchase through coupons, etc.
(5) *Protection against Predatory Journals*: Butler in his article advocates that the big and established publishers help in the protection of authors against the predatory open-access publishers who do not provide any service but charge from them.[5]
(6) *Increase Purchasing Capacity*: Electronic publishing allows libraries to form library consortia, increasing the purchase capacity and expanding resource accessibility.

Librarians and others counterviews

Librarians and scholars frequently complain that commercial publishers invariably raise subscription rates every year without enhancing the quality and despite the fact that electronic publishing offers

the possibility of a 10-fold decrease of basic journal publishing cost.[6] Association of Research Library (ARL) reported that "during 1986–2002, the unit cost for journal subscriptions grew at a rate of 7.7 percent per year, which is more than twice the growth rate of the unit cost of monographs, 3.6 percent."[7] Hence, the librarians and others expressed the following observations to counter the logic put forward by the publishers:

- *Infrastructure Cost*: New technology-based electronic publishing does need more money to produce the first copy but subsequent copies require almost zero marginal cost. Moreover, the distribution cost is also zero due to www.[8] It is not only the service provider but also the service receivers who have to invest in the new technology to access the resources, which is an extra burden on libraries besides the subscription cost. Moreover, libraries need to employ expert personnel in technology.
- *Free Labour*: Michael Eisen, co-founder of *PLOS*, showed his anger when he said: "the biggest travesty is that the scientific community carries out peer review – a major part of scholarly publishing – for free, yet subscription-journal publishers charge billions of dollars per year for scientists to read the final product."[9] A researcher conducts research in the organization that has employed him, compiles and types the findings on a computer and sends it for publication so that his research can be made available to the public. The experts, employees of their respective organizations, in the editorial board of a journal, also work as independent researchers and fulfill their scholarly duties. On top of this, these experts also arrange and select referee/referees, not on the payroll of the publishers. After the evaluation by the referee/referees, the editorial board decides whether or not to accept the submission. The accepted articles are passed on to the publisher for publication. Thus, the major intellectual work of any research work is performed by the free labor.
- *Cheaper Hosting and Computing Data Storage*: It is getting cheaper by the day with the introduction of cloud computing. According to Citigroup "in the past three years, prices have gone down by around a quarter and further significant falls look all but inevitable."[10]
- *Less Distribution Cost*: Publishing and distribution costs have changed drastically with the introduction of computer-based publishing, www and Internet technologies, etc. These new technologies, though not free, are much less expensive, especially with respect to the publishing and distribution cost.
- *Cheaper Hardware*: The investment in technology, which is advancing, is becoming cheaper every day. The US Bureau of Labour Statistics, which analyzed the prices for the last 18 years for broad categories, reported that the "prices had dropped dramatically in almost every tech sector. The drop in computer hardware is particularly steep."[11]
- *Cheaper Data Tariff*: Data tariff has been reducing with an even faster speed. Telecom Regulatory Authority of India (TRAI) reported that average wireless data tariff (outgoing) has drastically declined from Rs.268.97 per GB in 2014 to a mere Rs.19.35 in 2017.[12] Furthermore, mobile phone tariff has decreased from Rs.16 per minute in July 1995 to almost free in 2018.
- *No Qualitative Difference*: Both types of journals, i.e., not-for-profit and for-profit journals, cannot be differentiated by their contents, look, etc. In some cases, the contents of not-for-profit journals are of a higher quality than for-profit journals. Bergstrom reported that "the six most cited journals in economics are owned by non-profits, and only five of the twenty most-cited are owned by for-profit publishers."[13]
- *Auditors Observations*: The auditors insist on using technology during auditing, since it is easier to monitor the use of each title with its use. This ends up raising questions for the titles which have less/no use.

Big deal

Kenneth Frazier coined the term 'Big Deal', which means licensing all the journal content that a publisher distributes digitally.[14] In a big deal, libraries sign an arrangement to get electronic access to either the complete collection (like freedom collection of *Science Direct*) or a part of it at a substantial discount. In the early years of big deals, the libraries agreed that they would not cancel the titles in print. In a 'big deal', the publishers charge the same amount from the libraries that they are already paying for their journals, with some increment for electronic access. Varied terms like bundled subscription, one price one size, all-or-nothing, etc., are used in the libraries to refer to the big deal.

- *Advantages to Publishers*: In case of big deals, the service provider benefits the most as the simplicity of a single price results in lowering marketing and selling costs. Big deals also help in speeding up the execution process and selling a larger volume of products. Another advantage of big deals for the publishers is that they sell their weaker or inferior journals with its more mainstream or stronger journals. This helps them to charge a higher price for such journals.
- *Advantages to Libraries*: The libraries/scholars get an extra number of journals to access with the same price or with a minor increase, resulting in the expansion of their collection. The big deal also helps smaller journals where some of the most interesting science starts.

Voices against big deals

Since starting, a majority of the librarians have been raising their voice against the big deal, as it serves only the big publishers. Librarians have been requesting to "not sign on the Big Deal or any comprehensive licensing agreements with commercial publishers."[15] The reasons put forward by the librarians are:

- *Price Discrimination*: The loudest voice is against the price discrimination, which was not present earlier in case of a print subscription. The publishers bind the libraries through a legal agreement to not disclose their prices to other libraries. However, some libraries have been able to acquire the price information of other libraries. For example, in 2007, one of the most reputed science publishers charged $1,961,938.75 for unlimited access to its entire collection of journals from one university and $442,224.78 for unlimited access for the same collection of journals from another university.[16]
- *Faulty Pricing Strategy*: The publishers have a faulty pricing strategy when libraries wish to consider a big deal. The libraries subscribing bulk titles of a collection pay more to have access to the complete collection whereas the libraries subscribing a small number of titles of that collection pay a lesser amount for the same complete collection.[17]
- *Weakens Collection*: The big deal weakens the journal collection as the libraries are forced to subscribe journals which they neither need nor want.
- *Dependence on Few Publishers*: Big deal increases the dependence of libraries on big publishers, who have already captured information marketplace, hence curtailing the purchasing capacity of libraries to subscribe to journals from small publishers. This also creates a barrier in accessing information from the new journals or smaller publishers.
- *Denial of Cancellation*: The contract signed between the libraries and the publishers denies libraries the right to cancel the less used or unused journals under big deals.

Core collection

With the decreasing financial resources of the libraries, it would be useful for the libraries to identify heavily used (Core) titles in a bundle among other titles. Core collection, a well-established collection

management concept, was introduced by Hardesty and Mack.[18] The core collection refers to the titles which form the center of the collection around which the whole collection is developed for that subject. It could be core literature, a core list or a core journal. Intner defined core as "the nucleus of needed materials no self-respecting library would be willing to do without."[19] Core collection has been defined as "material that represents the major information interests of the users of a library" and core literature as "material considered essential for the study of a particular subject"[20] or "A list of the best books, periodicals, etc., on a subject or in a discipline … "[21]

Methods to identify core collection

Nisonger lists out 10 methods to identify core collection – subjective judgment by experts, usage, coverage by abstracting and indexing services, overlapping library holdings/subscription, citation data, total citation count in a set of source documents, coverage in the journal citation reports, discipline impact factor, citation network/co-citation analysis, and production of articles.[22] In the present study, one of the methods, i.e., usage has been used to identify the core collection for the University of Delhi.

Use of usage metric

A variety of metrics are used to highlight the quality and the influence of journals. The impact factor, based on citation metrics, is the frequently used and influential metric. However, doubts have been raised on this metric because (i) it is biased towards US journals; (ii) it can be manipulated by self-citation; (iii) it can be overused, misused, and over-interpreted; (iv) it has time-lag; (v) new journals have no impact factor; (vi) it reflects only author use of journals not other groups, such as students; etc. Considering the criticism of impact factor, the librarians, authors, and publishers are of the opinion that a credible alternative to impact factor should be used which can provide a more universal, quantitative and comparable measure of journal value. Here it is worth mentioning the findings of a research project funded by United Kingdom Serial Group (UKSG) which states "It is generally acknowledged that no such alternative currently exists, but the usage data could be basis of such a measure in the future. Around 70 percent of authors surveyed would welcome a new, usage-based measure of the value of scholarly journals."[23]

However, collecting usage statistics is a cumbersome task as "some vendors may be reluctant to supply usage statistics because they are fearful that low or uneven usage may cause libraries or consortia to demand changes in pricing, terms, and scope of content when license agreements come up for renewal."[24] As no metric is perfect, there are some issues with the usages statistics metric as well, like (i) non-compliance of standards; (ii) different format; (iii) double clicking – one for HTML and other for PDF; (iv) inflated statistics; (v) lack of web-based interface; (vi) various interfaces; (vii) different reporting platforms; and (viii) duplication of journals in different databases/bundles.

80/20

Vilfredo Pareto, an Italian economist, states that roughly 80% of the effects come from 20% of the causes.[25] The "usual Pareto 80–20 distribution predicts that 80 percent of the usage (effects) would be related to 20 percent of the journals (causes)."[26] Richard W. Trueswell was the first librarian who tried to apply Pareto's 80–20 rule to the library collection when he found that approximately 80% use of books and serials in libraries was accounted for by about 20% of the collection.[27] Nisonger in his study also concluded that roughly 25% of the material would be required to satisfy 80% of usage.[28] Singson and Hangsingre affirmed the same trend by conducting a study of the usage of electronic journals across academic institutions in India.[29]

Importance of pareto's principle

The big commercial publishers merge weak (irrelevant) titles with strong (relevant/core) titles. This tactic is alternatively referred to as "the Law of the Vital Few and the Trivial Many."[30] To identify the vital few (Core/relevant) titles "80/20 version seems to have become not merely an accepted rule for a well-run library but even a yardstick against which the efficiency of a library might be judged."[31] However, Crawford counters the 80/20 concept and said that "good libraries work to handle the special needs of the other 20 percent." He also emphasized that "libraries are most vital for the 20 percent who cannot afford to buy their own material."[32]

The libraries are forced by the big commercial publishers to accept their offer of bundled subscription in the pretext that the users will have more number of journals to access, resulting in increase in the number of irrelevant journals. In the process, libraries are compromising on the quality of collection even while increasing the quantity of collection. Hence, the library should evaluate the usage and identify the numbers (qualitative titles) which meet the maximum information needs of the users. The library should also rethink whether a one-size-fits-all approach to providing access to library materials is sensible, or whether only the most used 20% should be prioritized in terms of access.

In the context of the application of the Pareto's principle – 80% information needs of the users can be met by 20% of documents available in libraries, as has been shown by various research studies that have been conducted. Michael Chiorazzi, in his study, concluded that 90% of the research needs can be met by 20% of the material. He also emphasized the need to relook at the collection development policies and considered the cost of new space and the cost of acquiring seldom-used material.[33]

Young reported that only 16.4% periodicals in bibliographic databases meet the demands of 80% of the users in University College of Cape Breton.[34] Another study by Black and Sisson reported that only 12.44% to 18.37% of full-text databases of four colleges satisfy the information needs of 80% of their users.[35]

E-shodhsindhu

e-ShodhSindhu consortium came into being on December 1, 2015, after merging three consortia, i.e., UGC-INFONET Digital Library Consortium (2004), NLIST (2014), and INDEST-AICTE (2003). The recommendation for merging was proposed by an expert committee and the order was issued by the Ministry of Human Resources, Government of India.

Till December 2017, the e-ShodhSindhu was financing all resources available through the consortium but from January 2018, after an executive order, the consortium financed only selective resources. For the remaining, the respective institutes were asked to make arrangement of funds from their own resources. However, the consortium continued to negotiate with the publishers for its members, which included 91 central government-funded technical institutes, 217 universities, 108 All India Council for Technical Education recognized institutes and 3000 colleges. The INFLIBNET Centre, an autonomous Inter-University Center of University Grants Commission, had been assigned the responsibility for execution and operation of the consortium.[36]

The main objective of the consortium is to provide access to qualitative electronic resources including full-text, bibliographic and factual databases to member institutes at a lower rate of subscription. The consortium is intending to develop a national electronic library with electronic journals and electronic books as its major building blocks.

University of Delhi

The University of Delhi, which began as three colleges with two faculty members and 750 students, is now one of the India's largest institutions for higher education. Currently, it has 90 colleges, 16

faculty members, 82 departments and 20 centers. In 2017, it was ranked first in India by the Centre for World University Ranking and eight by the National Institutional Ranking Framework. QS BRICS has also ranked the University of Delhi among top 10 Indian public educational institution/universities. The h-index for the University of Delhi in 2017 touched 157, one of the highest among the Indian universities. Till 2015 the University maintained its first rank in using the e-ShodhSindhu resources. After 2015, the consortium stopped assigning ranks to the institutes.

During 2016–17, 666,176 students were enrolled in the University of Delhi, out of which 197,510, were in the undergraduate program, 25,745 in the under postgraduate (including M/Phil/Ph. D) and 6147 in the certificate/diploma/PG diploma. Around 413,339 students enrolled themselves in distance learning mode and 23,435 in Non-collegiate Women Education Board.[37]

Methodology

Resource-wise usage statistics have been collected from e-ShodhSindhu's portal, namely infistats: usage statistics portal, for e-resources after login of University of Delhi for the period 2010–2017 in the COUNTER JR1 report.[38] The usage of *Science Direct* for 2013 was not available through the portal and was thus collected from Elsevier directly. The usage statistics have been analyzed through different parameters, such as average download per year, percentage of journals accounting for 80% of downloads, number of titles contributed by each resource into the 80% downloads, identifying the core titles across the resources covering 80% of downloads and identification of less/unused titles and core titles with the help of tables and graphs.

Literature survey

e-ShodhSindhu, formerly known as UGC-INFONET, has been providing e-resources, related to various disciplines, to higher learning academic institutes in order to meet their information needs. Some studies have been carried out using different parameters to analyze the usage of electronic resources being provided by the consortium. Singson and Hangsing conducted a study in 2015 on the implications of 80/20 rule in electronic journals usage of UGC-INFONET. It confirmed, if not precisely, the applicability of 80/20 rule in a large consortium.[39] Chauhan and Mahajan, while surveying the use of UGC-INFONET resources by the social scientists, tried to identify the problems they were facing in accessing them. They also highlighted the efforts made by INFLIBNET to spread awareness about these resources.[40] In another study, the authors concluded that UGC-INFONET is the most popular library consortium with respect to users and the availability of resources. They reported that while cost has been constantly increasing, the cost of per download has been decreasing every year.[41] Bhatt, in his survey, concluded that more than 90% respondents in the field of History and Political Science in the University of Delhi are aware of the resources provided by the UGC-INFONET because of information literacy programs organized by the library professionals of University of Delhi. The study also revealed that e-journals and online databases are the most common e-resources consulted by the History department (68%) and the Political Science department (67%) for teaching purposes.[42] Singson and Leeladhran, in their study, reported that the usages of UGC-INFONET resources at Pondicherry University seem to be dominated by sciences and the most sought after databases were American Chemical Society and JSTOR. This might be because science fraternity is more aware and are avid users of electronic resources.[43] Another study was conducted to quantify the usage of electronic resources by the Aligarh Muslim University. In the study, it was concluded that about 73% of respondents use the electronic resources provided by the UGC-INFONET.[44] A study of researchers and faculty members of National Institute of Technology, Tiruchirappalli, reported that 67% of researchers and 33% of faculty members regularly use the electronic resources for various purposes.[45]

No study that has been done before is as comprehensive and empirical as the present study. The usage for a span of 8 years has been taken into consideration and efforts have been made to identify the core collection for the University of Delhi.

Data analysis and conclusion

- **Average Downloads**: The following Table 1 and Graph 1 provide an insight into the total number of titles and total downloads for them.

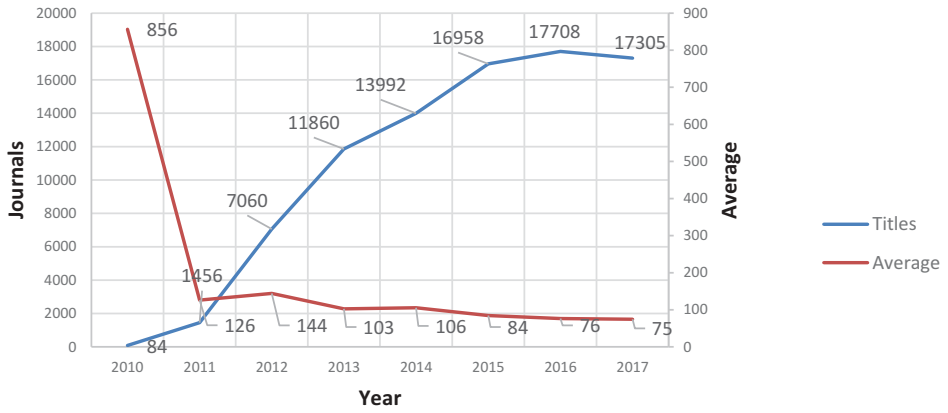
Table 1 and Graph 1 show that as the e-shodhSindhu consortium has continued to include new journals every year in its collection and the downloads have also increased every year at the University of Delhi. However, there are specific trends which show that the average downloads have been decreasing every year. In 2010, the average downloads were 856 per journal, from 84 journals, which drastically decreased to 126, though the number of titles increased to 1456 in 2011. There has been a decreasing trend in average downloads since 2010 as they have gone down from 856 in 2010 to 75 in 2017.

*Inference: More number of titles does not mean better average of downloads, as the average downloads per journal decreased although the collection has increased during the period of study.*Percentage of journals that cover 80 percent of downloads (resources wise)

Velfredo Pareto, an economist and sociologist, reported that in Italy 80% of wealth had been occupied by 20% of the people and remaining 20% wealth by the rest 80% of the people. Richard W Trueswell tried to apply Pareto's theory to libraries and found that approximately 80% of the usage of books and serials in libraries was accounted for by about 20% of the collection.[46] To find whether the same theory is applicable in electronic journals, the number of titles in each database

Table 1. Average downloads.

Year	Titles	Downloads	Average
2010	84	71,929	856
2011	1,456	183,522	126
2012	7,060	1,016,851	144
2013	11,860	1,217,921	103
2014	13,992	1,477,174	106
2015	16,958	1,430,560	84
2016	17,708	1,352,903	76
2017	17,305	1,290,721	75



Graph 1. Average download per journal.

was identified that account for 80% of the total downloads from the same. Table 2 and Graph 2 show the percentage of journals from each database which cover the 80% of downloads year wise.

An analysis of Table 2 and Graph 2 reveals that less than 20% of titles in almost half of the databases (ACM, AIP, CUP, IOP, JSTOR, Nature, OUP, Science Direct, Springer Link, and Wiley) account for 80% downloads from the respective databases. Around 20–30% Titles from four databases (Emerald, Project Muse, RSC, and T & F) account for 80% of their downloads. In case of Project Euclid, in the initial years (2011 and 2012) a higher percentage of titles accounted for 80% downloads but later the percentage of titles in 80% downloads decreased to 22% in 2017. In case of remaining databases (ACS, APS, Annual Review, Portland Press, Project Euclid, and SIAM), the percentage of titles accounting for 80% downloads was found to be more than 30% of the total number of their titles.

Inference: In case of bundled deal the Pareto's theory does not seem to apply as more than half the databases contain more than 80 percent titles which are not relevant or useful for University of Delhi users. It should also be noted that more number of titles does not mean a higher percentage of usage of all the titles. The databases with a higher percentage of journals constituting 80 percent of downloads also have comparatively lesser number of titles in comparison to the other databases.

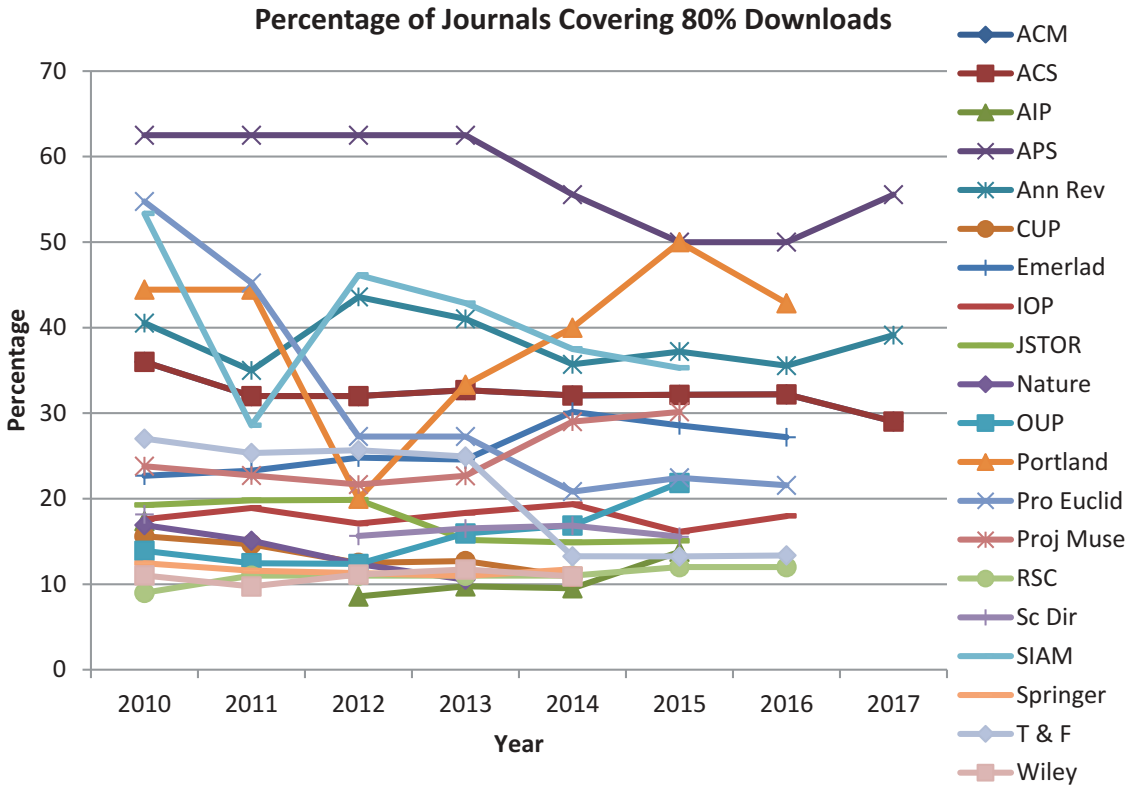
Titles covering 80% of downloads

A ranking list of all titles as per downloads from them, irrespective of their resources, has been prepared to know titles that cover 80% downloads. Table 3 and Graph 3 provide information about the number of titles of various resources which account for 80% downloads.

An analysis of Table 3 and Graph 3 shows that during the initial 2 years (2010 and 2011) of the study, the American Chemical Society contributed the highest number of titles to account for 80% of downloads. However, after the appearance of JSTOR and Science Direct in subscription list from 2013, the picture has changed as the majority of titles in the 80% downloads have been contributed by these two resources. In 2012, out of 884 total titles which account for 80% downloads, 400 and 325 belong to Science Direct and JSTOR, respectively, amounting to 82% of 884 titles. This trend

Table 2. Percentage of journals covering 80% of downloads.

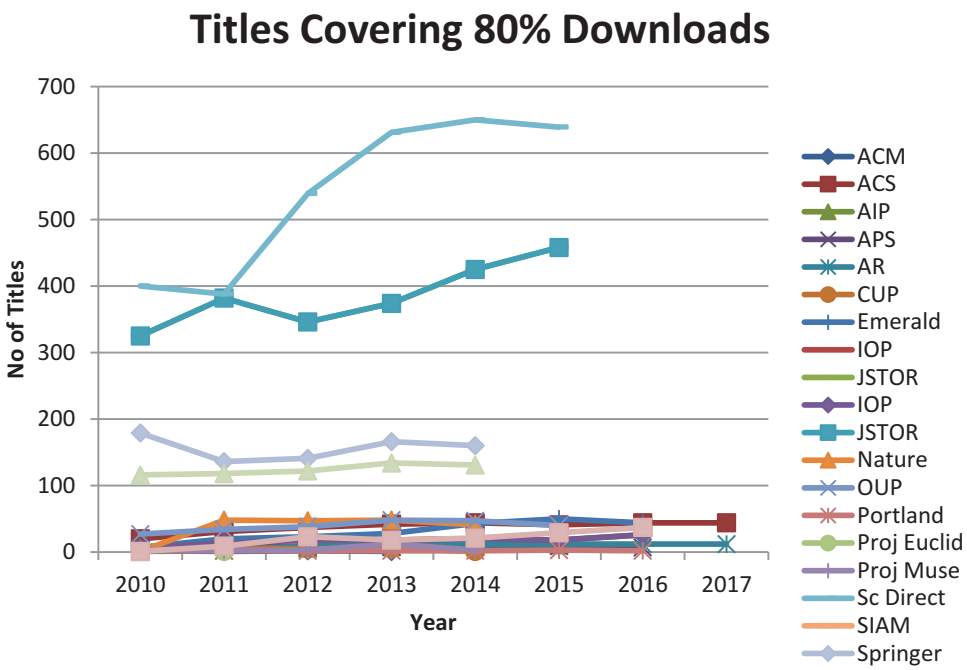
	Year							
Resources	2010	2011	2012	2013	2014	2015	2016	2017
ACM Digital Library				10	5	7	8	
American Chemical Society	36	32	32	33	32	32	32	29
American Physical Society	63	63	63	63	56	50	50	56
American Institute of Physics			17		9	10	10	14
Annual Reviews	41	35	44	41	36	37	36	39
Cambridge University Press			16	15	13	13	11	
Emerald Publishing		23	23	25	25	30	29	27
Institute of Physics		18	19	17	18	19	16	18
JSTOR			19	20	20	15	15	15
Nature					17	15	12	11
Oxford University Press			14	12	12	16	17	22
Portland Press		44	44	20	33	40	50	43
Project Euclid		55	45	27	27	21	22	22
Project Muse			24	23	22	23	29	30
Royal Society of Chemistry		21	23	22	22	20	22	22
Science Direct			18	23	16	17	17	16
SIAM			53	29	46	43	38	35
Springer Link				12	12	11	11	12
Taylor and Francis		27	25	26	25	13	13	13
Wiley				11	10	11	12	11



Graph 2. Percentage of journals covering 80% downloads.

Table 3. Titles covering 80% downloads.

	Year							
Resources	2010	2011	2012	2013	2014	2015	2016	2017
ACM Digital Library				2	1	2	1	
American Chemical Society	20	31	37	42	44	41	44	44
American Institute of Physics			6		6	8	8	11
American Physical Society		7	8	8	8	9	9	7
Annual Reviews	1	2	10	17	11	12	12	12
Cambridge University Press			3	6	5	5	1	
Emerald Publishing		7	20	23	28	43	50	44
Institute of Physics		6	13	18	18	20	18	26
JSTOR			325	382	346	374	425	458
Nature				1	48	47	48	41
Oxford University Press			27	34	38	48	47	40
Portland Press		2	2	2	2	2	3	2
Project Euclid								1
Project Muse				4	3	3	12	3
Royal Society of Chemistry		11	24	26	25	27	28	31
Science Direct			400	388	539	631	650	639
SIAM								1
Springer Link				179	136	141	166	160
Taylor and Francis		1	9	23	18	21	29	37
Wiley				116	118	122	134	131
Titles Covering 80 percent Downloads	**21**	**67**	**884**	**1271**	**1394**	**1556**	**1685**	**1688**
Total Titles of All Resources	**92**	**1456**	**7060**	**11860**	**13992**	**16958**	**17,708**	**17305**



Graph 3. Titles covering 80% downloads.

continued until 2017. It is to be noted that resources like Project Euclid and SIAM had a negligible contribution in 20% of titles accounting for 80% downloads. It is also noticeable that as the number of titles grew, the percentage of titles covering 80% downloads decreased, and as the total number of titles increased from 92 in 2010 to 17,305 in 2017, but the percentage of titles covering 80% downloads decreased from 22% in 2010 to 9.75% in 2017.

Inference: After analysis, it can be concluded that if the resources have large number of titles, then a higher number will contribute to 20 percent of the titles accounting for 80 percent of downloads. Another inference from the analysis is that as the total number of titles grows, the percentage of titles covering 80 percent downloads decreases. This again reiterates that Pareto's theory has lost its relevance in case of bundled subscription because in bundled subscription, titles which are not useful to the institute but are included by the publishers create the marketing for fewer use titles.

Core titles

An analysis has also been done to know the core titles among the total titles put together from all the resources. The percentage of core titles has been worked out by identifying the titles that meet the information needs of the users. Table 4 and Graph 4 provide an insight into the percentage of titles needed to meet a certain percentage of demand.

After an analysis of Table 4 and Graph 4, it has been revealed for the library needs, less than or equal to 10% of total titles put together from the resources meet the information needs of 80% of the users, as they account for 80% of total downloads. Even if the library wishes to satisfy 90% demands in most of the cases, it needs to subscribe less than 20% of total titles subscribed in all bundles. If the library further stresses itself and raises the level to satisfy 95% demands of the users, it needs to subscribe less than 30% to total titles. In other words, the libraries have been forced to subscribe more than 70% titles at present which account for only 5% of the total downloads.

Inference: In case of bundled subscription the Pareto's theory fails. In this study, it has been shown that 90 percent demands can be met by subscribing even less than 20 percent of the titles or 10 percent titles satisfy 80 percent of the information needs. Therefore, the libraries should put serious thought while considering a bundled subscription.

Table 4. Core titles.

	Percent of Titles Cover … percent Downloads		
Year	80	90	95
2010	25	39	51
2011	5	10	18
2012	13	22	32
2013	10	19	30
2014	10	18	27
2015	9	17	25
2016	10	17	26
2017	10	18	27

Less/unused titles

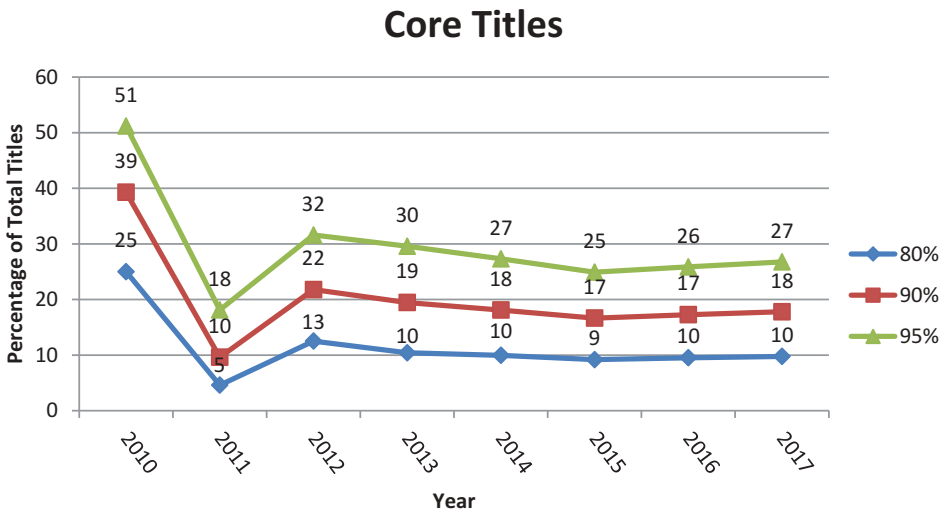
In the study, an attempt has been made to identify the less-used titles. The criterion for identifying less/unused tiles is that the title should be downloaded 12 or lesser number of times in a year (one download per month). In this analysis, it has also been worked out how much the library will gain if these titles were to be discontinued and how much library suffers loss regarding downloads from these titles.

The analysis of Table 5 and Graph 5 shows that as the number of title grows, the less/unused titles also grow. Except for 2010 and 2012, almost 60% titles were such from which either 0 downloads or less than 12 downloads per year were made. Regarding the loss of downloads, libraries lose less than 2% of total downloads.

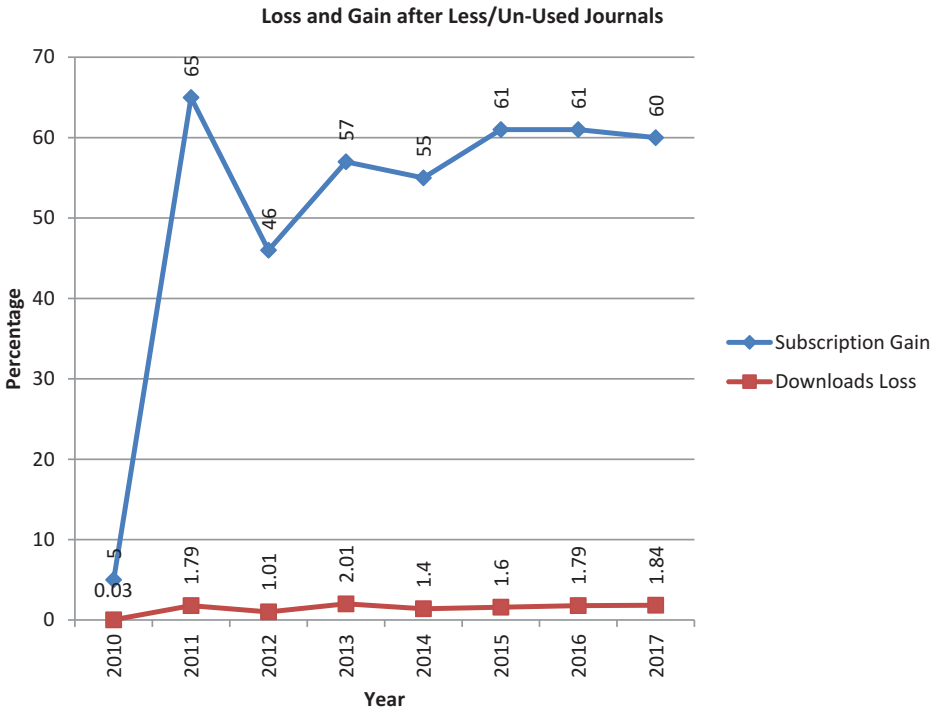
Inference: More number of titles does not mean more use, rather it shows the reverse trend, with more less/unused titles. Hence, before renewing these resources, the library can ask the vendor to replace these non-relevant titles with relevant titles. To meet the demands with these non-relevant titles, the library can use resource sharing techniques with other libraries or use like pay per view module.

Conclusion

More number of titles does not mean a better number of average downloads, as evident in the present study. The study concludes that the average downloads per journal decrease as the numbers



Graph 4. Core titles.



Graph 5. Loss and gain after less/un-used journals.

Table 5. Less or unused titles.

			0 or Less Than 12 Downloads			
Year	Total Downloads	Total Titles	No. of Titles	Downloads	Percent of Total Downloads	Percent of Total Titles
2010	71,929	84	5	24	0.03	5.95
2011	183,522	1,456	941	3,280	1.79	64.63
2012	1,016,851	7,060	3251	10,240	1.01	46.05
2013	1,217,921	11,860	6310	17,388	1.43	53.20
2014	1,477,174	13,992	7740	20,665	1.40	55.32
2015	1,430,560	16,958	10,334	22,768	1.59	60.94
2016	1,352,903	17,708	10,832	24,199	1.79	61.17
2017	1,290,721	17,305	10,455	23,802	1.84	60.42

of titles in a bundled subscription grow. The Pareto's 80/20 theory has been used to ascertain whether 80% usage correlates with 20% titles for each bundled subscription. The results, in most cases, have not substantiated. As it has been found that more than half of the databases contain even more than 80% titles which can be deemed as irrelevant or less-used titles for the users in/of the University of Delhi. However, the databases containing fewer titles seem to confirm with Pareto's theory as compared to databases having a larger number of titles.

The present study also concludes, after combining all the titles of all bundled subscriptions, 90% demands can be met by less than 20% of the titles. The study also reveals that only 10% of titles are sufficient to meet 80% demand of the users. Analysis of the usage also identified 60% of the titles as less-used or unused.

The libraries can use Pareto's simple statistical technique to identify the core titles responsible for 80% of the downloads. This technique can also be used to identify not only the databases having a higher number of less-used/unused titles but also the root cause(s) of lesser use of these titles. To enhance the usage of less-used or unused databases, the libraries can undertake various measures like database-specific awareness and literacy program.

To arrive at an informed decision to retain or cancel a journal for libraries, usage statistics is and will remain an important parameter. However, critics feel that statistics are mere figures, and sometimes they do not provide the real picture. Hence, other factors like feedback from the library user, price, impact factor and reputation of the publisher should also be considered for deciding retention or cancellation of a journal. Galbraith stated, "retention decisions should be based on the following factors: use patterns over the past five years, population served, availability of other titles covering the same topic, impact factor, coverage in indexes owned by the library, availability in other local libraries, available document option and faculty input."[47] In his report, Best observed that "Usage data is and should not be the main driver for decisions on selection and retention. Selection and retention decisions must be made with a proper understanding of the users being served, the users' needs, and the overall strength of the resources of the entire collection."[48]

The publishers/aggregators may also be informed about these types of studies, who may, in turn, re-think whether the concept of 'one size fits for all' is feasible in the future or an alternative module, like tailor-made bundles for each library, is needed.

Notes

1 A. S. Edlin and D. L. Rubinfeld, "Exclusion or Efficient Pricing? The "Big Deal" Bundling of Academic Journals," *Antitrust Law Journal* 72 (2004): 119–57.
2 American Library Association, "Acquistion of e-Journals," *Library Technology Report*, March–April 2003, 28–43.
3 D. Haank, "Is Electronic Publishing Being Used in the Best Interests of Science? The Publisher's View" (Proceedings of the Second ICSU-UNESCO International Conference on electronic publishing in science, Paris, 2001).
4 R. V. Noorden, "The True Cost of Science Publishing: Cheap Open-Access Journals Raise Question about the Value Publishers Add for their Money," *Nature* 495 (March 28, 2013): 426–29.
5 D. Butler, "The Dark Side of Publishing: The Explosion in Open-Access Publishing has Fuelled the Rise of Questionable Operators," *Nature* 495 (March 28, 2013): 433–35.
6 A. Odlyzko, "Open Access, Library and Publisher Competition, and the Evolution of General Commerce," October 18, 2013, [Online], http://www.dtc.umn.edu/~odlyzko/doc/libpubcomp.pdf (accessed August 18, 2018).
7 D.-S. Jeon and D. Menicucci, "Building Electronic Journals and Competition among Publisher," *Journal of European Economic Association* 4, no. 5 (September 2006): 1038–83.
8 H. R. Varian, "Pricing Electronic Journals," June 1996, [Online], http://people.ischool.berkeley.edu/~hal/Papers/dlib.html (accessed August 18, 2018).
9 Noorden, "The True Cost of Science Publishing.
10 S. Kambayashi, "The Cheap, Convenient Cloud: As Cloud-Computing Prices Keep Falling, the Whole IT Business will Change," April 18, 2015, [Online], https://www.economist.com/business/2015/04/18/the-cheap-convenient-cloud (accessed August 20, 2018).
11 M. Rosoff, "Why is Tech Getting Cheaper?" October 16, 2015, [Online], https://www.weforum.org/agenda/2015/10/why-is-tech-getting-cheaper/ (accessed August 18, 2018).
12 Telecom Regulatory Authority of India, "Yearly Performance Indicators of Indian Telecom Sector," May 4, 2018, [Online], https://www.trai.gov.in/sites/default/files/YPIRReport04052018.pdf (accessed August 19, 2018).
13 T. C. Bergstrom, "Free Labor for Costly Journals," 2001.
14 K. Frazier, "What's the Big Deal?" *The Serial Librarian* 48, no. 1–2 (Septmber 2005): 49–59.
15 K. Frazier, "The Librarians' Dilemma: Contemplating the Costs of the "Big Deal"," *D-Lib Magazine* 7, no. 3, March 2001 [Online], www.dlib.org/dlib/march10/frazier/03frazier.html (accessed August 19, 2018).
16 "Big Deals' and Publisher-Library Competition," February 14, 2013, [Online], https://piracy.americanassembly.org/publisher-library-competition/ (accessed August 19, 2018).
17 Edlin and Rubinfeld, "Exclusion or Efficient Pricing?".
18 L. Hardesty and C. Mack, "Searching for the Holy Grail: A Core Collection for Undergraduate Libraries," *Journal of Academic Librarianship* 19 (January 1994): 362–71.
19 S. A. Intner, "The Meaning of Core Collections," *Technicalities* 13 (July 1993): 4.
20 S. Keenan and C. Johnston, *Concise Disctionary of Library and Information Science* (London: Bowker Saur, 2000), 61.
21 J. M. Reitz, *Dictionary for Library and Information Science* (London: Unlimited, 2004), 183.
22 T. E. Nisonger, "Journals in the Core Collection: Definitions, Identification, and Applications," *The Serials Librarian* 51, no. 3–4 (2007): 51–73.

23 P. Shepherd, "The Feasibility of Developing and Implementing Journal Usage Factors: A Research Project Sponsored by UKSG," *Serials: The Journal for the Serials Community* 22, no. 2 (July 2007): 117–23.
24 T. Peters, "What's the Use? The Value of e-Resource Usage Statistics," *New Library World* 103, no. 1172–1173 (2002): 39–47.
25 B. Tracy, "The 80 20 Rule Explained," [Online], https://www.briantracy.com/blog/personal-success/how-to-use-the-80-20-rule-pareto-principle/ (accessed August 20, 2018).
26 J. Schopfel and Leduc, Claire, "Big Deal and Long Tail: e-Journal Usage and Subscriptions," *Library Review* 61, no. 7 (2012): 497–510.
27 R. L. Trueswell, "Some Behavioral Patterns of Library Users: The 80/20 Rule," *Wilson Library Bulletin* 43 (January 1969): 458–61.
28 Nisonger, "Journals in the Core Collection".
29 M. Singson and P. Hangsing, "Implication of 80/20 Rule in Electronic Journal Usage of UGC-Infonet Consortia," *The Journal of Academic Librarianship* 41, no. 2 (2015): 207–19.
30 J. M. Juran, "Universals in Management Planning And Controlling," *Management Review* 43 (November 1954): 748–61.
31 Q. L. Burrell, "The 80/20 Rule: Library Lore of Statistical Law?" *Journal of Documentation* 41, no. 1 (March 1985): 24–39.
32 W. Crawford, "Exceptional Institutions: Libraries and Pareto Principle," *American Libraries* 32, no. 6 (June-July 2001): 72–74.
33 M. Chiorazzi, "Books, Bytes, Bricks and Bodies: Thinking about Collection Use in Academic Law Libraries," *Legal Reference Services Quarterly* 21, no. 2–3 (2002): 1–28.
34 I. R. Yourng, "The Use of a General Periodicals Bibliographic Database Transaction Log as a Serials Collection Management Tool," *Serial Review* 18 (Winter 1992): 49–60.
35 S. Black and A. Sisson, "Bradford's Distribution, the 80/20 Rule, and Patterns of Full Text Database Use," *Against the Grain* 15 (December–January 2003-2004): 20–24.
36 "e-ShodhSindhu: A Consortium for HE e-Resources," 2018, [Online], http://ess.inflibnet.ac.in/ (accessed August 18, 2018).
37 "Annual Reports," 2017, [Online], http://www.du.ac.in/du/uploads/05022018_Annual_1.pdf (accessed August 18, 2018).
38 "INFLIBNET Centre," 2018, [Online], https://www.inflibnet.ac.in/infistat/ (accessed August 18, 2018).
39 Singson and Hangsing, "Implication of 80/20 Rule".
40 S. K. Chauhan and P. Mahajan, "Use of UGC-Infonet E-resources by Social Science Academics in Indian Universities: An Evaluation Study," *The Journal of Academic Librarianship* 40, no. 3–4 (May 2014): 359–66.
41 S. K. Chauhan and P. Mahajan, "Library Consortia in India with Special Reference to UGC-Infonet Digital Library Consortium," *The International Information and Library Review* 45, no. 3–4 (December 2013): 127–38.
42 R. Bhatt, "Use of UGC-Infonet Digital Library Consortium Resources by Research Scholars and Faculty Members of the University of Delhi in History and Political Science: A Study," *Library Management* 31, no. 4–5 (2010): 319–43.
43 M. Singson and M. Leeladharan, "Use of Scholarly Resources among Research Scholars in Pondicherry University," Februray 2010, [Online], https://www.academia.edu/643207/Use_of_Scholarly_Resources_among_Research_Scholars_in_Pondicherry_University (accessed August 20, 2018).
44 M. M. Raza and A. K. Upadhyay, "Usage of E-Journals by Researchers in Aligarh Muslim University: A Study," *The International Information & Library Review* 38, no. 3 (September 2006): 170–79.
45 P. Mounissamy and M. Kaliammal, "Promoting Effective Use of Electronic Resources using Library Web Sites by IITs and NITs: A Comparative Study," *IASLIC Bulletin* 51, no. 4 (2006): 213–20.
46 Trueswell, "Some Behavioral Patterns of Library Users".
47 B. Galbraith, "Journal Retention Decisions Incorporating Use-Statistics as a Measure of Value," *Collection Management* 27, no. 1 (2002): 79–90.
48 R. Best, "Lies, Damn Lies, and Usage Statistics: What's a Librarian to do?" in *Usage Statistics of e-Serials* (New York: Haworth Information Press, 2007), 199–214.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Narender Kumar http://orcid.org/0000-0002-5988-300X

2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)

A Novel Framework for Neural Architecture Search in the Hill Climbing Domain

Mudit Verma, Pradyumna Sinha, Karan Goyal, Apoorva Verma and Seba Susan
Department of Information Technology
Delhi Technological University
Bawana Road, Shahbad Daulatpur, Rohini, Delhi, 110042
Email: {mudit.verma2014, pradyumna014, karangoyal2k96, 19apoorva.v}@gmail.com, sebasusan@dce.ac.in

***Abstract*—Neural networks have now long been used for solving complex problems of image domain, yet designing the same needs manual expertise. Furthermore, techniques for automatically generating a suitable deep learning architecture for a given dataset have frequently made use of reinforcement learning and evolutionary methods which take extensive computational resources and time. We propose a new framework for neural architecture search based on a hill-climbing procedure using morphism operators that makes use of a novel gradient update scheme. The update is based on the aging of neural network layers and results in the reduction in the overall training time. This technique can search in a broader search space which subsequently yields competitive results. We achieve a 4.96% error rate on the CIFAR-10 dataset in 19.4 hours of a single GPU training.**

***Index Terms*—Neural Architecture Search, Hill-Climbing, Optimization**

I. Introduction

Deep Neural Networks have demonstrated impressive performance on many machine-learning tasks such as computer vision, speech recognition and natural language processing [1] [2] [3] [4]. The high performance of deep neural networks is strengthened by the cost of large-scale engineering and checks to find the best architecture for a given problem. High-level design decisions such as depth, units per layer, and layer connectivity are not always obvious and use hit and trial approach which is exhausting and time-consuming. This led to a growing interest in the automatic designing of neural networks. Since the architecture search space is discrete, traditional optimization algorithms such as gradient descent [20] are not applicable, hence approaches based on reinforcement learning algorithms [13] and evolutionary algorithms [21] are used for automated designing of neural networks. But these approaches are either very costly (requires hundreds and thousands of GPU days) or yield non-competitive results. In this work, we aim to reduce the computational cost by using a simple hill-climbing approach for the optimization of search space and our variant of network morphism [18] for making the network deeper and complex. We have also added gradient stopping, linear morphism and several implementational improvements for reducing the training time. Our work uses a single GPU, and we validate our method on the CIFAR-10 dataset. It takes our framework 19.4 hours to reach 4.96% error. When we use this same architecture to train and validate on MNIST dataset, we obtain an error percentage of 0.28. We have compared two ways of weight initialization, with and without Gradient Stopping.

In the following sections, we will discuss the previous attempts at Neural Architecture Search (NAS). Further, we present our NASGraph, explain its various nodes and morphism operations, and propose our methodology. We describe our implementation of morphism operations, gradient stopping technique, linear morphism and finally give a flow chart for our algorithm. In the end, we present our experiments and showcase our results in comparison with handcrafted architectures on CIFAR-10 as well as other Neural Architecture Search schemes, and also gain insights about using our generated deep neural network on different datasets. We then conclude our work and present a few pointers for possible future research opportunities based on our work.

II. Related Work

Research on neural architecture search started in the 1980s. Initially, evolutionary algorithm based approaches were proposed which promises to find out both the architecture and weights of the neural network in [5]. However, they have been unable to match the performance of the hand-crafted network. More recent work [6] [7] [8] [9], use an evolutionary algorithm only to find out the structure and stochastic gradient descent to find out the parameters of the network. [9] and [10] use evolutionary algorithms to generate a complex architecture from a smaller neural-network iteratively, but [9] used a large number of resources (250 GPUs and ten days) whereas [10] is limited to relatively smaller networks due to handling a population of networks.

RL-based approach. [12] proposed Neural Architecture Search (NAS) with the help of reinforce algorithms [13] to learn a network architecture. It generates a sequence of actions representing the architecture of a CNN. NASNet [14] uses proximal policy-optimization (PPO) in-place of reinforce algorithms [13] which further improves neural architecture search. It searches the architecture of a small unit which they referred to as block. It repeatedly concatenated itself to form a complete model. Other works in the field use approach like policy gradient in [14], Q-learning in [15] for finding the model architecture. Unfortunately, training a reinforcement

978-1-7281-1488-0/19/$31.00 ©2019 IEEE
DOI 10.1109/AIKE.2019.00009

IEEE computer society

learning agent for searching architecture is extremely expensive: both [16] and [17] required over 10.000 fully trained networks, requiring hundreds to thousands of GPU days.

Network morphism/ transformation : Using the concept of transfer learning, network morphism was introduced and first published in 2015 by Chen [18]. To make the network more complex, the author described function preserving operation such as 'deeper' or 'wider' with the goal of speeding up training and exploring network architectures. [18] further proposed additional operations, e.g., for handling non-idempotent activation functions or altering the kernel size and introduced the term network morphism. [19] presents a neural architecture search by hill climbing (NASH) and used morphism operations for making the network more complex. Out of all the methods, [19] resembles the most to our techniques of neural architecture search by hill climbing (NASH) with network morphism. We extend their work by adding gradient stopping, linear morphism, several implementational improvements like the selection of nodes upon which morphism is to be applied, the addition of more number of operators for growth of performance using less resources, time (1 GPU and less than one training day) and at the same time yielding higher accuracy.

III. NAS Graph

NASGraph represents an instance of neural-architecture in our state space of neural network architectures for the dataset that is given to the hill climbing algorithm. A specific set of attributes define a NASGraph. It is implemented as a Graph data structure where each node can either be a convolutional block (ConvLayer-BatchNorm-ReLU), a max pool node, a Merge node or a linear node. It has an adjacency list, the matrix that defines the connections between the nodes. It also stores a list of node identifiers and a mapping between the identifiers and the actual instances. There is another list that maintains the node identifiers of its parents and children. Morphism operations have been defined on the NASGraph, and during the hill climbing process, they are called to "morph" the parent network into a child network. At an abstract level, the graph is initialized with image (from the input dataset) dimensions. A Convolution node is added with random parameters initially, and linear layers are stacked at the end for classification. Finally, the NASGraph is transformed into a network, and its parameters are provided to the SGDR [23] optimizer.

Topological Ordering Condition: According to this condition, it is a must for the NASGraph to remain acyclic at all times. Therefore any connections made from node x to y in the NASGraph must be such that x lies before y in the topologically sorted order of the NASGraph.

A. NASGraph Nodes

NASGraph Nodes are the fundamental building blocks of all possible neural architectures through our framework. These nodes have dual behavior; they are the usual graph nodes and therefore participate in lists maintained for graph representation. However, they are also network layers; this limits the position of edges in the graph. Also, to implement the framework, algorithms such as topological sort, cycle detection, etc. have been performed on these which are used later to maintain the topological ordering condition.

There are a total of 4 NASGraph Nodes as follows.

1) **Convolution :** This node is composed of a 2-D convolution layer followed by a Batch Normalization [24] layer and ReLU activation.
2) **Maxpool :** This node is composed of a max pool layer.
3) **Merge :** This node may be an Add layer, which takes up multiple node outputs as it's input, where all inputs must be of the same dimension and output their matrix addition. Additionally, this node may be a Merge layer, where different inputs are stacked one below another along channel dimension and fed as one input to the next layer.
4) **Linear :** This is a fully connected neural network layer.

The first node of the NASGraph is supposed to be randomly created, and after this, the iterative hill climbing begins. The first node is a convolution node with kernel sampled uniformly from set $\{3, 5\}$, padding, uniformly from set $\{True, False\}$, $stride = 1$, and channels from set $\{8, 16, 32\}$. Beyond this point, the only way to modify the properties of an existing NASGraph node is via morphism operations.

B. NASGraph Morph Operations

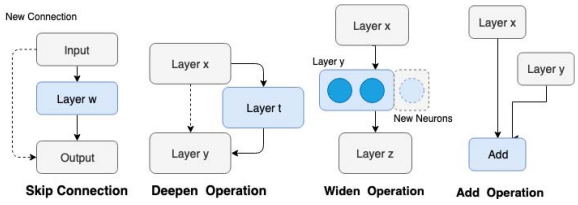


Fig. 1. Some of the NASGraph Operations

Our morphism technique differs from that of [19] in that we do consider layers without padding. Convolution layers without padding have proven to be useful as seen in popular architectures such as VGG16 [25]. Therefore, this factor must be included in the search space.

We allow our NASGraph to perform one of the following six operations for the generation of the child node in the hill climbing algorithm :

1) Skip Operation: Skip Operation allows the NASGraph to have skip connections in the neural network. Skip connections help traverse information in neural networks. As the NASGraph is found to generate very deep neural architectures, it is possible for the gradient information to be lost since it passes through so many layers, thus skip connections enable us to pass feature information to lower layers to help classify directly. As we also provide the provision for max-pooling layers, max pooling operation itself causes some amount of spatial information to be lost, skip connections help in this as well.

The idea for performing Skip is first to find all the node pairs in which the first nodes output is the same as the second nodes input. Such a situation may cause the second node to have multiple parents each, providing this node with same dimension inputs. Therefore, we use an Add node, described below, to add all such parents and then supply the combined result to this node in Fig. 1. We can see an example of a skip connection being applied between the input and output layer shown by the dashed line connection.

2) Deepen: Deepen, as the name suggests, is a layer to deepen the existing network. This layer, in our case, is a Convolutional Block composed of a Convolutional Neural Layer, a Batch Normalization Layer, followed by a ReLU activation function. The layer may also be a Linear Layer, that is a fully connected layer.

To perform the deepen operation between nodes x and y, let us assume, layer x has a neurons, and layer y has b neurons. Then, the weight matrix is of shape (a,b). The new layer t must have input dimension of (b,b) so that the composition of layers x and t gives the final output (a,b), same as before.

Selection of Nodes for Deepen is given in section IV-B. Figure 1 shows a new node t being added in between nodes x and y. The addition of a layer adds another level of abstraction that cannot be as simply contained within shallow layer architecture of the same number of parameters if they can be contained at all.

3) Widen: Widen operation increases the channels of a convolution node by a factor of 2 or 4. Widening requires changing the parent node weight matrix as well as the subsequent child node weight matrix to adjust for the widening changes. To simplify the process, we, therefore, create a set of those node pairs, in their topological ordering, (x, y) where has only one child y, and y has only one parent x. Not using this condition would require us to perform changes in the architecture recursively, i.e., if the parent weight changes, then all its children weight matrices have to change recursively.

Widen Operation has also been used for fully connected layer widen morphism which is described in detail in section IV-E.

4) Add: When a node in the graph has more than one parent, one of the options to combine the inputs is to do an element-wise sum of the incoming tensor. However, this is only possible if the incoming tensors are of the same shape.

5) Merge: The other option to combine inputs from more than one parent is to concatenate the channels of the image tensors. For this method to be used, it is required that the height and width of the incoming image tensors be the same.

6) Maxpool: The max pool property is essential to building CNN and is used extensively in object detection tasks like in [25]. [19] does not allow for dynamic addition of new max-pool nodes in the architecture, i.e., max pool layer must be originally present in the seed architecture for it to exist in the final architecture. We ease this restriction and allow for dynamic addition of max pool nodes in our NASGraph. Similar to the addition of a Convolution Node, we first find all node pairs, in topological order, which can accommodate a max pool layer in between them and maintain all previous connections. Then, we uniformly sample one such node pair and add a max pool.

IV. Proposed Work

A. Morphism

Morphism has been used in [19] before for Neural Architecture Search; however, in approaches like NASH [19], the use of Morphism restricted the network state space, i.e., all Convolution Layers are required to be padded. They treat each convolution layer as a function, which they split into two functions using the morphism operation, thereby the final input and output dimensions remain the same as before the morphism was applied. We, however, allow layers to be added without padding, which dynamically changes the output dimension of the layer.

Morphism functioning for different operations has been explained in Section III-B. The morphism has been applied in two ways, first with default Pytorch version 0.4.1 [28] initialization of layer weights (Xavier Initialization for Convolutional Layers), and the second, 0/1 initialization as in NASH. In 0/1 initialization when we use widen operation, the added neurons are set to value 0, and for deepening process, the added layer has all values initialized to 1. We show our results with both types of initialization methods and find that they do not affect the total time to get a trained network, and default initialization performs better than 0/1 initialization only marginally.

B. Selection of Nodes For Morphism

NASH [19] states the morphism technique and the operations; however, the selection of the nodes upon which the morphism operation would take place is described as "random." This random selection works for the NASH procedure because they begin with a seed architecture and each added layer has padding enabled, thereby keeping the output always the same. In our case, however, compatibility of nodes between which an operation is to be performed is an issue. For example, in the Skip operation, we need to pick two nodes A and B such that, $output(A) = input(B)$. Moreover, any such additional connection must meet the topological order condition of NASGraph as in section III.

We always keep a topological ordering of the NASGraph nodes, and for any request of connecting two nodes via Skip, compatible node pairs ordered topologically are found, and one such pair is selected uniformly randomly. This method of 'randomly' picking nodes for the morphism operation can be termed as fairly random for all practical purposes because it is essentially selecting nodes for the operation randomly from the set of all compatible nodes.

For adding a Convolution Layer with a given kernel shape and other parameters, say c, we find a set of all topologically ordered node pairs (x,y) such that $c(x.output).output = y.input$. Now we can randomly pick any element of this set for our morphism operation.

C. Random Selections

Many parameters of the NASGraph nodes are sampled randomly and we describe all such parameters below.

We need to select various properties of the NASGraph nodes, for example, the kernel, stride, etc. of the Convolution Node. As in [19], we choose these properties randomly from a predefined set of values. And finally, since we allow some of the layers to exist without padding, we mark padding True or False, with a probability of 0.5 each.

For deepen operation which is the addition of a Convolution Node, we select the kernel from the set $\{3, 5\}$. The number of channels is chosen to be equal to the number of channels of the closest preceding convolution.

For widen operation which is the expansion of a Convolution Node, the number of channels increases by a widening factor, which is sampled randomly from the set $\{2, 4\}$.

The parameters for Maxpool operation are also required to be chosen uniformly. Kernel for the same is selected uniformly from a set $\{2, 3\}$.

The morphism operation applied is also chosen randomly from the set of all possible morphisms. We describe morphism operations in the section III-B. Finally, if a certain operation cannot be performed at some point of time, then we temporarily remove that operation from the set and again sample uniformly from this new operations set.

D. Gradient Stopping

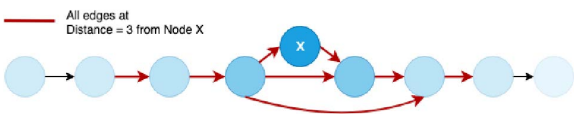


Fig. 2. Illustrating the nodes involved in gradient stopping for distance = 3

Repeated application of NASGraph morphism operations for a large number of times can result in very deep neural networks. The problem with the conventional approach to handling these dynamically growing neural architectures is overtraining of specific 'old' layers. Additionally, we try to analyze the addition of a layer via a water drop ripple analogy. A drop of water falling in a pool of water creates water ripples around the drop location, thus leaving far away locations unaffected. Hence, the addition of a convolution layer should effectively cause major changes in the weights of near-by layers only, and training of the far-away layers may be stopped. However, defining a set of nodes as 'far away' is difficult, so we use a Normal Distribution Curve to solve this.

By iterative generation of complete Neural Networks, we found the average depth of the graph is around 36. We arbitrarily choose to let the training of nodes continue which lie within $1/3$ of the average length of nodes which comes to be approximately 12. Thus after adding a convolution node, we find all the nodes connected up till distance = 6 (which would give at least 12 nodes as the total from either side). An example of distance = 3 is presented in Fig. 2 where all

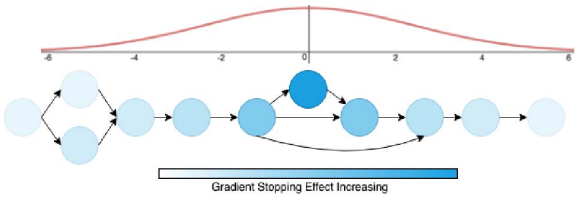


Fig. 3. Gradient Stopping, using the normal distribution to update new learning rates for neural network nodes.

the edges at a distance less than or equal to three have been marked red. Now, all the nodes connected via these marked red edges can be denoted via their distances.

To determine the learning rate of a node, a value from the Normal distribution equation $f(distance) = \frac{6}{\sqrt{2\pi}\cdot 2.4} e^{\frac{-(x)^2}{2s^2}}$ corresponding to the distance assigned to the node is obtained, say $alpha_n$. The distribution is set so that this $alpha_n$ lies between the range $(0, 1]$ and tends to zero as the distance value reaches 6. Beyond $distance = 6$ we set the value of $alpha_n = 0$. Now, we obtain the learning rate value $alpha$ according to the SGDR [23] update rule and get the new $alpha = alpha * alpha_n$. Note that in the next iteration, $alpha$ is again obtained from SGDR rule, as it would be without Gradient Stopping, and the changes are performed on that value. All the nodes beyond distance 6 do not update their gradient until the next non-layer-addition morphism operation.

Fig. 3 gives a graphical representation of gradient stopping, where the darkest circle represents the newly added node, for which the learning rate is unchanged, as $alpha_n = 1$. As we move away from this node, the circle shade fades off representing reduction in their learning rates, and beyond a specific number of nodes (6 in this case), the training completely stops. Gradient Stopping is a considerably good example of layer-wise training of a neural network with dynamic changes to the architecture at the run time.

E. Linear Morphisms

As mentioned previously, our framework supports convolution nodes to be added without padding, as well as the addition of max pool layers. The output dimensions of these layers are strictly less than the input dimensions; therefore, the final output shape of tensor exiting the 'NasGraph' (graph of Convolution, MaxPool, Add, Merge Nodes) bubble reduces. For example, if the previous output dimension is (28,28), now with the addition of Convolution Node at the very end with kernel size (5,5) without padding, the output dimension becomes (24,24). Hence the size of the flattened layer changes as in Figure 4. Due to this change, we must make the immediate following Fully Connected layer (FC layer) have a reduced number of neurons.

In our framework, we use two Fully Connected Layers for classification. The first one, Variable FC Layer, is connected to the flattened NASGraph matrix and is morphed to reduce in size dynamically. The number of neurons in this Variable FC Layer is equal to the number of neurons in the Flattened

Layer. We can do this because the image size of the CIFAR-10 dataset is (32,32). Thus, the maximum number of neurons in the Flatten layer can be $32 * 32 = 1024$. So now Variable FC also has 1024 neurons. Therefore the size of the weight matrix becomes $1024 * 1024$ which is around 1 Million parameters, feasible to handle by today's computing standards. The next and the final layer has ten neurons corresponding to the ten classes in the CIFAR-10 dataset. Therefore, any changes in the NASGraph will be held until the Variable FC layer and we can have a constant size final Fully Connected layer for Classification.

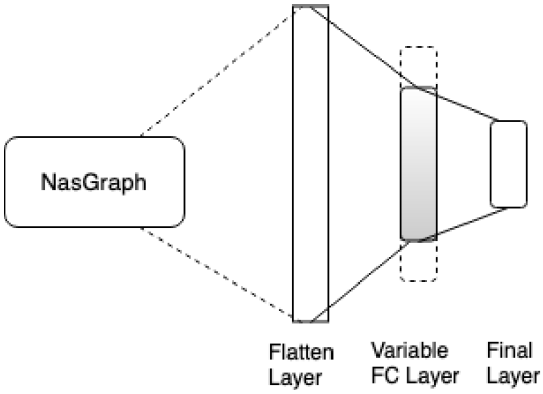


Fig. 4. Linear Widening Morphism

F. Process Flow

The flowchart in Fig. 5 shows the steps of our algorithm in a nutshell. The graphical architecture of the model is referred to as a NASGraph. First, a random architecture is generated to be the starting point of the state space search, called the NASGraph seed. The Hill climbing operation is carried for n_{steps}. In each step, n_{neigh} neighbors are generated for the current best model. To create a neighbor, n_{NM} morphism operations are applied as described in section IV-A, IV-B and IV-C. The resultant neighbors imbibe the properties of the parent model along with having new attributes. These neighbors are then trained for $epoch_{neigh}$. Finally, the current best model is updated by comparing with the results of the neighbors on the validation dataset. After n_{steps}, the best model is trained for $epoch_{final}$ and the trained model is reported as the output of the algorithm.

V. EXPERIMENTAL RESULTS

A. Dataset Overview

For neural architecture search, we use CIFAR-10 dataset [29]. Convolution Neural Networks have shown to be very effective in working on the image problem domains. Thus, we showcase our results also on image datasets. To test how well does the architecture generated by our method on CIFAR dataset performs on images of a different domain, we use the handwritten digits MNIST dataset. Baseline results on Neural Architecture Search are usually on CIFAR / MNIST datasets. Therefore, a comparison of our method along with other methods is suitable for the same.

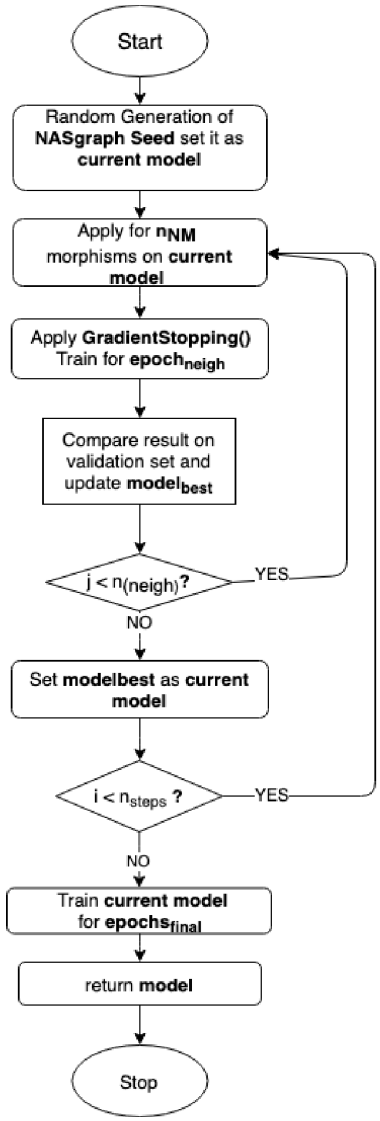


Fig. 5. Flowchart of our Iterative Hill Climbing technique for Neural Architecture Search

B. Discussion on Results

TABLE I
RESULTS COMPARISON WITH DIFFERENT WEIGHT INITIALIZATION TECHNIQUES.

Learning Rate	Xavier Initialization	0/1 Initialization
	Time (hrs) / Error (%)	Time (hrs) / Error (%)
SGDR + Gradient Stopping	19.4 / 4.96	19.6 / 5.2
SGDR	23.4 / 5.8	23.3 / 5.8

TABLE II
PARAMETERS FOR NASH AND OUR METHOD.

Variable	NASH-2	Ours
n_{steps}	8	10
n_{NM}	5	5
n_{neigh}	8	8
$epoch_{neigh}$	17	16
$epoch_{final}$	100	64
λ_{start}	0.05	0.1
λ_{end}	0.0	0.0

TABLE III
COMPARISON OF NEURAL ARCHITECTURE SEARCH ON CIFAR-10 ON DIFFERENT MODELS.

Model	Resource	Error(%)
Shake-Shake [11]	2 days, 2 GPU	2.9
RL-NAS [17]	? days, 800 GPUs	3.65
NASH-1	0.5 days, 1 GPU	5.7
NASH-2	1 day, 1 GPU	5.2
Ours	19.4 hours, 1 GPU	4.96

For similarities in our methods, we choose the NASH [19] as our baseline.

Parameters for NASH used to compare the results are as provided in [19]. The set values are given in Table II. NASH requires us to feed a simple seed architecture which they mentioned as : Conv-Max pool-Conv-Max pool-Conv-FC-Softmax where Conv = Conv + BatchNorm + ReLU. This architecture has already been trained for 20 epochs. Using these parameters, the best possible error rate achieved by NASH is 5.2% in one day of GPU training.

Table II also mentions the value of the parameters we have used, which is the same for all our architectures presented in this paper. We explore the same number of neighbors at any depth of the hill climbing as that of NASH. However, the exploration depth is higher in our case which in turn enables us to apply a higher number of morphisms on our network, and reach more depth, without requiring much time.

We first explain the parameters used as follows.

1) n_{steps} is the depth of the hill climbing tree.
2) n_{NM} is the number of times the morphism operation is applied on a hill climbing node.
3) n_{neigh} is the number of children generated in the hill climbing tree.
4) $epoch_{neigh}$ is the number of epochs for which each child is trained for.
5) $epoch_{final}$ is the number of epochs for which the final $model_{best}$ is trained.
6) λ_{start} and λ_{end} are the parameters required for SGDR.

Since the objective of this paper is to propose a newer faster baseline, we must focus our scrutiny on three areas primarily, amount of resources used, the time required for the complete training process, and the final error percentage. In all the three fronts we either match the NASH or are better. Table III gives a comparison of our framework (with Linear Morphisms, Updated Operations & Gradient Stopping with SGDR) with the two NASH models proposed in [19], the Shake-Shake [11] model, and a reinforcement learning based Neural Architecture Search Scheme [17]. NASH-1 has $n_{steps} = 5$, while NASH-2 parameters are given in Table II. Shake-Shake [11] is a manually crafted state-of-the-art model in CIFAR-10 classification which uses 2 GPUs and gives 2.9% error in 2 days. However, on comparing with a Neural Architecture Search scheme like that of [17], that although gives a low error of 3.65%, but uses a mammoth 800 GPU system, is quite unrealistic for practical purposes. Our baseline, NASH scheme, uses 1 GPU and one day of training time to give a 5.2% error, against our training time of 19.4 hours to produce better results at 4.96% error. The reported error for our model is the average of three full runs on CIFAR-10 classification which yielded 5.1%, 4.9%, 4.9% errors respectively averaging 19.4 hours of training time. We also note that our method searches for comparatively more of the neural architecture space in less time and produces better accuracy.

C. On a Different Dataset

We also performed classification using the architectures produced with CIFAR-10, on the MNIST dataset. It was done to gain insights regarding how the architecture performs on other image datasets. Low error percentages by architectures generated from one type of dataset, upon another dataset, would mean either similarity in the dataset, or the robustness of the architecture itself. Since MNIST and CIFAR-10 are known to be very different datasets, where MNIST consists of handwritten digits, and CIFAR-10 consists of images of various real-life objects, we chose MNIST for this purpose. Upon using the same architecture, for which the results on CIFAR-10 are mentioned in Tables I and III, we obtain 0.28% error on MNIST dataset. The current state-of-the-art on MNIST is Rmdl: Random multimodel deep learning for classification [22] which boasts an error percentage of 0.18. However, our architecture would still stand in the top 10[1] error percentages reported until June 2018.

VI. CONCLUSION & FUTURE WORK

We introduced a Neural Architecture Search in Hill Climbing domain using morphism operations that use our novel Gradient Update scheme. Our method is not just fast and uncomplicated, but it also outperforms many existing automated architecture search methods with lesser GPU days and resources. Our experiments performed on the CIFAR-10 and MNIST datasets and yielded satisfactory results.

Our approach can quite comfortably deliver as a basis for the development of more sophisticated techniques that yield further improvements in the existing performance. Certain operations are being executed randomly in our method, such as the selection of a morphism operation or selection of a node for performing the morphism steps. One can make these selections intelligently by using reinforcement learning or genetic algorithms. Essential criteria for selection of nodes

[1]https://paperswithcode.com/sota/image-classification-mnist-handwritten-d

would be to pick convolution blocks from those regions of the network which have been having significant changes in their weights in the recent iterations. Again, there are different ways to initialize the weights in the morphism operations. We experimented with two of many possibilities, and expect further research in weight initialization would yield better results.

Apart from improvements in the existing work, we find research in Learning Curve Extrapolation for dynamic neural networks in its initial stages, but would immensely speed up the overall hill climbing iterative algorithm. However, it is currently there for static [26], [27] neural network graphs.

REFERENCES

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[2] Karpathy, A., & Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3128-3137).

[3] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Dieleman, S. (2016). Mastering the game of Go with deep neural networks and tree search. nature, 529(7587), 484.

[4] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).4

[5] Schaffer, J. D., Whitley, D., & Eshelman, L. J. (1992, June). Combinations of genetic algorithms and neural networks: A survey of the state of the art. In [Proceedings] COGANN-92: International Workshop on Combinations of Genetic Algorithms and Neural Networks (pp. 1-37). IEEE.

[6] Liu, H., Simonyan, K., Vinyals, O., Fernando, C., & Kavukcuoglu, K. (2017). Hierarchical representations for efficient architecture search. arXiv preprint arXiv:1711.00436.

[7] Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., ... & Hodjat, B. (2019). Evolving deep neural networks. In Artificial Intelligence in the Age of Neural Networks and Brain Computing (pp. 293-312). Academic Press.

[8] Real, E., Aggarwal, A., Huang, Y., & Le, Q. V. (2018). Regularized evolution for image classifier architecture search. arXiv preprint arXiv:1802.01548.

[9] Real, E., Moore, S., Selle, A., Saxena, S., Suematsu, Y. L., Tan, J., ... & Kurakin, A. (2017, August). Large-scale evolution of image classifiers. In Proceedings of the 34th International Conference on Machine Learning-Volume 70 (pp. 2902-2911). JMLR. org.

[10] Suganuma, M., Shirakawa, S., & Nagao, T. (2017, July). A genetic programming approach to designing convolutional neural network architectures. In Proceedings of the Genetic and Evolutionary Computation Conference (pp. 497-504). ACM.

[11] Gastaldi, X. (2017). Shake-shake regularization. arXiv preprint arXiv:1705.07485.

[12] Zoph, B., & Le, Q. V. (2016). Neural architecture search with reinforcement learning. arXiv preprint arXiv:1611.01578.

[13] Williams, R.J.: Simple statistical gradient-following algorithms for connectionist reinforcement learning. Mach. Learn. 8(3-4) (May 1992) 229256

[14] Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8697-8710).

[15] Cai, H., Chen, T., Zhang, W., Yu, Y., & Wang, J. (2018, April). Efficient architecture search by network transformation. In Thirty-Second AAAI Conference on Artificial Intelligence.

[16] Baker, B., Gupta, O., Naik, N., & Raskar, R. (2016). Designing neural network architectures using reinforcement learning. arXiv preprint arXiv:1611.02167.

[17] Bello, I., Zoph, B., Vasudevan, V., & Le, Q. V. (2017, August). Neural optimizer search with reinforcement learning. In Proceedings of the 34th International Conference on Machine Learning-Volume 70 (pp. 459-468). JMLR. org.

[18] Wei, T., Wang, C., Rui, Y., & Chen, C. W. (2016, June). Network morphism. In International Conference on Machine Learning (pp. 564-572).

[19] Elsken, T., Metzen, J. H., & Hutter, F. (2017). Simple and efficient architecture search for convolutional neural networks. arXiv preprint arXiv:1711.04528.

[20] Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

[21] Miller, G. F., Todd, P. M., & Hegde, S. U. (1989, June). Designing Neural Networks using Genetic Algorithms. In ICGA (Vol. 89, pp. 379-384).

[22] Kowsari, K., Heidarysafa, M., Brown, D. E., Meimandi, K. J., & Barnes, L. E. (2018, April). Rmdl: Random multimodel deep learning for classification. In Proceedings of the 2nd International Conference on Information System and Data Mining (pp. 19-28). ACM.

[23] Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983.

[24] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

[25] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[26] Klein, A., Falkner, S., Springenberg, J. T., & Hutter, F. (2016). Learning curve prediction with Bayesian neural networks.

[27] Domhan, T., Springenberg, J. T., & Hutter, F. (2015, June). Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves. In Twenty-Fourth International Joint Conference on Artificial Intelligence.

[28] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... & Lerer, A. (2017). Automatic differentiation in pytorch.

[29] Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images (Vol. 1, No. 4, p. 7). Technical report, University of Toronto.

APPENDIX
SAMPLE OUTPUT OF NASGRAPH

Fig. 6 shows a sample graph produced by the NASGraph using the parameters mentioned in Table II. The generated network can be seen to have convolution blocks as represented by red bubbles, and Merge layers (which can be either the Merge or the Add function) represented by the blue add bubbles. The graph clearly holds the Topological Ordering condition as mentioned a must for the NASGraph. In addition, it is seen to have Skip connections for example node conv0 is connected to add27 providing input to conv1, along with conv13, conv19, conv7, conv17, conv11. The NASGraph can generate quite complicated neural networks and provide good accuracy rates on CIFAR-10.

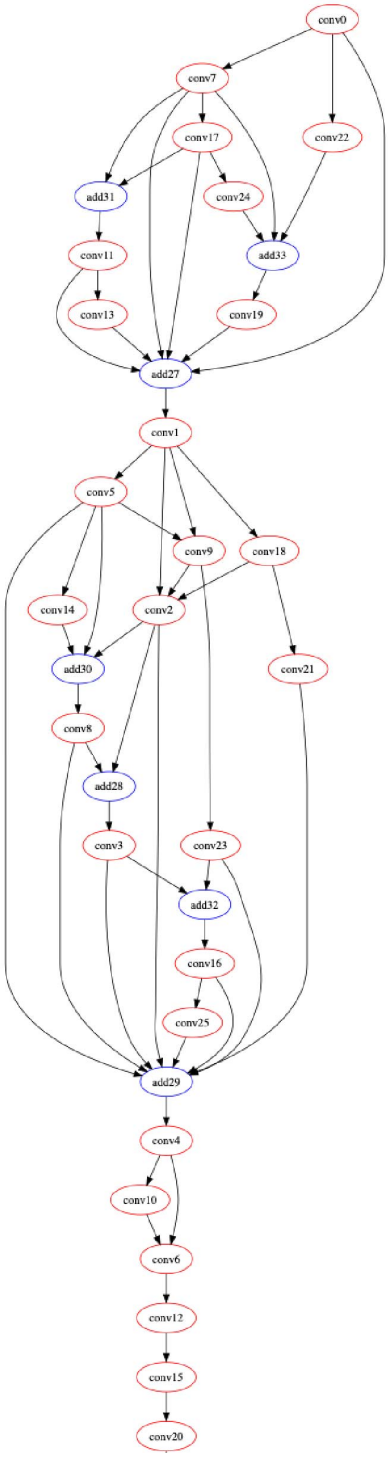


Fig. 6. Sample output of NASGraph on CIFAR-10 dataset

Iran Journal of Computer Science
https://doi.org/10.1007/s42044-019-00045-y

ORIGINAL ARTICLE

Anomaly detection in group activities based on fuzzy lattices using Schrödinger equation

Rajiv Kapoor[1] · **Om Mishra**[1] · **M. M. Tripathi**[2]

Received: 21 April 2019 / Accepted: 24 July 2019
© Springer Nature Switzerland AG 2019

Abstract
This paper proposes a new method to detect abnormality in a video captured for surveillance applications in real-time and also overcome the problem of the curse of dimensionality. To extract features related to any change in the video, nonlinear Gaussian fuzzy lattice functions have been applied on each frame of the video which results in the formation of fuzzy lattices. These fuzzy lattices have been expressed in the form of Schrödinger equation to find the kinetic energy involved corresponding to any change in the video. A number of the fuzzy lattice has been used as a dimension of the feature. It reduces the dimensionality significantly as compared to other state-of-the-art methods. Finally, the kinetic energy parameter is classified into normal and abnormal activities with the help of SVM-based classifier.

Keywords Anomaly detection · Fuzzy lattices · Schrödinger equation · SVM

1 Introduction

Recently, we have seen that analysis of crowd activities is very important specifically in the part of South Asia, due to a large population of countries such as India, China, and Pakistan. It is very tough to monitor crowd activities of busy places. In India, so many incidents of stampede occurred in different parts of the country. Terrorism is also a major problem in this part of the world. To monitor these kinds of activities, there is an increasing demand for automatic visual surveillance systems. In the case of video behavior profiling, video captured by CCTV camera is analyzed to detect any abnormality in the scene. Abnormality [1–3] means any suspicious activity or any activity which possibility of happening is very low. In other words, activities having higher probability are termed as normal and with lower probability are abnormal. Research is going on in computer vision to develop effective video behavior profiling algorithms in such a way that abnormal and normal activities in the video can be detected effectively.

✉ Rajiv Kapoor
rajivom2009.dtu@gmail.com

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Bawana Road, Delhi, India

[2] Department of Electrical Engineering, Delhi Technological University, Bawana Road, Delhi, India

Various approaches have been proposed to detect an anomaly in the video. They can be broadly categorized according to the type of scene representation. One very popular category is based on trajectory modeling [4–12]. The trajectory-based translation invariant feature is used to understand the videos [4]. Authors have used motion pattern and size pattern of the object as the features to detect the anomaly in the scene [6, 7]. The trajectory of the agent is used to find out his behavior pattern in Ref. [5]. In [10], the likelihood of the trajectory of an object is calculated to understand the scene. A Bayesian approach is used for an anomaly in the scene in Ref. [11] and in Ref. [12] statistical distance measures are used to find an event detection in two videos. In these methods, moving targets are tracked and classified into different categories which make use of image features and image-sequence-based tracking results. The main drawback of these methods is their sensitivity to occlusion and tracking errors. To overcome this occlusion problem, multiple views of the scene are captured from different cameras. Trajectory-based features are used as global anomaly features for the scene capturing form different cameras [8, 9]. All the features from the different cameras are fused. But the problem becomes profound in case of very crowded scenes because frequent occlusion that occurs at the public place makes tracking difficult.

To avoid these problems, various authors have proposed alternative motion representations that avoid tracking

Published online: 09 August 2019

[13–19]. Spatial and temporal anomalies in the scene are modeled by the appearance and dynamics of the scene in Ref. [14] and [17]. In [13], anomalous events in the scene are detected by creating a number of local monitors. Local statistical data from different monitors are fused to make a decision about an anomaly in the scene. The space–time Markov random field (MRF) is used to detect an unusual event in the video in Ref. [15]. The features thus extracted are directly used to create models of activities. These techniques suffer a curse of dimensionality which makes them noncompatible for real-time implementation. Most of the researchers used principal component analysis (PCA) where data are taken at linear subspace. There are also some nonlinear dimensionality reduction techniques such as LLE [20], Laplacian eigenmaps [21], and Isomap [22] which have been applied to reduce the dimensionality. The proposed method also uses a nonlinear dimensionality reduction technique. In our previous work, fuzzy lattice-based dimension-reduction method is used to recognize the dynamics of the face [16].

One very popular method for anomaly detection is a convolutional neural network-based approach. The approaches based on deep learning technique [23–25] are fully automated. Sabokrou and Fayyaz [23] use the 3D convolution to extract the features. This technique requires large dataset and the efficiency of these methods is based on the design of networks. These networks are made up of a large number of data and parameters that make its structure very complex. So it takes more time to train a model. Moreover, less availability of large datasets makes this method less generalized for anomaly detection. To overcome these problems, [26, 27] used new CNN architecture where two streams are used, one for spatial and other for the temporal network. The main disadvantage is that these descriptors for anomaly detection are based on supervised learning where we train the system for different anomalies but if a new anomaly is introduced, then the system will not detect it as an anomaly. Authors of Sabokrou et al. [28] overcame this problem using an unsupervised fully convolutional neural network. Although it is showing good result, the proposed method also gives comparable result to it.

The proposed method overcame both the problems in the following manner:

1. It drastically reduces the feature dimensions. So this speeds up the proposed method which will make the proposed system more suitable for real-time application.
2. The proposed method also supports unsupervised learning. Suppose we trained our system for a mall where people are walking normally and suddenly they start running. Walking normally is a normal class and running suddenly is an abnormal class. But if a new anomaly is introduced, for example, people start fighting, in that case, the proposed method will still detect anomaly as its kinetic energy will increase to a large extent as compared to a normal class.

So our contributions to the proposed method can be summarized as follows:

1. Fuzzy lattices are used for the first time to find out the anomaly in the crowd activities.
2. The concept of Schrödinger wave equation to find out the change in kinetic energy of the fuzzy lattices has been used for the very first time.
3. The proposed method is also a very good dimension-reduction technique which makes it suitable for real-time application.
4. The proposed method outperforms other existing state of the art in terms of accuracy and run time especially for crowd-related activities.

Three major complexities arise while detecting anomalies in videos. First is the diversity of data which creates action pattern variation in the same class. Second is the variation in the environment due to change in illumination and varying background and third is high-dimensionality problem. Due to these problems, there exists no linear relationship among neighborhood pixels. To deal with first and second problems we used Gaussian membership function with variation in variances and means. As varying Gaussian functions cover both high- as well as low-frequency components, they can represent complete information of activity performed in a scene. Unlike PCA, LLE, and ICE, Gaussian membership function covers two-dimensional information spread both horizontally and vertically and thus it is more accurate.

The proposed method uses a novel technique for feature extraction where Gaussian fuzzy lattice functions [29, 30] are used to form fuzzy lattices as features. To extract information regarding any significant change, kinetic energies of lattices are evaluated using the solution of the famous Schrödinger wave equation. To avoid high-dimensionality problem, the proposed method uses only a small number (n) of fuzzy lattices which have maximum energies because they contain maximum information of change in a scene. So having n number of fuzzy lattices to represent a scene reduces dimensionality drastically and thus makes proposed method real-time applicable. These features are classified with the help of SVM classifier [31–33] into normal and abnormal class. In our previous work, the proposed method has also been applied to recognize the facial expression by extracting nonlinear features of different expression [16]. The rest of the paper is structured as follows: Sect. 2 describes the proposed methodology, Sect. 3 discusses experimental results on challenging standard datasets and its performance evaluation, and real-time applicability and Sect. 4 discusses conclusion.

2 Proposed method

In the case of the anomalous behavior of the crowd in a video, it is desired to monitor and analyze the sequence of activities happening in the video. To do so, a large number of frames of video are analyzed. While monitoring such large data, a lot of redundant information is found. To reduce redundant information, the paper has proposed dimension reduction approach based on fuzzy lattices using the Schrödinger wave equation. Block diagram of the proposed method is mentioned in Fig. 1.

The proposed method can be represented with the following points:

(a) Representation of features in terms of lattices [3] with the help of optimized Gaussian fuzzy lattice function.
(b) Expressing/measuring lattice features in terms of kinetic energy by the Schrödinger wave equation.
(c) Classification of signal and decision-making process has been carried out to distinguish between normal and abnormal events by support vector machine.

2.1 Formation of fuzzy lattices

In every scene, there is nonlinearity in the image as the intensity value of pixels varies nonlinearly from one pixel to another. Therefore, this nonlinear relation between neighborhood pixels has been represented by performing nonlinear Gaussian fuzzy lattice function. This function has been given by the following equation and has been operated on each pixel:

$$f_{ij}(z) = e^{-\frac{(z-z_i)^j}{2\sigma_i^2}}, \tag{1}$$

where z_i and σ_i are the mean and variance of the lattice function $f_{ij}(z)$, respectively, where z is the gray-level value of the pixel at coordinate positions x and y of the image. The values of i and j are taken from 1 to 4, because beyond this value, Gaussian will become an impulse. By doing this, the total sixteen number of Gaussian curves having different centers and width will be formed as lattice functions. These lattice functions are applied to each pixel. If the order of pixel connectivity is considered as 8, then every pixel in the image except the pixels at an edge of the image has 8 neighborhood pixels corresponding to eight directions, i.e., 45°, 90°, 135°, 180°, 225°, 270°, 315°, 360° respectively. The order of connected pixel is decided by the value of lattice functions. Each Gaussian lattice function-modeled pixel has been compared with the gray value of its neighborhood pixels within some degree of lattice function. Figure 2a explains the one-to-one connectivity among the pixels. To find out the one-to-one connectivity, the first pixel is taken as test pixel and the neighboring pixels are found out. Now the connected neighboring pixel can be taken as the central pixel and the connected pixels are found out. This repetitive process yields a lattice of connected pixels and thus they form the sequence of connectivity. So the order of these features depends on the number of pixels connected in different directions. Figure 2b shows three lattices corresponding to three sequences of connected pixels. All the connected neighborhood pixels on the basis of Eq. (1) result in the formation of fuzzy lattices [16, 29].

Fuzzy lattices can be represented by equation:

$$l_t = \bigcup_{ij} e^{-\frac{(z-z_{mi})^j}{2\sigma_{zi}^2}} \quad \text{for } t = 1, 2, 3, \ldots, T, \tag{2}$$

where l_t represents lattice which is the union of all connected pixels satisfying the relation established by Eq. (1).

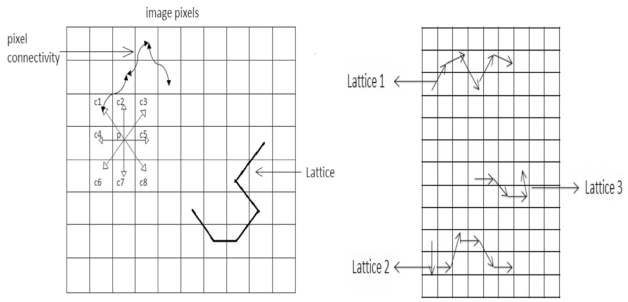


Fig. 2 **a** Connected pixels and **b** fuzzy lattices

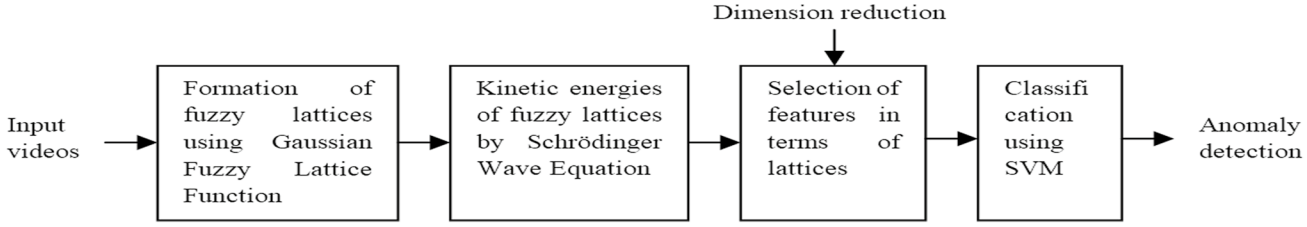


Fig. 1 Block diagram of the proposed method

In Eq. (2), T is the total number of lattices formed in an image representing a scene. Lattices behave like waves of the motion in the images and pixels are like wave particles. Therefore, these lattices are expressed in the form of Schrödinger wave equation [19, 34] to find the change in the kinetic energy (K.E.) of the lattices which corresponds to any change in the dynamic features of an image or frame. Any change that occurred in a scene due to the motion of the crowd can be represented by these fuzzy lattices. This change also produces a change in the kinetic energy of an image. This energy signal is a three-dimensional signal, where X and Y coordinates of the pixel are responsible for a change in K.E. due to change in position of X- and Y-directions respectively, while the grey-level value of that pixel is responsible for a change in K.E. in the Z-direction. To find X, Y and Z components of the K.E of these lattices, Gaussian lattice functions of X, Y and Z are given by the following three equations, respectively:

$$f_{ij}(x) = e^{-\frac{(x-x_{mi})^j}{2\sigma_{xi}^2}}, \tag{3}$$

$$f_{ij}(y) = e^{-\frac{(y-y_{mi})^j}{2\sigma_{yi}^2}}, \tag{4}$$

$$f_{ij}(z) = e^{-\frac{(z-z_{mi})^j}{2\sigma_{zi}^2}}, \tag{5}$$

where x_{mi}, y_{mi}, and z_{mi} are means of X, Y and Z, respectively, and σ_{xi}, σ_{yi} are the variances of x and y coordinates of all pixels of a lattice and σ_{zi} is the variance of the pixel value. Total Gaussian lattice function is given by the following equation:

$$F = f(x) \cdot f(y) \cdot f(z). \tag{6}$$

Suffixes i and j are ignored to reduce the complexity of the equation. Analysis of these lattices is done by the Schrödinger wave equation by finding the kinetic energy of lattices. These lattices can be treated as waves in different directions and pixels of the lattice are like free moving particles, so their potential energy is zero. Change in kinetic energy of lattices due to change in the position of pixels is found out by solving the Schrödinger wave equation (second-order differential equation). From the solution of Schrödinger wave equation [19, 34], kinetic energies of lattices formed in X-, Y- and Z-directions are given by the following equations:

$$k \cdot e_x = \frac{1}{f(x)} \frac{\partial^2 f(x)}{\partial x^2}, \tag{7}$$

$$k \cdot e_y = \frac{1}{f(y)} \frac{\partial^2 f(y)}{\partial y^2}, \tag{8}$$

$$k \cdot e_z = \frac{1}{f(z)} \frac{\partial^2 f(z)}{\partial z^2}. \tag{9}$$

To find out the kinetic energy of the fuzzy lattices in the form of Eqs. (7), (8) and (9), we have taken the second-order derivatives of fuzzy lattice functions expressed in Eqs. (3), (4) and (5). The first-order derivative of Eq. (3) with respect to X-direction is given by Eq. (10) and the second-order derivative is given by Eq. (11):

$$\frac{\partial f_{ij}(x)}{\partial x} = \frac{-j}{2\sigma_{xi}^2}\{(x - x_{mi})^{j-1}\} e^{-\frac{(x-x_{mi})^j}{2\sigma_{xi}^2}}, \tag{10}$$

$$\frac{\partial^2 f_{ij}(x)}{\partial x^2} = \frac{-j}{2\sigma_{xi}^2}\left\{\frac{-j}{2\sigma_{xi}^2}(x - x_{mi})^{2j-2} + (j-1)(x - x_{mi})^{j-2}\right\} e^{-\frac{(x-x_{mi})^j}{2\sigma_{xi}^2}}. \tag{11}$$

Similarly, second-order derivatives of Eqs. (4) and (5) with respect to Y- and Z-directions are given by the following equations, respectively:

$$\frac{\partial^2 f_{ij}(y)}{\partial y^2} = \frac{-j}{2\sigma_{yi}^2}\left\{\frac{-j}{2\sigma_{yi}^2}(y - y_{mi})^{2j-2} + (j-1)(y - y_{mi})^{j-2}\right\} e^{-\frac{(y-y_{mi})^j}{2\sigma_{yi}^2}}, \tag{12}$$

$$\frac{\partial^2 f_{ij}(z)}{\partial z^2} = \frac{-j}{2\sigma_{zi}^2}\left\{\frac{-j}{2\sigma_{zi}^2}(z - z_{mi})^{2j-2} + (j-1)(z - z_{mi})^{j-2}\right\} e^{-\frac{(z-z_{mi})^j}{2\sigma_{zi}^2}}. \tag{13}$$

To find out the kinetic energy of the fuzzy lattices in X-, Y- and Z-directions, the second-order derivatives from Eqs. (11), (12) and (13) are substituted in Eqs. (6), (7) and (8), respectively, to get the following equations:

$$k \cdot e_x = \frac{-j}{2\sigma_{xi}^2}\left\{\frac{-j}{2\sigma_{xi}^2}(x - x_{mi})^{2j-2} + (j-1)(x - x_{mi})^{j-2}\right\}, \tag{14}$$

$$k \cdot e_y = \frac{-j}{2\sigma_{yi}^2}\left\{\frac{-j}{2\sigma_{yi}^2}(y - y_{mi})^{2j-2} + (j-1)(y - y_{mi})^{j-2}\right\}, \tag{15}$$

$$k \cdot e_z = \frac{-j}{2\sigma_{zi}^2}\left\{\frac{-j}{2\sigma_{zi}^2}(z - z_{mi})^{2j-2} + (j-1)(z - z_{mi})^{j-2}\right\}. \tag{16}$$

The kinetic energy of a pixel of fuzzy lattice in X- and Y-directions given by Eqs. (14), (15) and (16) represents its intensity value, i.e., Z-direction. If a lattice constitutes 'N' numbers of pixels, the total kinetic energy of a lattice in X-, Y- and Z-directions of a lattice is given by the following equations, respectively:

$$\text{total K.E}_x \text{ of a lattice} = \sum_{i=1}^{N} k \cdot e_{xi}, \tag{17}$$

$$\text{total K.E}_y \text{ of a lattice} = \sum_{i=1}^{N} k \cdot e_{yi}, \tag{18}$$

$$\text{total K.E}_z \text{ of a lattice} = \sum_{i=1}^{N} k \cdot e_{zi}. \tag{19}$$

Total kinetic energy of a lattice given in the following equation is calculated by adding energies in X-, Y- and Z-directions from Eqs. (17), (18) and (19) respectively:

$$\text{total K.E} = \sqrt{(\text{total K.E}_x)^2 + (\text{total K.E}_y)^2 + (\text{total K.E}_z)^2}. \tag{20}$$

The kinetic energy of a lattice calculated in Eq. (20) may vary from one frame to another frame of a video due to motion.

2.2 Dimension reduction and classification

There are numbers of fuzzy lattices formed in the image due to motion. It increases the dimensionality of the feature. To deal with the dimensionality problem, we have adopted the following approach:

We have taken the appropriate number of fuzzy lattices having maximum kinetic energies. Suppose X is a set representing all fuzzy lattices given by the following equation:

$$L_\text{d} \varepsilon\, X \quad \text{where} \quad L_\text{d} = 1,\ 2,\ 3,\ \ldots, T. \tag{21}$$

The number of fuzzy lattices selected decides the dimension of the feature. Lattices having maximum kinetic energy show the maximum change in a scene. Out of all those lattices, suppose we are selecting n number of lattices having maximum kinetic energy where n will be the number of most prominent fuzzy lattices that are responsible for motion. The selection of n prominent fuzzy lattices is done on the basis of criteria mentioned in Eqs. (22)–(26) as follows: S_1 selects the lattice having maximum kinetic energy. S_2 selects the lattice having highest kinetic energy excluding S_1 whereas S_3 selects the lattice having highest kinetic energy excluding S_1 and S_2. Similarly, we calculate for n number of lattices S_n.

$$S_1 = \max_{L_d}\{\text{Kinetic energy } (L_1,\ L_2,\ L_3, \ldots,\ L_T)\}, \tag{22}$$

$$S_2 = \max_{L_d}\{L_1,\ L_2,\ L_3, \ldots, L_T/S_1\}, \tag{23}$$

$$S_3 = \max_{L_d}\{L_1,\ L_2,\ L_3, \ldots, L_T/S_1 \cup S_2\}, \tag{24}$$

$$S_4 = \max_{L_d}\{L_1, L_2, L_3, \ldots, L_T/S_1 \cup S_2 \cup S_3\}, \tag{25}$$

$$S_n = \max_{L_d}\{L_1,\ L_2,\ L_3, \ldots,\ L_T S_1 \cup S_2 \cup S_3 \cup S_{n-1}\}, \tag{26}$$

$S_1,\ S_2,\ \ldots,\ S_n$ are n fuzzy lattices having maximum kinetic energies from the set of fuzzy lattices X. Lattices having maximum kinetic energy (n) are used as the feature vectors. These features are fed to the radial basis function-based SVM classifier [32, 33] to classify the test video into abnormal or normal class.

The algorithm mentioned in Fig. 3 explains the process flow of the proposed technique wherein step 1 extracts the sequence of frames in an input video. In step 2, pixels of a frame are modeled by Gaussian fuzzy lattices function. The connectivity of these pixels formed different fuzzy lattices in a frame in step 3. Step 4 calculates the kinetic energy of these lattices using the Schrödinger wave equation. In step 5, dimensionality reduction is used to select the number of prominent lattices on the basis of their kinetic energies which is used as a feature vector. These feature vectors are fed to SVM classifier to detect the normal and abnormal classes.

3 Experimental results

The proposed method is implemented in the MATLAB R2015a and it is tested on a system having hardware configuration processor Intel(R) Core(TM) i5 CPU @ 2.40 GHz,

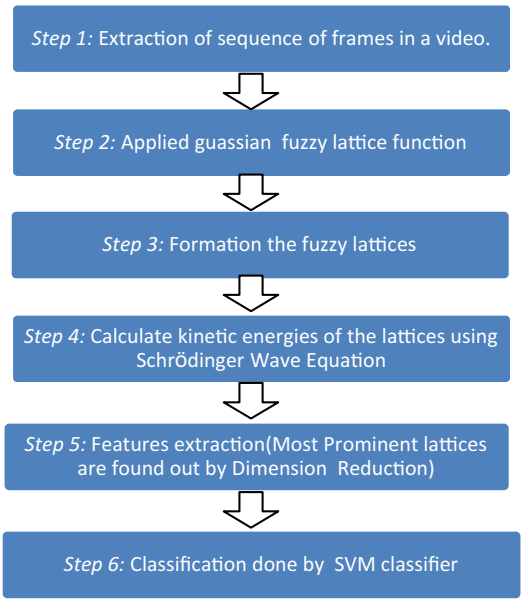
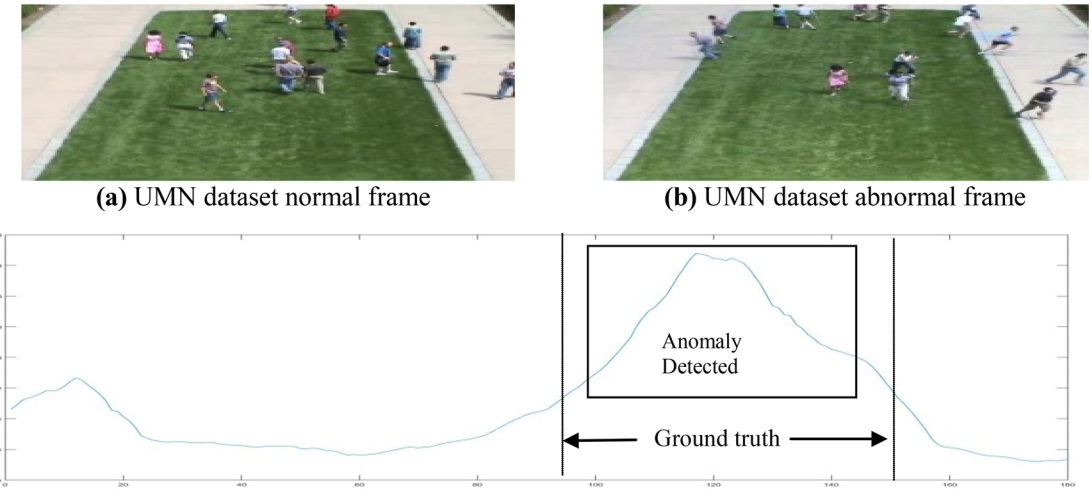


Fig. 3 Algorithm for making lattices

8 GB RAM and 64-bit operating system. For real-time analysis, all the experimental results are validated on the CUDA-enabled NVIDIA Titan Graphics Processing Unit using MATLAB 2015a. This method has been compared with the other state-of-the-art methods and applied for challenging and complex standard datasets such as UMN web dataset [35], UCSD [36] and UCF [37] for crowd activities.

The UMN dataset has three different scenarios where people are walking in different places and suddenly they started running to escape; this running activity is assumed to be an anomaly in the video. Figure 4 shows examples of normal and abnormal frames of this dataset where the first row of the figure represents normal events in the scenes and second row represents abnormal events.

UCSD dataset contains a variety of anomalies occurring in crowd activities shown in Fig. 5. It uses mounted elevated stationary camera to monitor the pedestrian walk. A number of people walking in a scene vary from less crowded to the more crowded scene. Normal event contains the walk of pedestrians and abnormal events contain people moving in random order, engaged in fights, etc. To represent anomalies, dataset uses nonpedestrian entities such as bikers, skaters, small carts, and wheelchairs. Another anomaly is also introduced by adding the

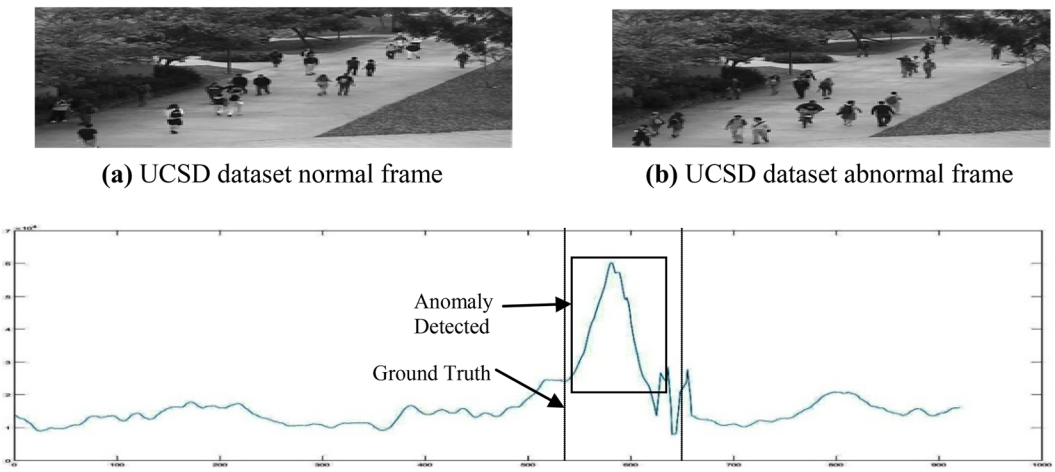


(a) UMN dataset normal frame

(b) UMN dataset abnormal frame

(c) Result on UMN dataset (X-axis represents frames and Y-axis represents the change in kinetic energy)

Fig. 4 **a** UMN dataset normal frame, **b** UMN dataset abnormal frame, and **c** result on UMN dataset (*x*-axis represents frames and *y*-axis represents the change in kinetic energy)



(a) UCSD dataset normal frame

(b) UCSD dataset abnormal frame

(c) Result on UCSD dataset (X-axis represents frames and Y-axis represents the change in kinetic energy)

Fig. 5 **a** UCSD dataset normal frame, **b** UCSD dataset abnormal frame, and **c** result on UCSD dataset (*x*-axis represents frames and *y*-axis represents the change in kinetic energy)

anomalous pattern of walking of the pedestrian. The videos of this dataset consist of two hundred frames in each video clip.

This dataset uses two scenarios.

Scenario 1 People are walking towards and away from the camera. For training purpose, 34 videos are used and 36 videos are used for testing purpose.

Scenario 2 People are walking parallel to the camera. Sixteen videos are used for training and 12 videos are used for testing purpose.

UCF web dataset is developed by the University of Florida. It is a promising dataset having real-world crowd scenario shown in Fig. 6. This dataset contains 20 videos having normal activities of the crowd which includes a pedestrian walk. It contains eight videos of abnormal activities which include panic escapes and crowd fighting.

We have trained these datasets simultaneously to make a large dataset having different types of anomalies in the crowded place.

Parameter sensitivity test Selection of the lattices is the most sensitive parameter of the proposed methodology. These lattices are selected on the basis of the maximum kinetic energies of the lattices based on the Schrödinger wave equation. Movement in the scene changes the kinetic energy of the lattices in *X*-, *Y*-, and *Z*-directions. We have used the UCSD dataset for the parameter setting. We have tested it on different numbers 3, 5, 10 and 15. It can be observed that the proposed methodology is best suitable for 5. Although greater the number of lattices greater is the accuracy, but for more than five lattices the overall accuracy varies from 2 to 3% only. The reason behind it is that lattices having less kinetic energy contribute less to detection anomaly in crowd motion. More lattices increase the complexity and dimension of the methodology. That is why we have selected five lattices having maximum kinetic energies. A similar setting is applied to the other dataset also.

3.1 Performance evaluation methodology

We used the frame-level performance evaluation methodology. For each frame in the video, lattices having maximum kinetic energy are selected. We have compared the proposed method with the other latest potential state of the art, qualitatively on frame-level evaluation. We have selected two parameters, area under curve (AUC) and equal error rate (EER), which give the relation between true-positive rate (TPR) and false-positive rate (FPR). Anomalous events are considered as "positive" and normal events are considered as "negative". TPR and FPR can be expressed by the following equations:

$$\mathrm{TPR} = \frac{(\text{True positive})}{(\text{True positive} + \text{False negative})}, \tag{27}$$

$$\mathrm{FPR} = \frac{(\text{False positive})}{(\text{False positive} + \text{True negative})}. \tag{28}$$

TPR are positive cases that are correctly classified and FPR are the negatives that are incorrectly classified as positive.

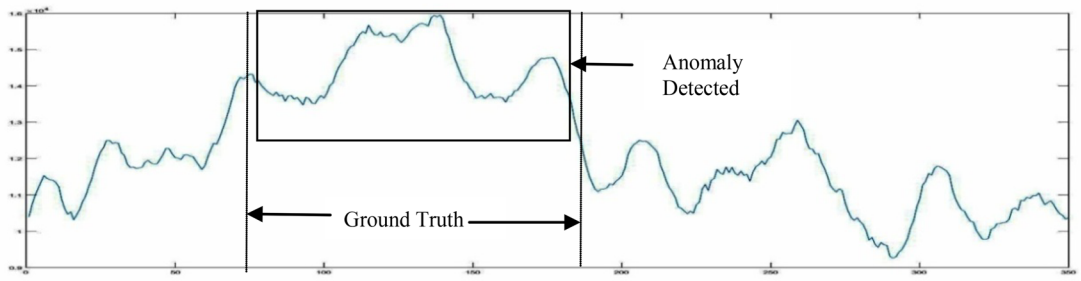
EER basically gives the misclassification rate. EER can be defined by the following equations:



(a) UCF dataset normal frame



(b) UCF dataset abnormal frame



(c) Result on UCF dataset (X-axis represents frames and Y-axis represents the change in kinetic energy)

Fig. 6 **a** UCF dataset normal frame, **b** UCF dataset abnormal frame, and **c** result on UCF dataset (*x*-axis represents frames and *y*-axis represents the change in kinetic energy)

$$\text{EER} = 1 - \text{Accuracy}, \tag{29}$$

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}, \tag{30}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative

We used these parameters for comparing the proposed method to the other state-of-the-art methods. Performance evaluation discussed above is used to compare the proposed method with other efficient state-of-the-art methods on different datasets.

3.2 Performance evaluation for UMN dataset

Figure 4a shows the frame where a group of people is walking normally and Fig. 4b shows the frame where people are running suddenly in different directions. Figure 4c shows that the ground truth for panic escape anomaly starts from frame no. 95 to frame no. 150. The proposed method detected the anomaly from frame no. 100 to frame no. 145. This means we are achieving high true positive from frame no. 100 to frame no. 145 and low false positive from frame no. 95 to 99 and from frame no. 146 to frame no. 150. The reason for this low false positive is that we are taking the running average of the kinetic energies of 20 consecutive frames. This results in nondetection of anomalies in the initial and the last five frames. If the running average is not considered, a single high spike may also show an anomaly which may increase the rate of true negatives. Consideration of running average also helps to detect anomaly accurately by showing high spike for a reasonable duration. Equal error rate (EER) and area under curve (AUC) have been used for performance evaluation. Table 1 shows the comparison of our proposed method with other state-of-the-art methods on UMN dataset. Table 2 gives the confusion matrix of the proposed method and other existing methods for a local dataset where N means normal and A abnormal.

AUC gives the area between true-positive rates and false-positive rate. Xu et al.'s [18] method shows good EER of 1.6 as compared to other methods [17, 18, 23]. AUC of almost all methods gives good result but AUC of our proposed method which is 99.8 shows the superiority over other existing methods.

Table 1 EER and AUC as anomaly detection parameters on UMN dataset

Methods	EER	AUC
Xu et al. [18]	4.8	99
Li et al. [17]	3.4	99.5
Sabokrou et al. [23]	2.5	99.6
Proposed method	1.6	99.8

Table 2 Confusion matrix of the proposed method and other existing method local dataset where N—normal and A—abnormal

Methods	N: normal; A: abnormal	N%	A%
Xu et al. [18]	N	98.1	1.9
	A	6.2	93.8
Li et al. [17]	N	98.3	1.7
	A	5.9	94.1
Sabokrou et al. [23]	N	99.4	0.6
	A	4.8	95.2
Proposed method	N	99.6	0.4
	A	2.4	97.6

3.3 Performance evaluation for UCSD dataset

Figure 5a shows the frame where a group of a pedestrian is walking normally and Fig. 5b shows the frame where a nonpedestrian (biker) enters the scene. Figure 5c shows the ground truth for nonpedestrian anomaly starts from frame no. 545 to frame no. 643. The proposed method detected the true positive from frame no. 550 to frame no. 630 and false positive from frame no. 545 to 549 and from frame no. 631 to frame no. 643. In UMN dataset example discussed above, the number of people is constant. Unlike UMN dataset, UCSD dataset has variation in the number of people in the scene. Some people are entering in the video whereas some are leaving, thus the curve in the resulting graph of this video is not as smooth as the curve of UMN dataset example. It has multiple changes in the spikes because kinetic energy changes every time a person enters or leaves the frame. But if the entry of a new person gives a sudden high spike in the resulting graph, it is detected as an anomaly. Table 3 shows the comparison of our proposed method with other state-of-the-art methods on UCSD dataset. Most of the methods have been verified on this standard dataset. This dataset

Table 3 EER and AUC as anomaly detection parameters on UCSD dataset

Methods	EER	AUC
Yachuang et al. [25]	27.4	71.3
Zhijun et al. [38]	26.6	73.4
Mousavi et al. [39]	24.2	77.5
Xu et al. [18]	18.5	80.4
Li et al. [17]	16.4	84.7
Sabokrou et al. [23]	14.2	87.2
Proposed method	5.6	95.8

is simple, so all methods show good result. The proposed method shows better EER of 5.6 as compared to other methods. AUC of almost all methods gives good result but AUC of our proposed method, that is, 95.8 shows the superiority over other existing methods. Table 4 shows the confusion matrix of the proposed method and other existing methods of the local dataset where N means normal and A abnormal. It is clear that for the proposed method only 2.8% normal class is confused with abnormal class and only 6.2% abnormal class is confused with the normal class.

3.4 Performance evaluation for UCF dataset

Figure 6a shows people gathered and the police are warning the criminal and Fig. 6b shows people running suddenly in all directions. Figure 6c shows that the ground truth for panic escape starts from frame no. 74 to frame no. 185. The proposed method detected anomaly from frame no. 80 to frame no. 180. Figure 6c also shows that just like UCSD dataset, the resulting curve is not smooth because of the variation in the total number of people in the video. Comparison of EER of the proposed method with the other state-of-the-art methods shows the superiority of the proposed method to over existing methods. The proposed method gives 12% EER evaluated in Table 5 which is very low as compared to other existing methods. Thus, it outperforms other methods. Table 6 shows the confusion matrix of our proposed method and other existing state-of-the-art methods. For our proposed method, only 2.6% normal class is confused with the abnormal class and 4.4% abnormal class are confused with the normal class. Thus, our proposed method clearly outperforms other existing state-of-the-art methods.

3.5 Performance evaluation YouTube video and new MOT 17 dataset

To validate the proposed method on another challenging environment, we applied it on the MOT 17 [40] dataset and YouTube video [41]. YouTube video is shown in Fig. 7a. Video shows normal chores going on in the shop and suddenly a customer becoming violent with the shopkeeper. There are 2900 frames in the video. We used 1600 frames for the training and 1300 frames for testing purpose. Figure 7c shows that the ground truth for the video starts from frame no. 470 to frame no. 578. The proposed method detected true positive from frame no. 480 to frame no. 570.

MOT 17 dataset is also analyzed for the anomaly in the crowd activity. MOT 17-09-FRCNN video is used for training purpose. It contains the crowd Pedestrian Street captured from low-angle stationary camera shown in Fig. 7b. It contains a total of 525 frames having resolution 1920×1080. For testing purpose, we used MOT 17-08-FRCNN which is also Pedestrian Street captured from a stationary camera. Figure 7d shows that there is no anomaly detected in the scene. The graph has some high spike for a very short duration which is not an anomaly because it occurred due to the entry of a set of new people in the scene. As people are disappearing from the scene it

Table 4 Confusion matrix of the proposed method and other existing method local dataset where N—normal and A—abnormal

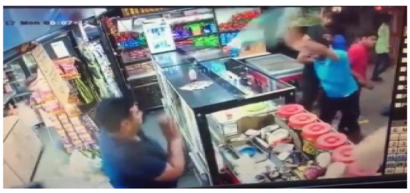
Methods	N: normal; A: abnormal	N%	A%
Xu et al. [18]	N	78.4	21.6
	A	30.2	69.8
Li et al. [17]	N	81.4	18.6
	A	19.4	80.6
Sabokrou et al. [23]	N	86.6	13.4
	A	20.4	79.6
Proposed method	N	97.2	2.8
	A	6.2	93.8

Table 5 Equal error rate for anomaly detection comparison for local dataset

Methods	EER
Yachuang et al. [25]	22
Zhijun et al. [38]	21
Mousavi et al. [39]	35
Xu et al. [18]	25
Li et al. [17]	20
Sabokrou et al. [23]	18
Proposed method	12

Table 6 Confusion matrix of the proposed method and other existing method local dataset where N—normal and A—abnormal

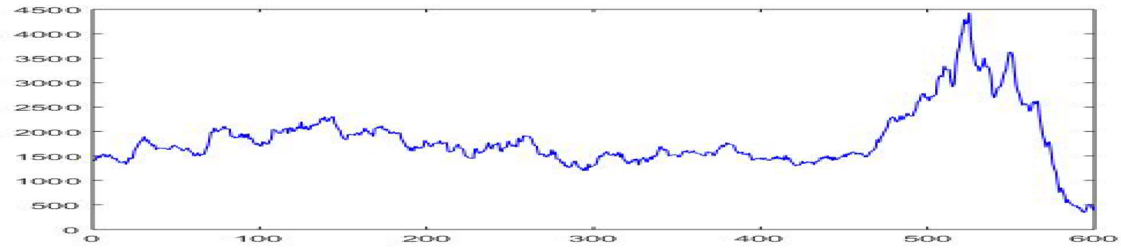
Methods	N: normal; A: abnormal	N%	A%
Yachuang et al. [25]	N	85.8	14.2
	A	17.4	82.6
Zhijun et al. [38]	N	86.8	13.2
	A	12.2	87.8
Mousavi et al. [39]	N	79.6	20.4
	A	38.4	61.6
Xu et al. [18]	N	75.6	24.4
	A	36.4	63.6
Li et al. [17]	N	86.4	13.6
	A	18.4	81.6
Sabokrou et al. [23]	N	94.7	5.3
	A	16.2	83.8
Proposed method	N	97.4	2.6
	A	4.4	95.6



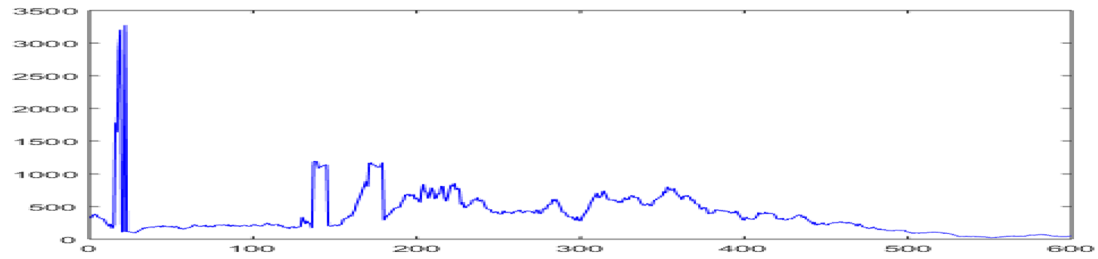
(a) Youtube dataset



(b) MOT dataset



(c) Result on Youtube video (X-axis represents frames and Y-axis represents the change in kinetic energy)



(d) Result on MOT 17 dataset (X-axis represents frames and Y-axis represents the change in kinetic energy)

Fig. 7 **a** YouTube dataset, **b** MOT dataset, **c** result on YouTube video (*x*-axis represents frames and *y*-axis represents the change in kinetic energy), and **d** result on MOT 17 dataset (*x*-axis represents frames and *y*-axis represents the change in kinetic energy)

becomes very less crowded. This results in the low value in the curve.

3.6 Real-time applicability

We analyzed the performance of the proposed method on UCSD datasets to find out the running time. We then compared our result with other state-of-the-art methods. Frame resolution considered is 238×158. For real-time analysis, we have used CUDA-enabled Nvidia Titan GPU MATLAB 2015a having parallel computing toolbox. We have calculated time duration for different steps used in the proposed method.

Step to be performed according to the proposed method and the time taken by the proposed method in seconds.

1. Formation of Fuzzy lattices: 0.00132.
2. Calculation of kinetic energy using Schrödinger wave equation: 0.00102.
3. Selection of most prominent lattices: 0.0002.
4. Classification: 0.0001.

Total time for anomaly detection: 0.00264.

We are processing a single frame in 0.00264 s. The proposed method is using 378 frames per second for the processing which makes it real-time applicable. Table 7 shows a comparison in the speed of the proposed method with other state-of-the-art methods.

The proposed method is giving far superior result as compared to the other state-of-the-art methods. Even if compared with recent methods based on CNN [23, 26,

Table 7 Run-time analysis of different methods

Methods	Run time (s)
Xu et al [18]	Offline
Li et al [17]	1.34
Sabokrou et al. [23]	0.04
Karpath et al. [26]	0.007
Meng et al. [27]	0.004
Sabokrou et al. [28]	0.0027
Proposed method	0.00264

Table 8 Comparison of the proposed method with other dimension-reduction technique

Method	Dimensions	Recognition accuracy (%)
PCA	42	80.5
LLE	32	86.4
Laplacian	15	78.4
Isomap	10	88
Proposed method	5	96.6

27] with fast running time, our method shows dominance. FCN-based method [28] that is an extended version of CNN is achieving very good real-time applicability (340 fps). But the proposed method is performing better as it is using 378 fps. One more reason this method is real-time applicable is its low dimensionality.

Table 8 shows a comparison between different dimension-reduction techniques. The proposed method uses only five dimensions.

4 Conclusion

The proposed method is a novel technique for anomaly detection in a video, where the concept of mathematics, physics, and computer science along with signal processing has been combined to model a detector for anomalies. It is capable of determining anomalies in real-time because it considerably reduces the dimensionality of a feature. Nonlinear Gaussian membership functions (GMF) have been applied to video for finding the dynamic features in terms of fuzzy lattices and their corresponding tensors. This method used the concept of the Schrödinger wave equation for modeling these features. Energy signal, obtained after applying the Schrödinger wave equation is classified into normal and abnormal events. To validate the proposed method, standard datasets such as UCF web dataset, UCSD dataset, and UMN dataset are used. The experimental result shows the superiority of our proposed method over other state-of-the-art methods for anomaly detection. The proposed method has also been applied to YouTube video and MOT17 dataset which gives good results.

Funding No funding is received.

Compliance with ethical standards

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

References

1. Cui, P., Sun, L., et al.: A sequential Monte Carlo approach to anomaly detection in tracking visual events. In: IEEE Workshop on Visual Surveillance, pp. 1–8 (2007)
2. Song, X., Wu, M., et al.: Conditional anomaly detection. IEEE Trans. Knowl. Data Eng. **19**(5), 631–645 (2007)
3. Serra, J.: A lattice approach to image segmentation. J. Math. Imaging Vis. **24**, 83–130 (2006)
4. Hervieu, A., Bouthemy, P., et al.: A statistical video content recognition method using invariant features on object trajectories. IEEE Trans. Circuits Syst. Video Technol. **18**(11), 1533–1543 (2008)
5. Dee, H., Hogg, D.: Detecting inexplicable behaviour. In: Proceedings of British Machine Vision Conference, pp. 477–486 (2004)
6. Basharat, A., Gritai, A., et al.: Learning object motion patterns for anomaly detection and improved object detection. In: Stauffer, C., Grimson, E. (eds.) CVPR. 1-8 (2008): Learning Patterns of Activity Using Real-time Tracking. IEEE Transactions on PAMI, vol. 22, pp. 747–757 (2000)
7. Stauffer, C., Grimson, E.: Learning patterns of activity using real-time tracking. IEEE Trans. PAMI **22**, 747–757 (2000)
8. Wang, X., Tieu, K., et al.: Correspondence-free activity analysis and scene modeling in multiple camera views. IEEE Trans. Pattern Anal. Mach. Intell. **32**(1), 56–71 (2010)
9. Zelniniker, E., Gong, S., et al.: Global abnormal behavior detection using a network of CCTV cameras. In: Eighth International Workshop on Visual Surveillance (2008)
10. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. Phys. Rev. E. **51**(5), 4282–4286 (1995)
11. Oliver, N., Rosario, B., et al.: A Bayesian computer vision system for modelling human interactions. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 831–843 (2000)
12. Zelnik-Manor, L., Irani, M.: Event-based video analysis. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2001)
13. Adam, A., Rivlin, E., et al.: Robust real-time unusual event detection using multiple fixed locations monitors. PAMI **30**(3), 555–560 (2008)
14. Mahadevan, V., et al.: Anomaly detection in crowded scenes. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1975–1981 (2010)
15. Kim, J., Grauman, K.: A space-time mrf for detecting abnormal activities with incremental updates. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2928 (2009)
16. Kapoor, R., Gupta, R.: Non-linearity dimensionality reduction using fuzzy lattices. IET Comput. Vis. **7**(3), 201–208 (2013)
17. Li, W., Mahadevan, V., et al.: Anomaly detection and localization in crowded scenes. IEEE Trans. Pattern Anal. Mach. Intell. **36**(1), 18–32 (2014)
18. Xu, D., Song, R., et al.: Video anomaly detection based on a hierarchical activity discovery within spatiotemporal contexts. Neurocomputing **143**, 144–152 (2014)
19. Schrodinger, E.: An undulatory theory of the mechanics of atoms and molecules. Am. Phys. Soc. Rev. **28**(6), 1049–1070 (1926)
20. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000)
21. Belkin, M., Niyogi, P.: Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. Advances in Neural Information Processing Systems, pp. 585–591. MIT Press, Cambridge (2001)
22. Tenenbaum, J., Silva, V., et al.: A global geometric framework for nonlinear dimensionality reduction. Science **290**(5500), 2319–2323 (2000)
23. Sabokrou, M., Fayyaz, M., et al.: Deep-cascade: cascading 3D deep neural networks for fast anomaly detection and localization

in crowded scenes. IEEE Trans. Image Process. **26**, 1992–2004 (2017)
24. Sabokrou, M., Fathy, M., et al.: Real-time anomaly detection and localization in crowded scenes. In: International Workshop on Computer Vision Pattern Recognition, pp. 56–62 (2015)
25. Yachuang, F., Yuan, Y., Xiaoqiang, L.: Learning deep event models for crowd anomaly detection. ACM J. Neurocomput. **219**(1), 548–556 (2017)
26. Karpathy, A., Toderici, G., et al.: Large-scale video classification with convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1725–1732 (2014)
27. Meng, B., et al.: Human action recognition based on quaternion spatial-temporal convolutional neural network and LSTM in RGB videos. Springer Multimed. Tools Appl. **77**(20), 26901–26918 (2018)
28. Sabokrou, M., Fayyaz, M., et al.: Deep-anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes. Computer Vision and Image Understanding. Elsevier, Amsterdam (2016)
29. Kapoor, R., Gupta, R.: Fuzzy lattice based technique for classification of power quality disturbances. In: European Transactions on Electrical Power, pp. 147–156 (2012)
30. Zhang, T., Lu, H., et al.: Learning semantic scene models by object classification and trajectory clustering. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1940–1947 (2009)
31. Piciarelli, C., et al.: Trajectory-based anomalous event detection. IEEE Circuit Syst. Video Technol. **18**(11), 1544–1554 (2008)
32. Cristianini, N., Shave-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, Cambridge (2000)
33. Schokopof, B., Smola, A.: Learning with Kernels Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, Cambridge (2002)
34. Moore, W.: Schrodinger: Life and Thought. Cambridge University Press, Cambridge. ISBN 0-521-43767-9 (1992)
35. UNN dataset, access on September 18, 2018 (online). http://mha.cs.umn.edu/Movies/CrowdActivity-All.avi
36. UCSD dataset (online). http://www.svcl.ucsd.edu/projects/anomaly/dataset.html. Accessed 18 Sept 2018
37. Solmaz, B., Moore, B., et al.: Identifying behaviors in crowd scenes using stability analysis for dynamical systems. IEEE Trans. Pattern Anal. Mach. Intell. **34**(10), 2064–2070 (2012)
38. Zhijun, F., Fei, F., et al.: Abnormal event detection in crowded scenes based on deep learning. ACM J. Multimed. Tools Appl. **75**(22), 14617–14639 (2016)
39. Mousavi, H., Nabi, M., et al.: Abnormality detection with an improved histogram of oriented tracklets. In: International Conference on Image Analysis Processing, pp. 722–732 (2015)
40. MOT challenge dataset (online). https://motchallenge.net/vis/MOT17-08-FRCNN. Accessed 18 Sept 2018
41. Youtube video (online). http://www.youtube.com/watch?v=pG3nIN_CxAY. Accessed 18 Sept 2018

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

Wireless Personal Communications
https://doi.org/10.1007/s11277-019-06683-6

Ant Colony Optimization and Excess Energy Calculations Based Fast Converging Energy Efficient Routing Algorithm for WSNs

Aarti Jain[1] · **Anuj Pathak**[1]

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Wireless sensor networks (WSNs) find their application in as diverse fields as collection of data for weather forecasting to detection of enemy activities by defense agencies. Considering the diverse and sensitive areas where WSNs are deployed, un-interrupted and timely delivery of data is as important as energy efficient delivery. This necessitates the requirement of a routing protocol that takes into account both the energy consumption and system delays while finding the best route to deliver packet/data from node to sink. In literature, a number of shortest path based algorithms viz. dikshatra, bellman ford, A*, floyd–warshall's and heuristic search based algorithms viz. Ant colony optimization (ACO), particle swarm optimization, evolutionary algorithms reinforcement learning have been proposed for enhancing the routing efficiency. ACO which is one of the heuristic search algorithms has proven to more efficient for routing methods due to its dynamic and flexible nature. In most of the ACO based routing algorithms total energy consumption and delay incurred by a path have been considered as two main optimization parameters for finding the optimal path between source and sink. However, due to little difference in the respective optimization parameters of different available paths, the convergence time of these algorithms is very high, which results in longer set up delay and higher energy consumption. In this paper, an ACO based routing algorithm has been proposed which considers excess energy (excess energy is that part of communication energy expenditure, which is used to move packet in direction perpendicular to the line of sight direction between source and destination) as one of the optimization parameters. The use of excess energy consumption as route selection parameter leads to faster convergence of the algorithm as well as results in finding more energy and delay efficient path. The proposed method has been simulated and compared with state-of-the-art ACO based routing methods i.e. deflection angle based ACO algorithm and E&D ANTS algorithm. The simulation results indicate that the proposed method has low convergence time, balanced energy consumption, lower time delay, high packet delivery ratio and leads to longer network lifetime.

Keywords Energy aware routing · Ant colony optimization · Optimality principle · Wireless sensor network · Convergence time

✉ Aarti Jain
aartijain@aiactr.ac.in

Extended author information available on the last page of the article

Published online: 23 August 2019

1 Introduction

WSNs have been a focus area of research in the field of wireless networking. WSNs include small sensor nodes deployed in either randomly or pre-planned manner in the geographical area under consideration. These small sensor nodes collect data with the help of sensors that are embedded onto the node itself. The collected data is first processed locally, converted into packets and then transmitted to the sink node for further analysis and processing. WSNs find extensive usage in applications that requires data gathering and monitoring. Few examples of such applications are defense, home security, smart home, weather forecasting and health and agriculture monitoring [1].

With the evolvement of data analytics, WSNs have gained a lot of popularity for collection of vast amount of data. Such popularity drives the need to develop protocols that address the constraints imposed by WSN networks while enhancing the required Quality of service (QoS) parameter of network performance. The constraints imposed by WSN network can be classified into two main groups 1. General constraints i.e. limited computational capabilities, limited energy, limited memory and small size. 2. Application specific constraints such as sensor node deployment, security of network, maintenance self-configuration. QoS parameters for WSN mainly requires routing of data from a originating node (source) to sink nodes (destination) with low end-to-end delay and with minimum number of intermediate nodes, energy efficient transmission, high packet delivery ratio and robustness [2–4]. This requires efficient routing protocol especially designed for WSNs that takes into account both the energy consumption and system delays while finding the best route to deliver packet/data from node to sink. In literature, a number of shortest path based algorithms viz. dikshatra, bellman ford, A*, floyd–warshall's and heuristic search based algorithms viz. Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Evolutionary Algorithms (EA) Reinforcement Learning (RL) have been used as optimization algorithms for finding efficient routing paths from sensor nodes to sink. ACO which is one of the Heuristic search algorithms has proven to more efficient for developing routing methods due to its dynamic and flexible nature. ACO is a biologically inspired problem solving technique proposed by Dorigo et al. [5–7]. In literature several ACO based routing algorithms have been proposed. In ACO based routing algorithms, limited database information is maintained at each node. Periodically each node generates some route build Agents Known as Ant packets. These ant packets are routed by utilizing local database information and eventually reach the destination node. During their backward journey these ant packets update local information at intermediate nodes. For subsequent transmission of local information, the updated local information is used to route the packet. Routing overhead reduces more and more as the algorithm progresses and converges to quality solution.

ACO have been a preferred choice of researchers for designing path finding algorithms in WSNs and results have proven that ACO based routing algorithms outperform the traditional dynamic programming based route finding algorithms [8]. Reviewed literature shows that in most of the ACO based routing algorithms total energy consumption and delay incurred by a path have been considered as two main optimization parameters for finding the optimal path between source and sink. However, due to little difference in the respective optimization parameters of different available paths, the convergence time of the algorithms becomes very high (exponential) or sometimes leads to stagnation, which results in longer set up delay and higher energy consumption.

In this paper, we propose a new Ant colony optimization based heuristic routing algorithm that considers the excess energy as one of the optimization parameter along with end-to-end delay being second parameter. The proposed algorithm compares the excess energy consumption vis-à-vis energy required by line of sight path to update the pheromone value. Excess Energy is the part of communication energy expenditure, which is used to move packet in direction perpendicular to the line of sight direction between source and destination. The domain/search space of the excess energy consumption parameter is smaller than the actual energy consumption parameter for any set of paths between source to destination. Thus, these parameters (excess energy, end to end delay) lead to faster convergence towards optimum solution as well as results in finding more energy and delay efficient path.

We have compared proposed algorithm against Deflection Angle Based ACO Algorithm [9] and E&D ANTS algorithm [10]. The simulation results indicate the capability of proposed algorithm to converge faster to optimum solution and to save network energy.

This paper has been divided into different sections based on the design of the proposed method. Section 2 presents the literature survey presenting the ACO based routing methods. Section 3 presents system modeling which contains network model, energy model and delay model used for simulations. Section 4 provides a brief introduction to the working of ACO. Section 5 introduces the proposed method, its description, mathematical background, working and pseudo-code. In Sect. 6 pertains to the performance evaluation of the proposed method based on the simulation results. The last Sect. 7 presents the conclusion.

2 Literature Survey

Ant colony based routing algorithms in WSNs have adapted the collective behaviors of natural ants for finding optimal route to food source. Ant usually achieve adaptive, scalable, and robust optimized paths between the nest and the source of food with little intelligence and capacities at each individual. This behavior of ants is very suitable from WSNs perspective which is composed of small sensors with limited capacities and resources in hostile and unpredictable environment.

In [11] Iyengar, the authors have investigated the application of several biologically-inspired methods viz. variants of ACO and (Genetic-Algorithm) GAs, in solving routing problems in WSNs. The implementation is done for various dynamic and heterogeneous networks. The investigations have proved that routing algorithms based on biologically-inspired methods have less congestion, end-to-end delay and computational-complexity. Similarly, survey presented in [12] has proven the advantages of ACO based routing algorithms over conventional and other biologically-inspired based routing algorithms.

ANTNET [13–16] is one of the first algorithms that utilize ACO phenomena to find optimum route in wired network. It utilizes agents namely forward ant that collect network information in its forward journey and backward ants which updates routing information by using forward ant's database. Although ANTNET outperform conventional algorithm for wired network, it does not perform well in case of wireless networks. ANTNET works fine for wired networks because of high speed and reliable connection. But the same is not true in case of wireless networks, where bandwidth is limited, reliability of the medium is a concern and link failures are frequent. Due to varying nature of wireless medium, delays

can vary even for the optimum path which leads to low convergence of ANTNET based routing algorithm.

EEABR [17] is a routing algorithm proposed to maximize WSN lifetime by taking energy consumption into the consideration while selecting optimum path. In EEABR, pheromone update rule is modified in such a way so that both energy consumption and hop count will contribute to pheromone accumulation. EEABR also tries to reduce the ant packet size to save energy consumption for transmitting those extra bits.

Liu et al. [9] has proposed a Deflection-Angle-Based ACO (DACO) routing algorithm that utilizes deflection angle to guide the ants from source to sink nodes. Forwarding a forward ant packet is a two-step process. The node first eliminates neighbor nodes whose cosine of deflection angle is less than zero, since such nodes, if selected, will move ant away from sink. One of the remaining neighbor nodes is selected probabilistically, depending upon the cosine of deflection angle and local probability table. Usage of cosine in hop selection procedure may result in slow divergence problem. In slow divergence problem, the path with lower initial divergence is selected without taking into consideration the overall divergence leading to longer path, thus neglecting the shorter paths having higher divergence at the start.

E&D ANTS algorithm was proposed by Wen et al. [10] to address the energy constraint of WSNs. In WSN, node has to operate on battery source that may not get replaced. Thus it requires a special algorithm that should maximize network lifetime while reducing end-to-end delay. The E&D ANTS algorithm proposes some significant underlying changes in traditional ACO algorithm to enhance energy efficient transmission of packets. The algorithm includes energy expenditure as a metric to decide the quality of a solution. In traditional ACO algorithm delay is utilized to calculate pheromone trail. This helps the network to operate for longer time period with some trade-offs with end to end delay. However, E&D ANTS method depends on Reinforcement Learning method for its convergence, which requires initial database collection for validated mean calculations of path. It also considers overall energy expenditure of path, which results in insignificant difference between good and best solutions and hence leads to slower convergence of the routing algorithm.

In [18] the authors have proposed a multi-path reliable routing method. The proposed algorithm has utilized ACO as optimization algorithm for finding multiple paths between source and BS. The method has considered energy and path length as prime factor while selecting between different paths between source and BS. In [19], a ACO aided routing protocol AntSensNet has been proposed for a cluster based WSNs. The AntSensNet has considered four QoS parameters viz. delay, energy, memory and packet loss for the selection of paths from clusters heads to sink node. Similarly in [20] a ACO aided cluster based topology (AVAPCR-ACO) has been considered to find out the energy efficient routes between different cluster heads.

In [21] a fuzzy aided ACO based routing algorithm named as FACOR has been proposed. In this algorithm, the next hop selection by ant is determined on the basis of fuzzy amount possessed by a neighbouring node. The fuzzy amount is calculated by using fuzzy inference rule, while remaining energy, distance, connectivity and pheromone values are used as inputs. The fuzzy inference system is itself computational complex and added more complexity to ACO algorithm and thus is not suitable for WSNs. Similarly in [22], ACO and fuzzy inference system has been used for efficient path selection, while using path length, residual energy and end-to-end delay as optimization parameters.

In [23] a cross layer approach named as BIOSARP has been proposed and metrics viz. link quality, remaining energy and transmission time has been taken as link selection optimization parameters. In this method ACO has been used as optimization algorithm to find routes between source and destination nodes. The results presented shows that the method results in improved packet delivery ratio and energy efficiency.

In [24] three routing algorithms which are improvements over basic ant based routing have been proposed. The first proposed algorithm has been named as SC. In SC the pheromone updation is dependent upon both the cost of transmission from source to sink as well as energy efficiency of the path. The second proposed algorithm has been named as FF. In FF algorithm, controlled flooding of forward ants is initiated to find multiple paths from source to sink. The pheromone updation of multiple paths is done as done in SC routing algorithm and finally for data transmission best path is selected on the basis of accumulated pheromone value. The third proposed method is FP. In FP, multiple paths are found as proposed in FF routing algorithm and multipath routing is used to simultaneously transmit both control packets (ants) and data packets.

In [25], the authors have proposed two routing methods viz. AR (Ant-based Routing) and IAR (Improved Ant-based Routing). In the first routing algorithm i.e. AR, reinforcement learning concept as proposed in [26] has been included to modify the routing algorithm proposed by Zhang et al. [24]. The basic AR algorithm is ant based routing algorithm which has considered the distance between the originating node and end node as only optimization factor to update the pheromone value. While, the second routing algorithm viz. IAR modifies the basic AR algorithm by including the distance between neighbour nodes of originating node and end node as additional optimization factor for updating the pheromone value. The simulation results show that IAR outperforms the AR algorithm.

In [27] authors have proposed location information based routing algorithm (ACLR). ACLR has used ACO for route finding between sink node and sensor nodes. The algorithm has used the location information of nodes and residual energy for pheromone deposit as well as for evaporation. Kadri et al. [28] have proposed an ACO based routing algorithm for WSNs. The proposed algorithm is the modification to the basic ARA routing algorithm for MANETS [29]. Unlike ARA BS launches ants to initiate route discovery process. The method has introduced the concept of handshaking and symmetric key exchange for ensuring better security in WSNs with minimum energy usage.

Reviewed literature shows that in most of the ACO based routing algorithms, energy efficiency of the selected path is mainly addressed while total energy consumption and delay incurred by a path have been considered as two main optimization parameters. A little emphasis is given to convergence rate of the route finding algorithm, which is a basic parameter for real time implementation of WSNs

In this paper, we propose a new Ant colony optimization based heuristic routing algorithm that considers the excess energy as one of the optimization parameter along with end-to-end delay being second parameter. The proposed algorithm compares the excess energy consumption vis-à-vis energy required by line of sight path to update the pheromone value. These parameters lead to faster convergence towards optimum solution as well as results in finding more energy and delay efficient path.

3 System Models

3.1 Network Model

We have considered a WSN network with N number of sensor nodes and M number of sink nodes. Each data packet generated by any source sensor node is required to be transmitted to nearest sink node. Each sensor node have its positional information which can be obtained by utilizing any position estimation method i.e. Localization technique [30], by mean of hardware method like GPS [31] or can be made available at the time of network deployment. Each sensor node can transmit data within its transmission range. Thus if the distance between the two communicating nodes is greater than maximum possible transmission distance, packets must hop through intermediate nodes to reach the destination node. If there are L number of neighbor nodes in the transmission radius of node i, then there exists a link (i, j) between node i and node j, where $(j \in L)$. Each node maintains local database of its neighbor nodes that includes link properties i.e. location coordinates, residual energy and next hop local probability table for node j where $(j \in L)$. The next hop Local probability table is a $[L \times M]$ matrix, where M the number of is sink nodes in the network and L is number of neighbor nodes of node i. The next hop local Probability table is having following structure.

$$P^i_{jm} = \begin{bmatrix} P^i_{11} & P^i_{12} & \cdots & P^i_{1M} \\ P^i_{21} & P^i_{L1} & \cdots & P^i_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ P^i_{L1} & P^i_{L2} & \cdots & P^i_{LM} \end{bmatrix}$$

Here P^i_{jm} represents probability that node i selects node j as the next intermediate node in order to reach destination sink node m, where (i∈N), (j∈L), (m∈M).

3.2 Energy Model

We have considered that the transmission of data packets follows free space propagation model [32], accordingly power received P_r at distance d for a given transmitted power P_t is given by following equation:

$$P_r = \frac{P_t}{4\pi d^2} \tag{1}$$

For proper reception of data, receiver requires a minimum signal power level P_r^{min} to counterfeit the propagation losses of the channel. Thus for a given P_r^{min} transmitter needs to transmit data with some minimum transmission power P_t^{min}. According to free space propagation loss considerations [32], this P_t^{min} is calculated as follows.

$$P_t^{min} = P_r^{min} \times 4\pi d^2 \tag{2}$$

For a given power consumption, the energy consumption is calculated as:

$$Energy = Power \times Time \tag{3}$$

Thus for given transmission duration of time T_{Tx} and distance d_{ij} between nodes i and j.The minimum energy consumption in transmitting a packet from node i to node j is calculated as follows,

$$Ex_{ij}^{min} = P_t^{min} \times T_{Tx} \tag{4}$$

Using Eq. (2)

$$Ex_{ij}^{min} = 4\pi P_r^{min} d_{ij}^2 \times T_{Tx} \tag{5}$$

For a predefined P_r^{min} value and time period T_{Tx}, for which node is in transmitting mode, the above equation is restructured as follows:

$$Ex_{ij}^{min} = C_x \times d_{ij}^2 \times T_{Tx} \tag{6}$$

where C_x is constant having unit Watt per meter square. This energy Ex_{ij}^{min} is used to counter free space propagation losses that inherently reduce the EM wave power as it travel through the space. Further, minimum energy spent on receiving a data packet is calculated as follows:

$$Er_{ij}^{min} = C_r \times T_{Rx} \tag{7}$$

where C_r is a constant, having unit Watt and represents energy consumed on demodulating of information from received power. Assuming $T_{Tx} = T_{Rx}$, total energy consumed for communication of a packet from node i to node j is given by

$$E_{ij}^{min} = \left(C_x \times d_{ij}^2 + C_r \right) \times T_{Tx} \tag{8}$$

3.3 Delay Model

Based upon the delays encountered at different layers, the end-to-end delay can be segregated into three categories i.e. Queuing Delay, Media Access Delay and Propagation Delay [33, 34]. Queuing delay is experienced by packets at network layer. Queuing delay is the time period between entry of packet in the queue and when it is passed to MAC layer for transmission and is caused when a node generates/receives packets more than its transmission capacity. Media excess delay is experienced at data-link layer. It is the time difference between when a data reaches MAC layer and when it is passed to physical layer. This delay is caused is dependent on the contention period for a given node. Propagation delay occurs at physical layer and is the time difference between sender starting transmissions to receiver starting reception of data. This delay is time taken by of the EM waves to travel up to a distance and delay experienced due to transmitter and receiver hardware units. Accordingly, in this paper we have considered end-to-end delay as a sum of Queuing delay, Media Access Delay and propagation delay.

3.3.1 Queuing Delay Calculations

Let p_i represents the number of neighboring sensor nodes i.e. sensing only of node i on path P and q_i represents the number of relay nodes i.e. relaying only of node i for path P.

For a given data generation rate λ_{RT} for sensor nodes and data service rate (transmission rate) $r\zeta$ on sensor node, the total data produce by p_i (sensing node) will be $p_i\lambda_{RT}$ and total data load to be relayed by q_i (relay node) will be $\sum_{j=1}^{q_i} r_j\zeta$. Here r represents bandwidth ratio. It is basically an initial value set by the user and represents the bandwidth to be allocated for data transmission on a outgoing link and ζ represents the packet drop ratio. Then the total data load denoted as $\lambda_{RT}^{(i)}$ to be transmitted by a sensor node i is:

$$\lambda_{RT}^{(i)} = p_i\lambda_{RT} + \sum_{j=1}^{q_i} r_j\zeta \tag{9}$$

This leads to total queuing delay on node i is:

$$TQ_{RT}^{(i)} = \lambda_{RT}^{(i)}/r_i\zeta \tag{10}$$

Then average end-to end queuing delay for a particular path P is:

$$E[T_E] = \sum_{i\in Path} TQ_{RT}^{(i)} \tag{11}$$

$$\begin{aligned} &= \sum_{i\in path}\left[\frac{p_i\lambda_{RT} + \sum_{j=1}^{q_i} r_j\zeta}{r_i\zeta}\right] \\ &= \left[\frac{\lambda_{RT}}{\zeta}\right]\left[\sum_{i\in path}\frac{p_i}{r_i}\right] + \sum_{i\in path}\left[\frac{\sum_{j=1}^{q_i} r_j}{r_i}\right] \\ &= \left[\frac{\lambda_{RT}}{\zeta}\right]\left[\sum_{i\in path}\frac{p_i + \sum_{j=1}^{q_i} r_j}{r_i}\right] \end{aligned} \tag{12}$$

3.3.2 MAC Delay Calculations

The MAC delay is mainly consists of Carrier sense delay, back-off delay and transmission delay and is calculated as follows:

Suppose there are K hops from the source node to sink node. The carrier sense delay and back-off delay is random for each hop, and we denote their value at hop k by $t_{cs,k}$ and $t_{bd,k}$. Their mean value is obtained from contention window size and back-off timer limit and is denoted by t_{cs} and t_{bd}. The transmission delay has a fixed value for a fixed packet length and is denoted by t_{tx}.

The average delay at hop count n is $t_{cs} + t_{bd} + t_{tx}$. Accordingly, the total MAC delay for K hops is

$$T_{MAC} = \sum_{k=1}^{K} t_{cs,k} + t_{bd,k} + t_{tx} \tag{13}$$

So the average MAC delay over K hops is

$$E\left[T_{MAC}\right] = K(t_{cs} + t_{bd} + t_{tx}) \tag{14}$$

3.3.3 End to End Propagation Delay Calculations

The end to end average propagation delay for a path P is given by:

$$E[T_P] = \sum_{i,j \in Path} c \times dist_{ij} \tag{15}$$

Here, c is a constant and is equal to speed of wireless transmission and $dist_{ij}$ represents the distance between node i and node j

Finally total end-to-end delay is given by:

$$\begin{aligned} T_{E-E} &= E[T_{MAC}] + E[T_E] + E[T_P] \\ &= K(t_{cs} + t_{bd} + t_{tx}) + \left[\frac{\lambda_{RT}}{\mu}\right]\left[\sum_{i \in path} \frac{p_i + \sum_{j=1}^{q} r_j}{r_i}\right] + \sum_{i,j \in Path} c \times dist_{ij} \end{aligned} \tag{16}$$

4 Basic ACO Algorithm

Ant colony optimization (ACO) is a biological inspired algorithm that is proposed by Dorigo et al. [7, 8, 11]. ACO is one of the popular methods to solve the NP hard optimization problems by utilizing distributed learning process inspired from biological ants. This algorithm exploits stigmergy property of ant's learning that utilizes pheromone trails dropped along the path to obtain its favored degree of selection. In real world, biological ants wander randomly in the search area to find food source. Upon reaching the food source each ant travels back towards the nest and drops some volatile compound known as pheromone along the path. Being a volatile compound pheromone tends to evaporate with time and thus its intensity reduces with time. Any ant wandering in the search area reads the intensity of pheromone value and makes the decision to follow the path depending upon the same. Thus more the number of ants travelling back from food source, the more will be pheromone intensity and the more number of ants will follow the said path.

As the ants traveling longer path takes longer time as compared to ants traveling shorter path this leads to higher decay of pheromone dropped along longer path as compared to pheromone dropped along shorter path. Thus, this process continues to accumulate pheromone along the most favorable path. This proportional decay property of pheromone makes sure that ants will ultimately converge to the best path.

This biological behavior is modeled in an algorithm as under

1. The WSN is modeled as undirected graph $G(N, E)$, where N represents number of sensor nodes and E represents the edge/link between two nodes.
2. The source node generates some route setup packets which are termed as Forward Ant Packets
3. These Forward Ant Packets are forwarded to next node randomly utilizing probability matrix until they reach their destination node
4. In the forward journey, each packet creates a local database containing information regarding its own path. This database includes nodes it has visited, cost metric like delay and energy.
5. Upon reaching the destination node backward ant packet is generated by utilizing forward ant's database. This backward ant is traversed along the same path traversed in its forward journey, in reverse order (destination to source).

6. In its backward journey, ants will update pheromone value of each path at nodes database using cost metric parameters
7. At the end of an iteration probability matrix is updated by utilizing pheromone deposited by all ants using following equation

$$P_{jd}^{i} = \frac{\left[\tau_{jd}\right]^{\alpha} * \left[\eta_{ij}\right]^{\beta}}{\sum_{j \in L}\left[\tau_{jd}\right]^{\alpha} * \left[\eta_{ij}\right]^{\beta}} \tag{17}$$

where τ_{jd} is pheromone intensity of the edge $i \rightarrow j$ in order to reach destination d, η_{ij} is the heuristic value indicating the desirability of selecting node j by node i as next hop forwarder, α and β represents relative importance of pheromone τ_{jd} and heuristic value η_{ij} and L represents number of neighbor node of the node i.

Pheromone value for subsequent cycle is calculated as follows:

$$\tau_{jd}(t+1) = \rho\tau_{jd}(t) + \Delta\tau_{ij}^{best} \tag{18}$$

where ρ represents decay factor that simulates evaporation property of pheromone. Decay factor is incorporated to avoid stagnation and self-adaptability to the algorithm in respect to dynamic network characteristics. $\Delta\tau_{ij}^{best}$ represents the best solution obtained by an ant in the current iteration. Data packets are forwarded by paths having highest pheromone value.

8. Repeat steps 2–6 for next route discovery phase.

5 Proposed EwDACO Method

5.1 Mathematical Background

The E&D ANTS method proposed in [9] considers total energy and time spent by an ant while travelling from source to destination. As the path length increases, the difference between good and the best solution decreases and this leads to insignificant difference in the energy and delay cost parameter for differentiating good and the best solution. This makes the routing algorithm to take longer time to converge to best solution. In this paper, a new routing algorithm has been proposed, wherein the excess energy and delay are used

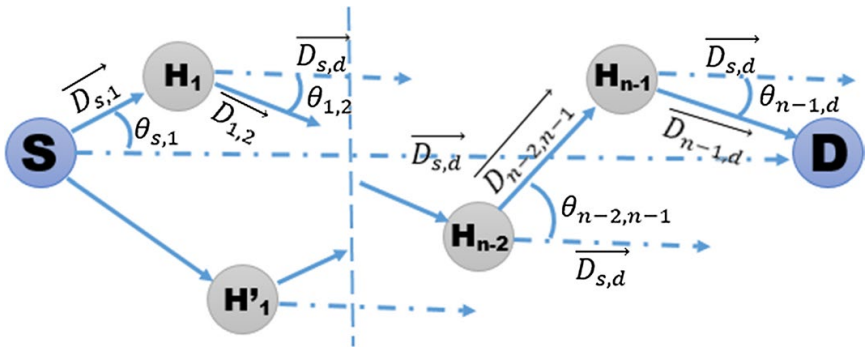


Fig. 1 Global routing problem

as a criterion for differentiating the optimum solution from a number of solutions obtained in iteration. Excess Energy is that part of communication energy expenditure, which is used to move packet in direction perpendicular to the line of sight direction between source and destination.

In order to differentiate between solutions obtained, the proposed method splits each solution in two parts i.e. 1. Displacement expenditure, which is the energy used to displace packet from source to destination along Line Of Sight (LOS) path. 2. Deviation expenditure, which is the energy used when a packet deviates away from LOS path. Displacement expenditure is energy parameter (similarity) that each solution must spend in order to reach from source to destination while deviation part varies depending on the quality of the solution. The proposed method eliminates this similarity among different solutions and considers this deviation expenditure as the path selection parameter. The same is elaborated with the following example:

Consider a multi hop network as given in Fig. 1, which is required to send a data packet from node source node S to destination node D. Given that each node has a maximum transmission range of R, data Packet from node S can reach destination node D by hoping through $n-1$ intermediate nodes. Minimum distance between any two points in space is known as displacement or Line of Sight (LOS) distance. For any Multi hop packet transmission from source to destination, this line of sight distance provides minimum distance to be travelled by the packet to reach the Destination. For data communication between node S and D, if the selected path for data transmission converges towards LOS, the following significant objectives can be achieved [35]:

1. Distance that a packet needs to travel to reach the destination decreases. Considering free space propagation loss, the energy expenditure for transmission reduces by a factor inversely proportional to D^2 (where D indicates distance). It also reduces propagation delay which is dependent on communication distance.
2. The shorter distance leads to lower number of hop counts and results in further reduction in energy expenditure, queuing delay and MAC delay.

Mathematically, let a packet is hopped from a node to another node, then the distance travelled by the packet can be represented as two elements. (1) Distance traveled in the direction of destination named as effective distance traversed (D^{eff}) and distance travelled in direction perpendicular to D^{eff} names as deviation distance (D^{dev}).

The $D^{eff}$and D^{dev} distance, when packet is hopped from node i to node $i+1$ can be calculated as under.

Given $\overrightarrow{D_{i,j}}$ is displacement vector from ith node to jth node and $\theta_{i,j}$ is the angle between $\overrightarrow{D_{i,j}}$ vector and $\overrightarrow{D_{s,d}}$ vector, then,

$$D_{i,j}^{eff} = \left|\overrightarrow{D_{i,j}}\right| \cos\left(\theta_{i,j}\right) \tag{19}$$

$$D_{i,j}^{dev} = \left|\overrightarrow{D_{i,j}}\right| \sin\left(\theta_{i,j}\right) \tag{20}$$

From the diagram, for a packet to travel from source S to D by hopping through $n-1$ intermediate nodes, following conditions must be satisfied

$$\sum_{i=0}^{n-1} Deff_{i,i+1} = \sum_{i=0}^{n-1} \left|\overrightarrow{D_{i,i+1}}\right| \cos\left(\theta_{i,i+1}\right) = Dmin \tag{21}$$

$$\sum_{i=0}^{n-1} Ddev_{i,i+1} = \sum_{i=0}^{n-1} \left|\overrightarrow{D_{i,i+1}}\right| \sin\left(\theta_{i,i+1}\right) = 0 \tag{22}$$

where $\theta_{i,i+1}$ can be calculated as follows

$$\theta_{i,i+1} = \cos^{-1}\left(\frac{\overrightarrow{D_{i,i+1}}.\overrightarrow{D_{s,d}}}{\left|\overrightarrow{D_{i,i+1}}\right|\left|\overrightarrow{D_{s,d}}\right|}\right) \tag{23}$$

Considering the energy model as discussed in Sect. 3.2, for a required minimum received power at receiver P_r^{min}, the minimum power required for transmission P_t^{min} is proportional to D^2. A part of this transmitted power is used to travel D^{dev} which does not contribute to the propagation of packet in the required direction. In the proposed method, the main focus is on the minimization of this component of energy (excess energy) to improve the network efficiency.

Using free space path loss propagation model, the excess energy utilized by a path while communicating from source to destination is calculated as:

$$Ew = \sum_{i=0}^{n-1} Ew_{i,i+1} \tag{24}$$

where $Ew_{i,i+1}$ is Excess Energy in hoping a packet from node i to node $i+1$ relative to LOS direction. This Excess Energy is calculated as follows by using Energy Model given in Sect. 3.2.

$$Ew_{i,i+1} = E_{i,i+1} \sin\left(\theta_{i,i+1}\right) \tag{25}$$

$$Ew_{i,i+1} = \left(C_x * d_{i,i+1}^2 + C_r\right) \sin\left(\theta_{i,i+1}\right) \tag{26}$$

$$Ew = \sum_{i=0}^{n-1} \left(C_x * d_{i,i+1}^2 + C_r\right) \sin\left(\theta_{i,i+1}\right) \tag{27}$$

$$Ew = \sum_{i=0}^{n-1} C_x d_{i,i+1}^2 \sin\left(\theta_{i,i+1}\right) + \sum_{i=0}^{n-1} C_r \sin\left(\theta_{i,i+1}\right) \tag{28}$$

Proposed method utilizes this Excess Energy parameter with end-to-end delay to analyze quality of solution. First term of Eq. (28) minimizes the distance traveled by a packet and the second term minimizes the hop count. For calculation of Excess Energy, each node must be able to calculate the angle between $\overrightarrow{D_{s,d}}$ vector and $\overrightarrow{D_{i,i+j}}$ vector. While the second vector $\overrightarrow{D_{i,i+j}}$ can be calculated by utilizing the local database of neighbor nodes, calculation of 1st vector $\overrightarrow{D_{s,d}}$ needs any one of the following.

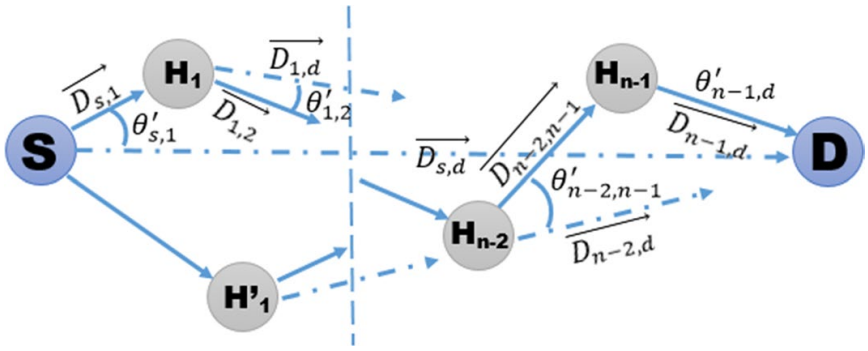


Fig. 2 Local routing problem

1. Every node know the position coordinates of every other source node in the network
2. Source node must send its position coordinate in ant packet.

But in WSN, due to limited buffer capacity of WSN node 1st condition cannot be achieved, and 2nd condition should be avoided to reduce packet size as larger packets require more communication energy.

Thus in order to calculate $\overrightarrow{D_{s,d}}$ Optimality Principle [36] has been applied to break down global problem into local problems that can be solved using local information. Optimality principle states that "if router j is on the optimal path from router i to router k, then the optimal path from j to k also falls along the same route." By using optimality principal, the problem of finding optimum path from source to destination is divided into problem of finding optimum path from any node itself to destination. Thus, each intermediate node is made to approach its own LOS rather than source to destination LOS.

Accordingly, Fig. 1 is modified to incorporate Optimality Principle in Fig. 2. In Fig. 2, the angle $\theta_{i,i+1}$ (in Fig. 1) which is the angle between vectors $\overrightarrow{D_{i,i+1}}$ and $\overrightarrow{D_{i,d}}$ has been updated to $\theta'_{i,i+1}$. Here $\overrightarrow{D_{i,i+1}}$ represents the vector between node i and its next hop node and $\overrightarrow{D_{i,d}}$ represents LOS from node i to destination node. Using given parameters, the excess energy used in deviating packet from node i to node $i+1$ can be calculated as follows

$$Ew_{i,i+1} = E_{i,i+1} \sin\left(\theta'_{i,i+1}\right) \tag{29}$$

Thus if a node i knows the minimum excess energy expenditure $Ew_{i+1,d}$ for transmitting a data packet by its neighbor nodes $(i+1)$ to destination node d, Node i can calculate its own excess energy parameter $Ew_{i,d}$ by adding one hop excess energy $Ew_{i,i+1}$ spent in hopping a packet from node i to neighbor node $(i+1)$ to the value $Ew_{i+1,d}$.

Thus the total energy wasted parameter can be calculated as follows

$$Ew_{i,d} = E_{i,i+1} \sin\left(\theta'_{i,i+1}\right) + Ew_{i+1,d} \tag{30}$$

Equation (30) is a chain equation with the chain terminating at last hop. In general, $Ew_{i,d}$ can be written as:

$$Ew_{i,d} = \begin{cases} 0, n = 1 \\ E_{i,i+1} \sin\left(\theta'_{\mathrm{i,i+1}}\right), n = 2 \\ E_{i,i+1} \sin\left(\theta'_{\mathrm{i,i+1}}\right) + Ew_{i+1,d}, n > 2 \end{cases} \tag{31}$$

where $\sin\left(\theta'_{\mathrm{i,i+1}}\right)$ can be calculated as

$$\sin\left(\theta'_{\mathrm{i,i+1}}\right) = \frac{\overrightarrow{D_{i,i+1}} \times \overrightarrow{D_{i,d}}}{\left|\overrightarrow{D_{i,i+1}}\right| . \left|\overrightarrow{D_{i,d}}\right|} \tag{32}$$

In WSN destination nodes are limited in numbers. Thus a node can maintain local database of fewer destination nodes and this will enable them to calculate $\overrightarrow{D_{i,d}}$ and consequently $\sin\left(\theta'_{\mathrm{i,i+1}}\right)$.

As energy being an important criterion in WSN network, in this paper it is proposed to minimize the excess energy spent in deviating a packet during source to destination journey. This objective has been achieved by updating pheromone value that is inversely proportional to Excess Energy parameter. Thus an ant will favor higher pheromone path to travel. Consequently, the path with lesser excess energy will be favored, which will further lead to conservation of energy.

5.2 Algorithm Description

In the proposed method, it is assumed that each sensor node knows its own location coordinate at any point of time. Each node broadcasts a hello packet containing its location coordinates at regular intervals. This packet is used for link setup if a hello packet is received from a node not present in local database. If a link is already established between the communicating nodes, its location coordinates are updated. Hello packets also re-initiate link expiry timer that helps in updating the link information status and prevent performance degradation caused by link failure. In WSN, nodes are either stationary or have less mobility, thus broadcast interval is chosen such that node position remains updated while preventing unnecessary hello packets broadcast. In WSNs node sink are fixed nodes with constant power supply. This enables sink node to broadcast packet to whole network which helps in maintaining sink related database at each nodes. These databases are used by nodes to calculate $\sin\left(\theta'_{\mathrm{i,i+1}}\right)$ Eq. (32) locally at each node and helps in reducing ants packet size of ant.

Each node generates an ant packet at fixed average time interval (μ). Ant's origin time is defined by normal distribution centered at average time (μ) and deviation (σ). This little randomness was introduced to prevent packet collision if all packets are generated at same fixed time. The ant packets are added to node buffer for subsequent transmission. Whenever a node has an ant packet to broadcast, node checks the nature of the ant packet and acts accordingly. The same is elaborated as under:

1. Forward Ant: The sender node selects one of its neighbor nodes according to local probability matrix. This selection procedure first eliminates nodes that have been already visited by the forward ant to build a list of candidate nodes then uses roulette wheel

selection method to select one of candidate nodes probabilistically. Before transmitting packet, sender node updates ant's database by making its own entry in the path database along with updating of delay information. After updating ant packet, node sends this packet to the selected node. Receiving node receives the packet and checks its destination address. If packet has reached its final destination, the node converts forward ant packet into backward ant packet and adds it to the buffer or else packet is added to buffer as it is.

2. Backward Ant: Node selects the previous node along the path the ant has followed using ant's database and transmits ant packet to the selected node after updating best cost information for the given sink node address in ant database. Upon receiving a backward ant, each node calculates the excess energy parameter by utilizing the local database. This excess energy along with the delay obtained from ant database is added to best cost obtained from ant's database to find best cost from current node (22).If Ant packet has not reached sender node, it is added to the buffer with highest priority else it is destroyed

When iteration timer expires, each node drops all the remaining ants' packets in order to prevent use of stale information. Each node utilizes best cost obtained to calculate pheromone value of each path. This pheromone values is used to update probability matrix. This probability matrix is used in next iteration.

$$f_{i,j}(t) = min\left\{ \sum_{i \in \{tab_k\}} \sum_{j \in L} \left(\left[Ew_{j,d}^{i} \right]^{a1} * \left[T_{j,d}^{i} \right]^{a2} \right) \right\} \tag{33}$$

$$\Delta \tau_{ij}^{best} = \frac{1}{f_{i,j}(t)} \tag{34}$$

Here $f_{i,j}(t)$ represents best cost of path originating from node i to destination d, $T_{j,d}^{i}$ represents end-to-end delay experience and $Ew_{j,d}^{i}$ represents Excess Energy parameter at node i if node j is selected as next hop in order to reach destination node d, a_1 and a_2 are constants defining optimization importance of each parameter.

As the nodes along the best path to sink will have lesser delay and excess energy expenditure, this result in higher pheromone accumulation along the path since higher pheromone value results in higher probability of selection, the subsequent iteration will have more ants on this path further increasing the pheromone value. Thus as the algorithm progresses, there is convergence towards optimum path. The next subsection presents the pseudo-code for the proposed EwDACO Method.

5.3 Pseudo-Code

1. Generate Ants
2. **WHILE** iteration timer 1 not expired
    a. **FOR** each Ant
        i. **IF** forward ant
            1. Calculate delay and Excess Energy
            2. Update path information in Ant's database
            3. **IF** destination reached
                a. Convert to backward ant
                b. Next node from ant's database
            ELSE
                c. Find next node using routing table
            END IF
            4. Hop to next node

        ii. **ELSE IF** backward Ant
            1. Update best cost information
            2. **IF** Source reached
                a. Destroy ant
            END IF
                a. Find previous node from ant's database
                b. Hop ant to previous node
            END IF
        END IF
    END FOR
END WHILE
3. Destroy all remaining ants
4. Update pheromone value
5. Update routing table
6. Wait till Timer 2 expires.
7. Repeat steps 1 to 5 for routing table updation.

5.4 Pseudo Code Description

For every new route discovery phase a random number of ants are generated in the beginning of each iteration (at 1), where each iteration is bounded by a timer 1(at 2). This timer is incorporate to reduce the effect of stale information. At each node, all ants (loop 2.a) are processed depending on their nature (branch 2.a.i). Forward ants collect path information (2.a.i.2) and update table value according to collected delay and excess energy value to next node (branch 2.a.i.3). If node has reached destination it is converted into backward ant and is traversed backward. Else next node is calculated from local routing table. Branch 2.a.ii handles backward node. Where backward node updates best cost information at current node. Branch 2.a.ii.2 handles lifetime of backward nodes. If ant has completed its cycles, it is destroyed by the node. Else ant is traversed backward using path information from its database. When iteration timer expires Loop 2 breaks and control reaches line 3 where all remaining ants at the nodes are dropped. Best cost information collected during the iteration is used to update pheromone value at each node (line 4). This pheromone change is transformed into routing information and is used to update routing table (at line 5). This routing table is used by nodes to forward data/information packets till timer 2 expires and this routing information is also used for future ants (at line 2.a.i.3.c) to update routing table. When timer 2 expires, next set of iterations for route discovery re-initiates itself.

6 Performance Evaluation

6.1 Simulation

Performance analysis is done with a discrete time simulator written in java. Nodes are placed randomly in an area of 500×500 m with approximate even density. Data, ant and hello packets are generated at constant average rate. Ants packets are generated every 2 s when node is in exploration state and every 30 s when nodes are in maintenance state. Hello Packets are generated every 2 min. Network traffic is assumed to be 10 packets per second with each packet having size of 64 Bytes. Packet generation incorporates little randomness around its average rate to prevent contention caused by packet generation at the same time. We have used CSMA/CA with Power Control as our MAC protocol. Each node has 1 J initial energy and transmission constant C_x is 21 mW while reception constant C_r is taken as 15 mW. We have implemented DACO, E&D ANTS and proposed EwDACO algorithm to compare the performance of each. Algorithm parameters are taken as $(\alpha = 0.8, \beta = 0.4, \rho = 0.9, Q = 100)$ for DACO, $(\alpha = 0.8, \beta = 0.4, \gamma = 0.1, x = 0.2, \rho = 0.9, \omega = 0.4)$ for E&D ANTS while parameters for our algorithm are used as $\left(a_1 = 0.8, a_2 = 0.6, \alpha = 0.8, \beta = 0.4, \rho = 0.9\right)$.

6.2 Results

We have simulated tests on different random network topology. Multiple tests are performed to minimize the effect of java random number generator on the simulation results. Individual test results are averaged to compile the final results. Figure 3 shows a random

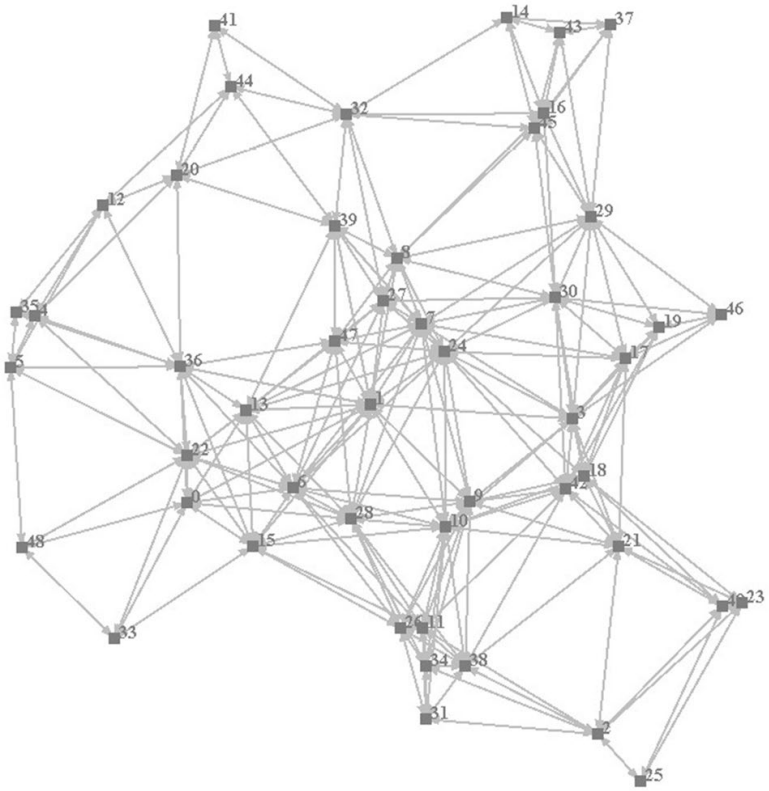


Fig. 3 Random graph for 49 sensor nodes

generated network topology. Performance of the proposed method is evaluated on the basis of algorithm convergence time, energy efficiency, PDR and E*D (energy delay metrics).

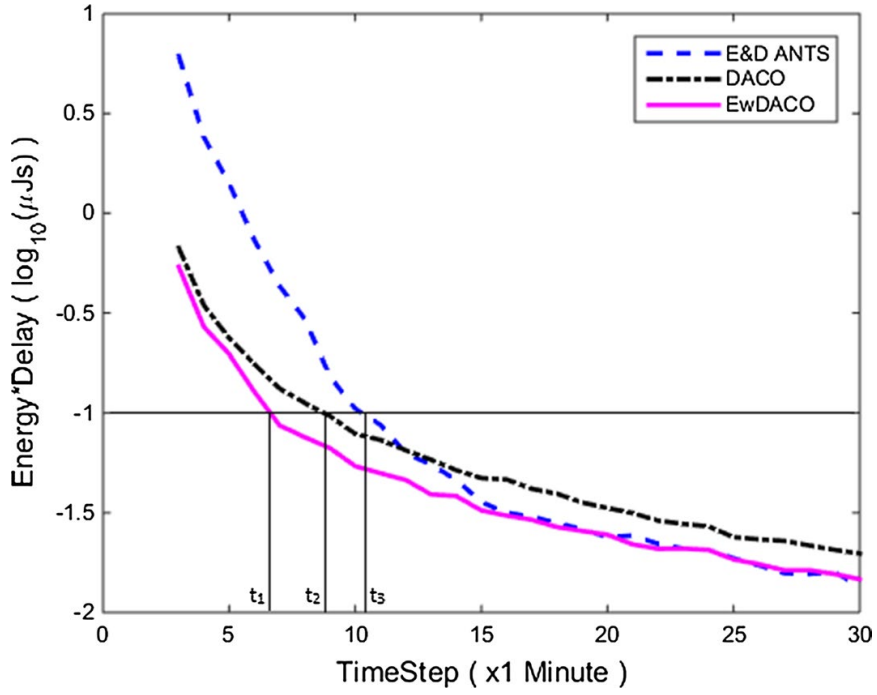


Fig. 4 E*D convergence test

1. Convergence test indicates the speed with which an algorithm reaches a given threshold of performance. The average Energy*Delay (E*D) cost of packet delivery is accounted after every minute of simulated time. Figure 4 shows the convergence of an algorithm to the optimum solution. Simulation result indicates that proposed EwDACO algorithm's convergence is faster than both DACO and E&D ANTS. As route setup agents are generated at constant rate and the time required in case of proposed EwDACO is minimum among the algorithms tested, it can be concluded that the proposed algorithm requires lesser routing overhead as compared to DACO and E&D ANTS for a given performance threshold. Figure 4 shows that for a performance threshold of -1, time required by algorithms EwDACO, DACO and E&D ANTS are in order $t_1 < t_2 < t_3$.

Figure 5 shows the average E*D expenditure experienced by a node at a given distance from the nearest sink node. Due to faster convergence of the proposed method towards optimum solution, E*D experienced in the case of EwDACO is minimum irrespective of the distance.

2. Residual Energy of network: WSN network are required to operate for longer time for a given amount of energy. This test analyzes the normalized residual energy of the network after every simulated minute. Figure 6 shows energy depletion curve of the network. It has been observed that EwDACO manages to save more energy as compared to DACO and E&D ANTS. At the end of 30 min of simulated time, proposed method is able to save nearly 15% energy as against other algorithms. Thus, the simulation results prove that the proposed method helps in extending the network lifetime.

We have also compared Energy distribution of network after 30 min of simulated time and the results are shown in Fig. 7. EwDACO algorithm provides higher mean Network energy in comparison to both DACO and E&D ANTS, which indicates better energy efficiency of the algorithm. Variance of network energy distribution shows the capability of algorithm to maintain even energy density across the network. From the simulation results

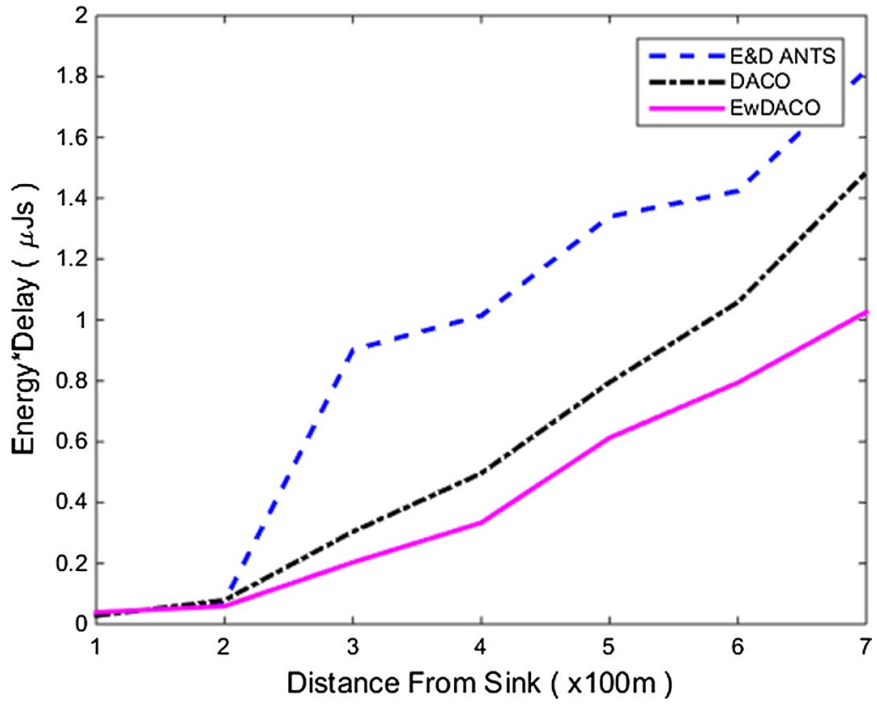


Fig. 5 E*D versus path length

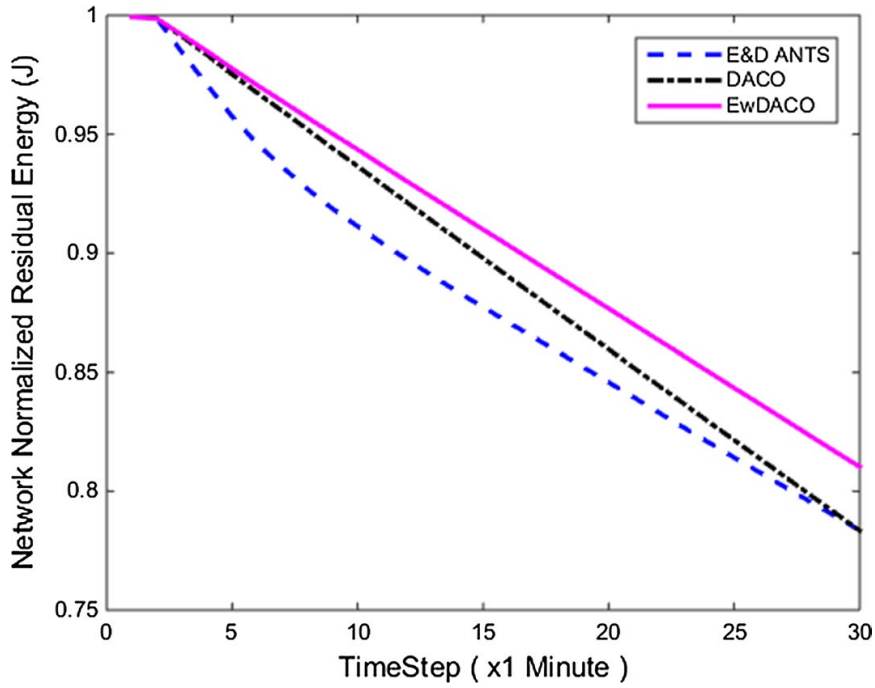


Fig. 6 Energy depletion of network over time

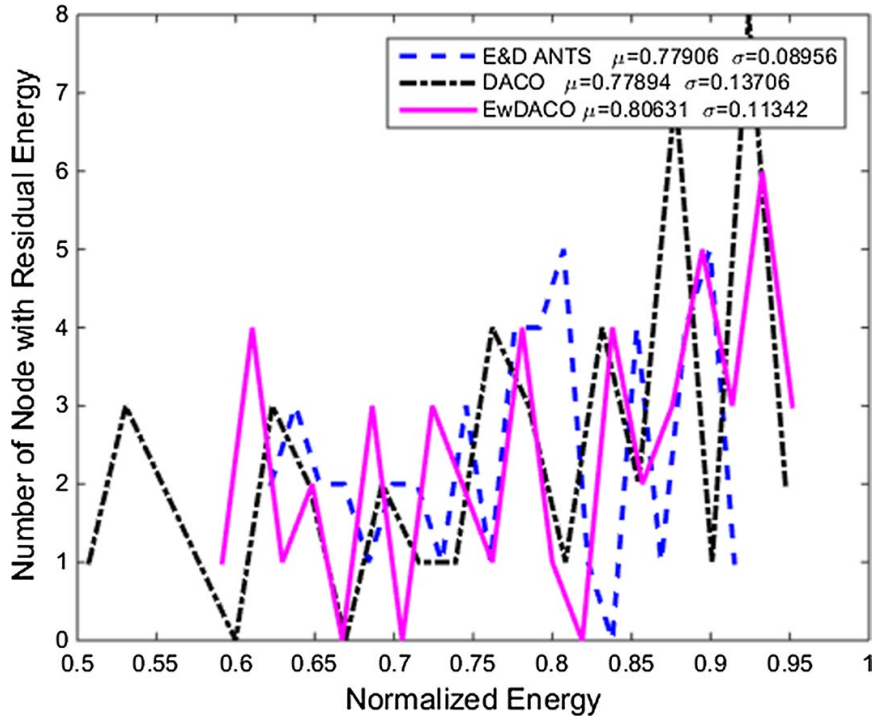


Fig. 7 Energy distribution of network

it has been observed that Variance in residual energy of the proposed algorithm is better than the compared algorithms (Figs. 8, 9).

3. PDR and E*D Comparisons In this test we have compared the Overall Performance parameters i.e. mean E*D expenditure and Packet Delivery Ratio of algorithms. Simulation results indicate that the proposed algorithm provides 57% reduction in E*D expenditure

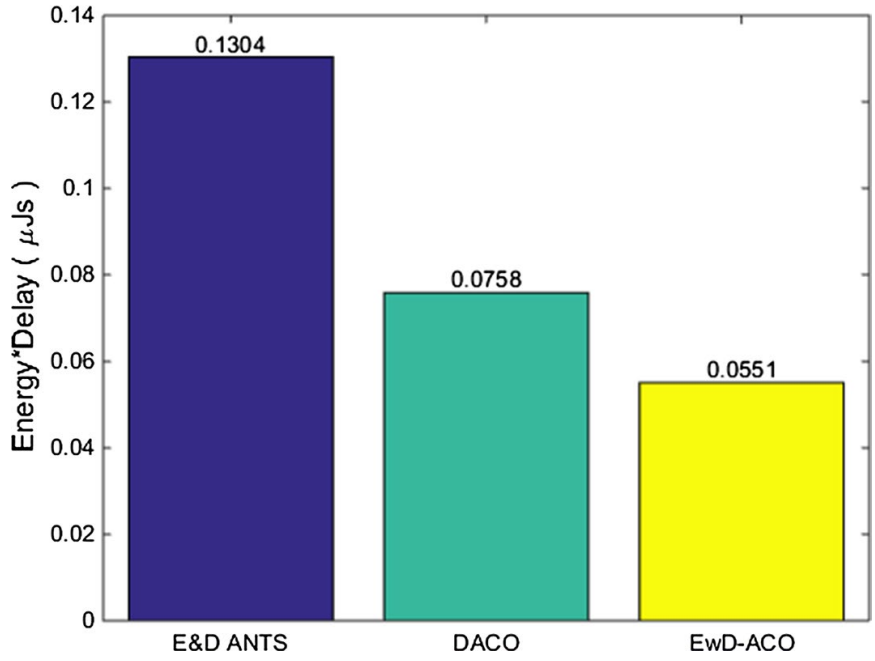


Fig. 8 Average E*D cost comparison

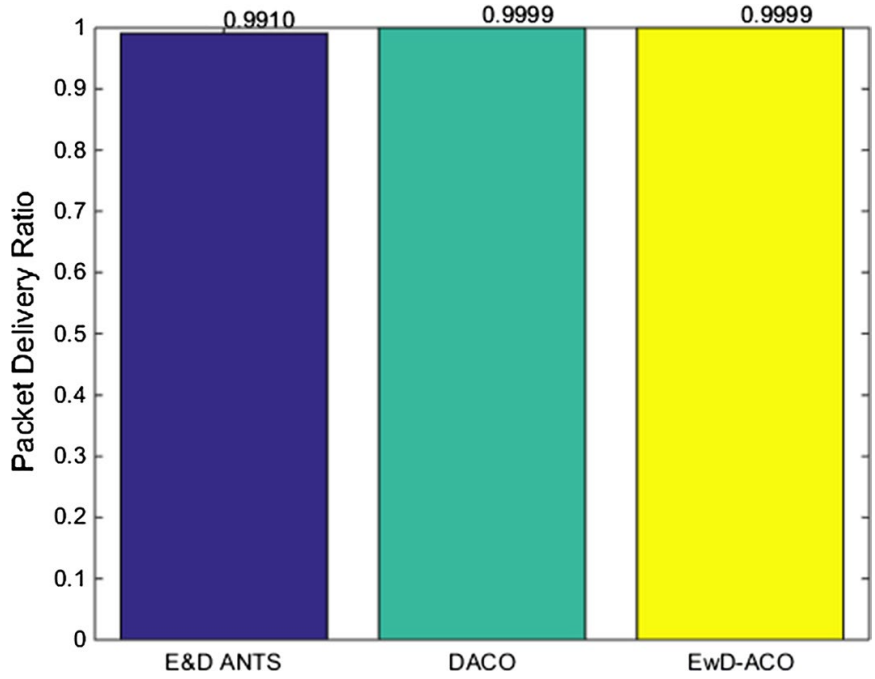


Fig. 9 Packet delivery ratio comparison

as compared to E&D ANTS and 27% reduction as compared to DACO. Thus in terms of overall performance, the proposed method provides better performance.

At the end of the simulation, we have also compared Packet Delivery Ratio of ACO, DACO, E&D ANTS, and EwDACO. EwDACO provides better PDR as compared to E&D ANTS and comparable to DACO.

7 Conclusion

In this paper, we have proposed a new routing algorithm that accounts the excess energy expenditure while communicating a packet from source to destination. The proposed method has utilized both energy and deflection angle for calculating the excess energy consumption. It can be concluded from the research work that consideration of excess energy consumption as against absolute energy consumption as optimization parameter has expedited the convergence of the algorithm to the optimum solution. The performance of the proposed method has also been compared with two state-of-the-art methods i.e. E&D ANTS and DACO in terms of various performance parameters i.e. convergence, energy consumption, packet delivery ratio and overall Energy*Delay metric. The simulation results indicate significant performance improvement with the proposed method. Moreover, it has been observed that proposed method quickly adapts the network deployment characteristics and hence converges fast to the optimal solution. In future the same work can be extended for efficient redeployment and movement of nodes towards destination in case of event driven mobile sensor networks.

References

1. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: A survey. *Computer Networks, 38*(4), 393–422.
2. Sha, K., Gehlot, J., & Greve, R. (2013). Multipath routing techniques in wireless sensor networks: A survey. *Wireless Personal Communications, 70,* 807–829.
3. Manap, Z., Ali, B. M., Ng, C. K., Noordin, N. K., & Sali, A. (2013). A review on hierarchical routing protocols for wireless sensor networks. *Wireless Personal Communications, 72*(2), 1077–1104.
4. Jain, A., & Reddy, B. R. (2015). Ant colony optimization based orthogonal directional proactive-reactive routing protocol for wireless sensor networks. *Wireless Personal Communications, 85*(1), 179–205.
5. Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 26*(1), 29–41.
6. Dorigo, M., Di Caro, G., & Gambardella, L. M. (1999). Ant algorithms for discrete optimization. *Artificial Life, 5*(2), 137–172.
7. Dorigo, M., Birattari, M., & Stutzle, T. (2006). Ant colony optimization. *IEEE Computational Intelligence Magazine, 1*(4), 28–39.
8. Kulkarni, R. V., Forster, A., & Venayagamoorthy, G. K. (2011). Computational intelligence in wireless sensor networks: A survey. *IEEE Communications Surveys & Tutorials, 13*(1), 68–96.
9. Liu, Y., Zhu, H., Xu, K., & Jia, Y. (2007). A routing strategy based on ant algorithm for WSN. In *Third international conference on natural computation, 2007 (ICNC 2007)* (Vol. 5, pp. 685–689). IEEE.
10. Wen, Y. F., Chen, Y. Q., & Pan, M. (2008). Adaptive ant-based routing in wireless sensor networks using Energy* Delay metrics. *Journal of Zhejiang University-Science A, 9*(4), 531–538.
11. Iyengar, S. S., Wu, H. C., Balakrishnan, N., & Chang, S. Y. (2007). Biologically inspired cooperative routing for wireless mobile sensor networks. *IEEE Systems Journal, 1*(1), 29–37.
12. Sim, K. M., & Sun, W. H. (2003). Ant colony optimization for routing and load-balancing: Survey and new directions. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 33*(5), 560–572.
13. Di Caro, G., & Dorigo, M. (1997). *AntNet: A mobile agents approach to adaptive routing*. Technical Report IRIDIA/97-12, IRIDIA, Université Libre de Bruxelles, Belgium.
14. Di Caro, G., & Dorigo, M. (1998). AntNet: Distributed stigmergetic control for communications networks. *Journal of Artificial Intelligence Research, 9,* 317–365.

15. Dhillon, S. S., & Van Mieghem, P. (2007). Performance analysis of the AntNet algorithm. *Computer Networks, 51*(8), 2104–2125.
16. Baran, B., & Sosa, R. (2000). A new approach for AntNet routing. In *Proceedings of the ninth international conference on Computer communications and networks, 2000* (pp. 303–308). IEEE.
17. Camilo, T., Carreto, C., Silva, J. S., & Boavida, F. (2006). An energy-efficient ant-based routing algorithm for wireless sensor networks. In *International workshop on ant colony optimization and swarm intelligence* (pp. 49–59). Berlin: Springer.
18. Okdem, S., & Karaboga, D. (2009). Routing in wireless sensor networks using an ant colony optimization (ACO) router chip. *Sensors, 9*(2), 909–921.
19. Cobo, L., Quintero, A., & Pierre, S. (2010). Ant-based routing for wireless multimedia sensor networks using multiple QoS metrics. *Computer Networks, 54*(17), 2991–3010.
20. Sun, L., Ma, H., & Hong, F. (Eds.). (2014). Advances in wireless sensor networks: In *7th China Conference, CWSN 2013, Qingdao, China, October 17–19, 2013*. Revised selected papers (Vol. 418). Springer.
21. Amiri, E., Keshavarz, H., Alizadeh, M., Zamani, M., & Khodadadi, T. (2014). Energy efficient routing in wireless sensor networks based on fuzzy ant colony optimization. *International Journal of Distributed Sensor Networks, 10*(7), 768936.
22. Jafari, M., & Khotanlou, H. (2013). A routing algorithm based an ant colony, local search and fuzzy inference to improve energy consumption in wireless sensor networks. *International Journal of Electrical and Computer Engineering, 3*(5), 640.
23. Saleem, K., Fisal, N., Baharudin, M. A., Ahmed, A. A., Hafizah, S., & Kamilah, S. (2010). Ant colony inspired self-optimized routing protocol based on cross layer architecture for wireless sensor networks. *WSEAS Transactions on Communications, 9*(10), 669–678.
24. Zhang, Y., Kuhn, L. D., & Fromherz, M. P. (2004). Improvements on ant routing for sensor networks. *Lecture Notes in Computer Science, 3172,* 154–165.
25. GhasemAghaei, R., Rahman, M. A., Gueaieb, W., & El Saddik, A. (2007). Ant colony-based reinforcement learning algorithm for routing in wireless sensor networks. In *IEEE instrumentation and measurement technology conference proceedings, 2007 (IMTC 2007)* (pp. 1–6). IEEE.
26. Lu, Y., Zhao, G., & Su, F. (2004). Adaptive ant-based dynamic routing algorithm. In *Fifth world Congress on intelligent control and automation, 2004 (WCICA 2004)* (Vol. 3, pp. 2694–2697). IEEE.
27. Wang, X., Li, Q., Xiong, N., & Pan, Y. (2008). Ant colony optimization-based location-aware routing for wireless sensor networks. In *International conference on wireless algorithms, systems, and applications* (pp. 109–120). Berlin: Springer.
28. Kadri, B., Feham, M., & Mhammed, A. (2014). Efficient and secured ant routing algorithm for wireless sensor networks. *IJ Network Security, 16*(2), 149–156.
29. Gunes, M., Sorges, U., & Bouazizi, I. (2002). ARA–the ant-colony based routing algorithm for MANETs. In *Proceedings of the international conference on parallel processing workshops, 2002* (pp. 79–85). IEEE.
30. Han, G., Xu, H., Duong, T. Q., Jiang, J., & Hara, T. (2013). Localization algorithms of wireless sensor networks: A survey. *Telecommunication Systems, 52,* 2419–2436.
31. Hofmann-Wellenhof, B., Lichtenegger, H., & Collins, J. (2012). *Global positioning system: Theory and practice*. Berlin: Springer.
32. Zhang, K. Q. (2015). *Wireless communications: Principles, theory and methodology*. New York: Wiley.
33. Akkaya, K., & Younis, M. (2003). An energy-aware QoS routing protocol for wireless sensor networks. In *Proceedings of the 23rd international conference on distributed computing systems workshops, 2003* (pp. 710–715). IEEE.
34. Ye, W., Heidemann, J., & Estrin, D. (2004). Medium access control with coordinated adaptive sleeping for wireless sensor networks. *IEEE/ACM Transactions on Networking (ToN), 12*(3), 493–506.
35. Golestanian, M., Azimi, M. R., & Ghazizade, R. (2014). Distributed cognitive routing in multichannel multi-hop networks with accessibility consideration. *International Transaction of Electrical and Computer Engineers System, 2*(6), 149–157.
36. Bagad, V. S., & Dhotre, I. A. (2009). *Data Communication Systems*. Technical Publications.

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



Aarti Jain graduated with a B.Tech. degree in Electronics and Communication Engineering from Beant Govt. Engineering College, Punjab, India in June 2002. Then she obtained her M.E. degree (with distinction) in Electronics and Communication from Delhi College of Engineering, University of Delhi, India in 2009. She has obtained her Ph.D. degree from Guru Gobind Singh Indraprastha University, Delhi in 2016 in the field of wireless sensor networks. Currently, she is working as Asst. Prof. at the Department of Electronics and Communication Engineering, Ambedkar Institute of Advanced Communication Technology and Research. She has contributed more than 30 research articles in the field of wireless sensor networks. Her current research interests are in Energy and quality management for wireless sensor networks, bio inspired computing and its application, IoT and node localization.



Anuj Pathak received his B.Tech. degree (with distinction) in Electronics and Communication in the year 2014 from Northern India Engineering College, GGSIPU, New Delhi and M.Tech. degree (with Distinction) in Digital Communications in the year 2016 from Ambedkar Institute of Technology, GGSIPU, New Delhi. He has worked as an Assistant Professor in Department of Electronics and Communication at AIACT&R, GGSIPU, New Delhi and Founded Bhawna Infinity Tech which provides innovative education platform. His research interests include Wireless Communications, Wireless Sensor Networks, Internet of Things and Ad hoc Networks.

Affiliations

Aarti Jain[1] · Anuj Pathak[1]

Anuj Pathak
iamanujpathak@gmail.com

[1] Department of Electronics and Communication Engineering, Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India

nanomaterials

MDPI

Article

Biofunctionalized Nanostructured Yttria Modified Non-Invasive Impedometric Biosensor for Efficient Detection of Oral Cancer

Suveen Kumar [1,2], Shweta Panwar [1], Saurabh Kumar [1], Shine Augustine [1] and Bansi D. Malhotra [1,*]

[1] Nanobioelectronics Laboratory, Department of Biotechnology, Delhi Technological University, Delhi110042, India

[2] Department of Chemistry, University of Delhi, Delhi110007, India

* Correspondence:bansi.malhotra@gmail.com

Received: 4 July 2019; Accepted: 17 August 2019; Published: 22 August 2019

Abstract: We report results of the studies relating to the development of an efficient biosensor for non-invasive detection of CYFRA-21-1 cancer biomarker. We used a low dielectric constant material (nanostructured yttrium oxide, nY_2O_3) for the fabrication of the biosensing platform. The nY_2O_3 was synthesized via solvothermal process and functionalized using 3-aminopropyl triethoxy silane (APTES). Electrophoretic deposition (EPD) of the functionalized nanomaterial (APTES/nY_2O_3) onto an indium tin oxide (ITO)-coated glass electrode was conducted at a DC potential of 50V for 60s. The EDC-NHS chemistry was used for covalent immobilization of −COOH bearing monoclonal anti-CYFRA-21-1 onto $-NH_2$ groups of APTES/nY_2O_3/ITO electrode. To avoid the non-specific interaction on the anti-CYFRA-21-1/APTES/nY_2O_3/ITO immunoelectrode, bovine serum albumin (BSA) was used. X-ray diffraction (XRD), transmission electron microscopy (TEM), and field emission scanning electron microscopy (FESEM) were utilized for structural and morphological studies, whereas Fourier-transform infrared spectroscopy (FTIR) was used for the bonding analysis. Cyclic voltammetry (CV) and electrochemical impedance spectroscopy (EIS) techniques were used for electrochemical characterization and response studies of fabricated electrodes. The fabricated immunosensor (BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO) exhibited linearity in the range of 0.01–50 ng·mL^{-1}, sensitivity of 226.0 Ω·mL·ng^{-1}, and lower detection limit of 0.01·ng·mL^{-1}. A reasonable correlation was observed between the results obtained using this biosensor and concentration of CYFRA-21-1 measured through ELISA (enzyme-linked immunosorbent assay) technique in salivary samples of oral cancer patients.

Keywords: non-invasive; dielectric constant; yttrium oxide; metal oxide; nanoparticles; CYFRA-21-1

1. Introduction

Cancer is an abnormal and uncontrolled cell growth [1,2]. According to a WHO report, cancer is currently the secondmost death-causing disease worldwide and about 8.8 million deaths were reported in 2015 due to cancer [3]. Lung, colon, breast, prostate, oral, and ovarian cancers are the most prominent cancers in the present-day world. Oral cancer (OC) is the sixthmost death-causing cancer and it can occur in lips, cheeks, gingiva, or palate part in the mouths of human beings. The main reason behind oral cancer is the mutation, which occurs at the gene level due to which uncontrolled cell cycle and differentiation do not take place [4]. Mutations in a gene may occur due to the use of a mutagenic agent which is widely present in tobacco, cigarettes, cigars, alcohol, etc. [5] The current detection techniques such as biopsy, cytopathology, visualization adjuncts, etc. can be used for detection of OC at advanced stages [6–9]. Moreover, these techniques require a tissue

sample for analysis, and hence are invasive, time-consuming, expensive, and require trained experts for sample collection as well as data analysis. To overcome these limitations, non-invasive [10] biosensors can play an important role since they offer many advantages such as painless detection, high sensitivity, cost-effectiveness, and require low sample volume.

For a non-invasive biosensor, salivary CYFRA-21-1 biomarker can be used for OC detection [11,12]. Kumar et al. recently used a number of materials including zirconia, hafnia, and nanocomposites for development of non-invasive biosensors. Tiwari et al. used lanthanum hydroxide for the fabrication of a biosensor for CYFRA-21-1 detection via CV or DPV (differential pulse voltammetry) techniques [13–18]. In these methods, even a slight change in volume of a biological sample and analytical solution can adversely affect the peak current response. Besides this, the accurate determination of active surface area of the immunoelectrode is an important issue. These limitations can, perhaps, be overcome by using an EIS technique.

In the fabrication of a biosensor, nanostructured metal oxides (NMOs), nanostructured metal sulfides (NMSs), reduced graphene oxide, carbon nanotubes, etc. have been extensively used as immobilization matrices [19–26]. The nanostructured oxides of metals (e.g magnesium oxide, zinc oxide, molybdenum trioxide, hafnia, titania, zirconia, etc.) have been found to have fascinating nano-morphological, functional, biocompatible, and electrochemical properties, and enhanced electron-transfer kinetics [19,27–33]. Among these NMOs, the nano-sized yttrium oxide (Yattria, nY_2O_3) exhibits high surface-to-volume ratio, fast oxygen ion mobility, efficient charge transfer ability, chemical inertness, sharp line emission bands, and biocompatibility [34–36]. In addition, it is known to have high quantum yield, excellent photo stability, and low dielectric constant (k), making it a potential candidate for application towards the development of biosensors [37–41]. The low dielectric constant (13) of yttrium oxide makes the thin film highly conductive [42]. Besides this, the oxygen moieties in Y_2O_3 can facilitate functionalization and covalent immobilization of antibodies. Rasheed et al. developed yttria-reduced graphene oxide nanocomposite based genosensor for detection of BRCA 1 (breast cancer) gene [40]. Efforts have also been made to use yttria–zirconia nanocomposite for the fabrication of sensors that can be used for detection of hydrogen, oxygen, and nitric oxide [41,43].

In this paper, we report results of the systematic studies relating to the development of a low dielectric material (nY_2O_3)-based electrochemical impedometric biosensor for detection of the salivary CYFRA-21-1 biomarker for OC detection. This immunosensor (BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO) exhibits higher linear range (0.01–50 ng·mL^{-1}) with remarkable sensitivity (226.0 Ω·mL·ng^{-1}).

2. Experimental Section

2.1. Preparation and Functionalization of Nanostructured Yttrium Oxide (nY_2O_3)

Solvothermal synthesis process of nanoparticles was used for the synthesis of nY_2O_3. Initially, we separately prepared 4 mM of yttrium (III) nitrate hexahydrate, 2 mM of CTAB, and 80 mM of urea in Milli-Q water. The CTAB solution was added dropwise into yttrium (III) nitrate hexahydrate solution at 40 °C with constant stirring of 300 rpm for 2 h. Subsequently, urea was added dropwise into this solution at the same temperature and stirring conditions. Further, the whole solution was transferred to a hot (80 °C) ultrasonication bath for 2h. The obtained precipitate was washed three times with Milli-Q water, followed by ethanol with the help of a centrifuge at 4500 rpm for 30 min, after which the precipitate was kept in a hot air oven at 80 °C to make it moisture-free. The dried product was transferred into the muffle furnace at 800 °C for 2h. The obtained white product was crushed usinga pestle and mortar and was stored in a desiccator at room temperature.

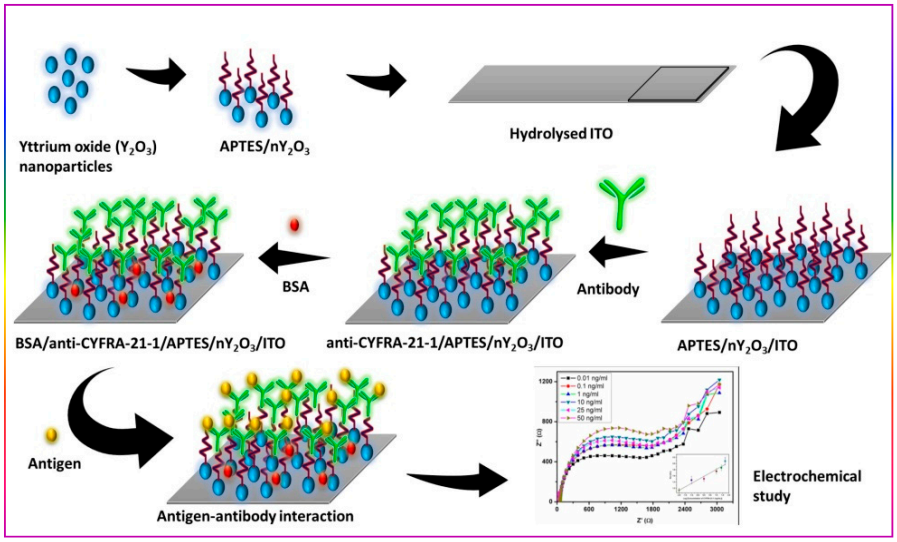
For functionalization of nY_2O_3 with APTES molecules, 100 mg of nY_2O_3 was dissolved in a minimal amount of isopropanol and then sonicated for 30 min. The solution was kept at constant stirring (300 rpm) at 50 °C until the nanoparticles were completely dispersed, after which we added 1 mL of APTES dropwise and 5 mL of Milli-Q water. This solution was kept in a magnetic stirrer at 200 rpm at 40 °C for 48 h. Subsequently, the solution mixture was washed with Milli-Q water (4500

rpm for 15 min) to remove any unbound APTES molecules. After centrifugation, the obtained powder was kept in a hot air oven at 50 °C until a moisture-free product was obtained.

2.2. Fabrication of Impedometric Biosensing Platform

A colloidal suspension of functionalized nanomaterial (APTES/nY_2O_3; concentration of 0.5 mg·mL^{-1}) in acetonitrile was prepared via ultrasonication. Thereafter, 10 mL of colloidal solution was poured into a two-electrodes system (platinum as reference and ITO as working electrode), kept 1 cm apart to each other. Further, the optimized DC potential of 50V was applied in the two-electrodes system using an electrophoretic deposition (EPD) unit for 60s. The uniform thin layer of APTES/nY_2O_3 was obtained on ITO electrode.

Immobilization of the monoclonal antibodies (anti-CYFRA-21-1) was carried onto APTES/nY_2O_3/ITO electrode via EDC-NHS coupling chemistry. For this purpose, we took 15 µL of anti-CYFRA-21-1 and mixed it with 7.5 µL of 0.4 M of EDC and 7.5 µL of 0.1 M NHS, and uniformly drop cast onto APTES/nY_2O_3/ITO electrode. This process helped in covalent bond formation between −COOH and −NH_2 group present on anti-CYFRA-21-1 and APTES/nY_2O_3/ITO electrode, respectively. Subsequently, the fabricated platform (anti-CYFRA-21-1/APTES/nY_2O_3/ITO) was washed with PBS to remove any unbound antibodies. For blocking of non-specific binding sites of the prepared electrode, we spread 30 mL of BSA (1 mg·dL^{-1}) molecule. Finally, the fabricated immunoelectrodes (BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO) were stored at 4 °C. Scheme 1 shows the stepwise fabrication process of the BSA/anti-CYFRA-21-1/APTES/nY2O3/ITO biosensor.



Scheme 1. Stepwise fabrication process of the BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO biosensor.

2.3. Biocompatibility Studies of nY_2O_3

The biocompatibility study of nY_2O_3 was carried out on HEK 293 cell line by using MTT colorimetric assay technique. In this technique, the yellow color of MTT dye got converted into purple color of the formazan product by mitochondrial succinate dehydrogenase enzyme present in the live cells. The colored formazan product was solubilized in a buffer, and further, intensity of color was measured at 540 nm by using an ELISA plate reader.

For biocompatibility assay, 96-well plates were used. First of all, we seeded the HEK 293 cells (cell density as 105 cells per well) in a 96-well plate and incubated it at 37 °C in a humidified 5% CO_2 incubator. After the cells attained 70% confluence, we added 10 mg·mL^{-1} to 250 mg·mL^{-1} of nY_2O_3 and again incubated it for 48 h in a CO_2 incubator. For the MTT assay, 200 mL of MTT solution (0.5 mg mL^{-1} in DMEM) was added to each well and incubated for 2h. Further supernatant aspirated and the buffer (100 mL isopropanol containing 0.06M HCl and 0.5% SDS) were added to each well to

solubilize the formazan crystals. With the help of an ELISA plate reader, we measured the absorbance of colored product obtained at 540 nm. The untreated cells were taken as control with 100% viability and wells with MTT reagent without cells were used as blank to calibrate the spectrophotometer to zero absorbance. The relative cell viability (%) compared to control cells was calculated using $[abs]_{sample}/[abs]_{control}\times100$. All the experiments were carried out in triplicate.

3. Results and Discussion

3.1. Structural and Morphological Studies

To investigate the structural and morphological characteristics of synthesized nY_2O_3 and the modified ITO electrodes, we conducted XRD, FESEM, and TEM studies. The obtained XRD pattern of the material is shown in Figure 1a. The diffraction peaks found at 2θ angle of 20.5, 29.2, 33.9, 40.0, 43.6, 48.5, 57.7, and 78.7 corresponded to the planes of (211), (222), (400), (332), (134), (440), (622), and (662), respectively. These peaks are in good correlation with the JCPDS NO. 65-3178, indicating successful synthesis of Y_2O_3 with a crystalline size (D) ~ 80 nm, calculated using the Debye Scherer method[14]. To confirm the size of the synthesized nanostructured material, we carried out TEM analysis at different magnifications (Figure 1b–d) and found that the synthesized nanomaterial was oval-shaped and had an average size of ~80 nm. The FESEM technique (Figure 2) was used at different magnifications to investigate the morphology at the surface of (i) nY_2O_3, (ii) electrophoretically deposited APTES/nY_2O_3, and (iii) anti-CYFRA-21-1/APTES/nY_2O_3 electrodes. The images (Figure 2a–c) observed at different magnifications confirm the presence of nanosized yttria, and these results are in agreement with the results of XRD and TEM studies. Further, in Figure 2d–f, the uniform deposition of APTES/nY_2O_3 onto the ITO electrode can be seen. Interestingly, the globular morphology with uniform pattern (dendritic-like) was observed (Figure 2e–g) after immobilization of the anti-CYFRA-21-1.

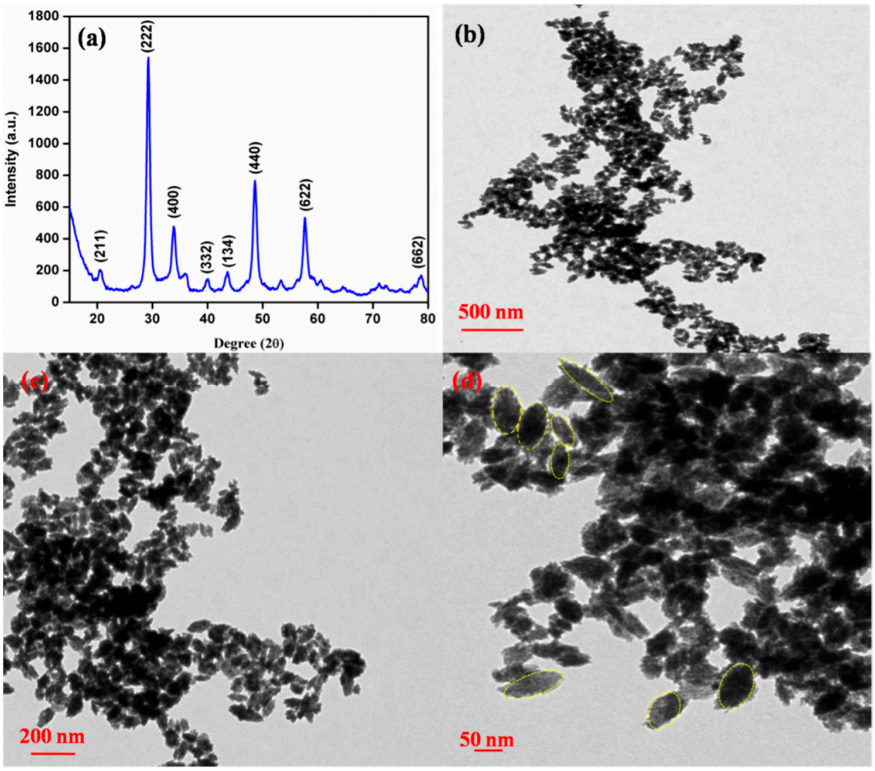


Figure 1. (**a**) XRD graph of nY_2O_3. TEM images of nY_2O_3 at scale bar of (**b**) 500 nm, (**c**) 200 nm, and (**d**) 50 nm.

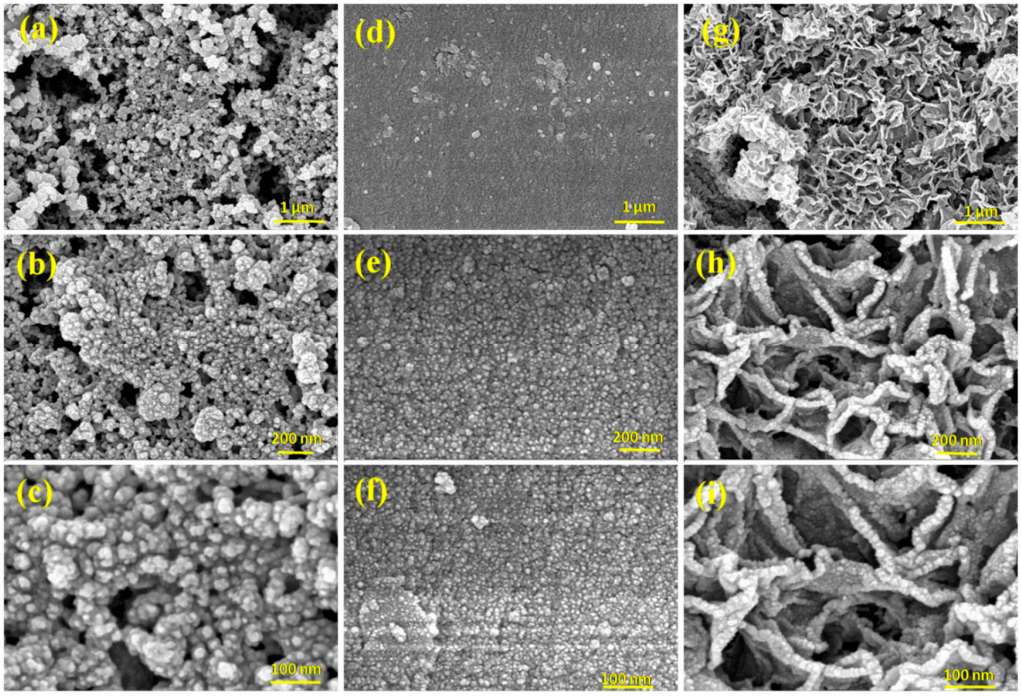


Figure 2. FE-SEM images of nY_2O_3 at (**a**) 1µm, (**b**) 200 nm, and (**c**) 100 nm; APTES/nY_2O_3/ITO at (**d**) 1µm, (**e**) 200 nm, and (**f**) 100 nm; and anti-CYFRA-21-1/APTES/nY_2O_3/ITO at (**g**) 1µm, (**h**) 200 nm, and (**i**) 100 nm.

3.2. Fourier Transform-Infrared (FTIR) Spectroscopy Study

To investigate the presence of functional groups and bonds at the surfaces of APTES/nY_2O_3/ITO and anti-CYFRA-21-1/APTES/nY_2O_3/ITO electrodes, FTIR studies were conducted and the results are shown in Figure 3a and Figure 3b, respectively. In Figure 3a, the band present at 1340 cm^{-1} is due to the presenceof C–N, and 1530 cm^{-1} is assigned to the presence of N–H (primary and secondary) bond present in amine molecules of the APTES-functionalized nY_2O_3[44]. The bands present between 3000 to 4000 cm^{-1} are attributed to the N–H stretching of amine, C–H molecules present in APTES, and O–H molecules present in PBS[44]. These results confirm the presence of amine ($–NH_2$) group onto the APTES/nY_2O_3/ITO electrode. Further, we investigated the formation of amide bond after the immobilization of anti-CYFRA-21-1 onto APTES/nY_2O_3/ITO electrode (Figure 3b). In Figure 3b, the presence of band at 1260 cm^{-1} confirms the formation of C–N (amide) bond between −COOH group present at the F_c region of the anti-CYFRA-21-1 and $-NH_2$ groups of APTES/nY_2O_3/ITO electrode[44]. The bands seen at 1560 $cm^{-1}$and 1632 cm^{-1} indicate the presence of N–H (primary and secondary) bending and C=O groups of amides. Bands found at 1055cm^{-1} and 1095 cm^{-1} are attributed to the C–O bands of carboxylic acid group, and bands present at 3350 cm^{-1} are assigned to the presence of N–H band of amines and O-H groups of PBS[44]. The FTIR results confirm the covalent immobilization (amide bond formation) of anti-CYFRA-21-1 onto the APTES/nY_2O_3/ITO electrode.

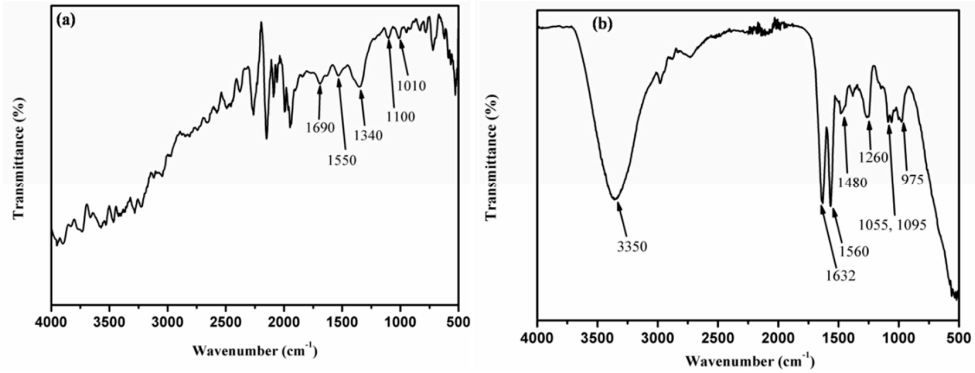


Figure 3. FTIR data of (**a**) APTES/nY_2O_3/ITO and (**b**) anti-CYFRA-21-1/APTES/nY_2O_3/ITO electrode.

3.3. Biocompatibility Studies

The biocompatibility studies of nanomaterials are known to provide an insight towards the method of immobilization and facilitate the selection of nanomaterial for development of implantable chip [15]. Besides this, biocompatible nanomaterials can be helpful in providing a favorable environment for immobilized biomolecules. We performed the biocompatibility study through MTT assay and the results are shown in Figure 4 and Table S1. The cell viability of HEK 293 cells was more than 95% upto 50 µg mL^{-1}concentration of nY_2O_3, after which the cell viability was found to be decreased. These results indicated that the synthesized nanomaterial (nY_2O_3) was favorable for the immobilization of biomolecules to be utilized for the fabrication of biosensors.

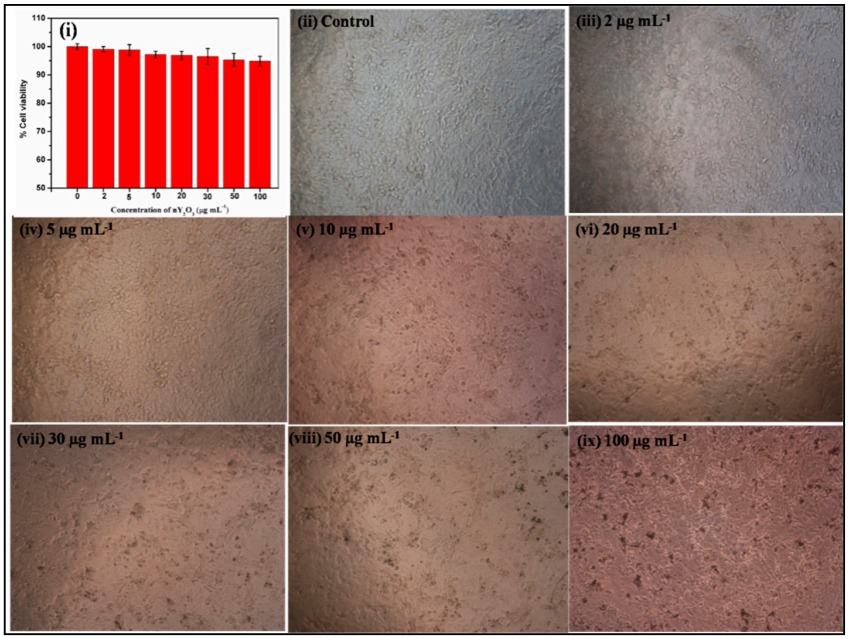


Figure 4. (**i**) Percentage of cell viability and (**ii–ix**) morphological view of HEK 293 cell line with various concentrations of nY_2O_3.

3.4. Electrochemical Characterization

Electrochemical investigations were carried out via EIS and CV studies using a redox mediator $[Fe(CN)_6]^{3-/4-}$(5 mM) in PBS. The effect of pH on the BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO immunoelectrode was conducted via CV. The maximum activity of the immunoelectrode was observed at pH 7.0 [$\{Fe(CN)_6\}^{3-/4-}$(5 mM) in PBS buffer]. In both acidic and basic pH, the peak current

decreased (Figure 5i). This result was ascribed to the maximum activity of anti-CYFRA-21-1 at pH 7.0 of PBS. In both acidic and basic pH, the antibodies became denatured, resulting in a decrease of the peak current[14,45].

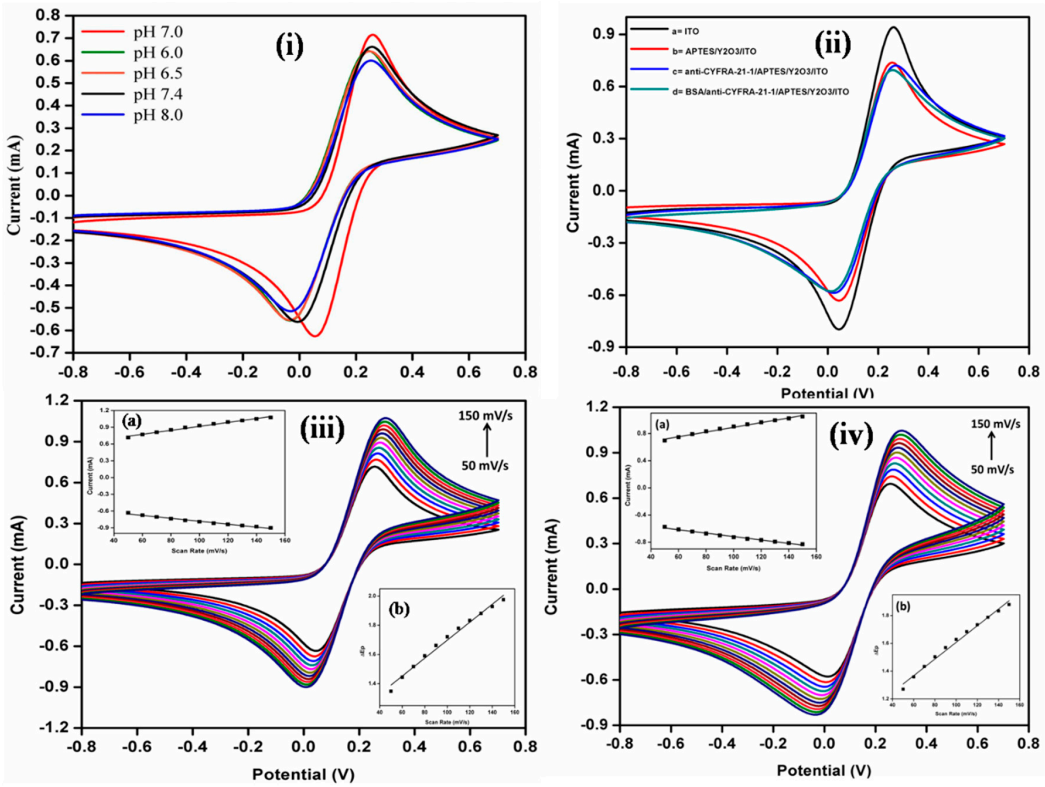


Figure 5. (**i**) CV studies of the BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO immunoelectrode on different pH of $[Fe(CN)_6]^{3-/4-}$containing PBS; (**ii**) electrochemical response of the different electrodes; (**iii**) CV response of APTES/nY_2O_3/ITO electrode; and (**iv**) the BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO immunoelectrode as a function of scan rate (50–150 mV/s). In both images, inset (**a**) magnitude of anodic (I_{pa}) and cathodic (I_{pc}) peak current as a function of scan rate (mV/s), and inset (**b**) difference potential $\Delta E_p = E_{pa} - E_{pc}$ as a function of scan rate.

The electrochemical studies of the ITO, APTES/nY_2O_3/ITO, anti-CYFRA-21-1/APTES/nY_2O_3/ITO, and BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO immunoelectrodes were investigated via CV. The peak current of the hydrolyzed ITO and APTES/nY_2O_3 electrodes was found to be ~0.9 mA and ~0.7 mA, respectively (Figure 5ii). After immobilization of the antibodies and BSA, the peak current decreased to ~0.68 mA and ~0.66 mA, respectively. This decrease in peak current was attributed to the insulating properties of the biomolecules (anti-CYFRA-21-1 and BSA)[46]. The results of scan rate studies of the APTES/nY_2O_3/ITO and BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO electrodes carried out from 50 to 150 mV/s revealed that the fabricated immunoelectrode showed uniform electrochemical behavior (Figure 5iii,iv).

The cathodic and anodic peak current increased linearly towards the higher positive and lower negative potential, respectively (Inset i, Figure 5iii,iv and Figure S1). The observed linearity with respect to the scan rate revealed that the fabricated electrodes exhibited diffusion control process, indicating that we can use scan rate in the range, 50 to 150 mV/s for the desired electrochemical studies. Equations (1)–(4) well describe the linearity obtained between cathodic and anodic peak current response with respect to scan rate[13,30].

$$I_1 = [3.53\ \mu A(s\ mV^{-1}) \times (\text{scan rate}[Mv\ s^{-1}])] + 562.91\ \mu A,\ R^2 = 0.987 \quad (1)$$

$$I_2 = -[2.62\ \mu A(s\ mV^{-1}) \times (\text{scan rate}[mV\ s^{-1}])] - 520.36\ \mu A,\ R^2 = 0.985 \quad (2)$$

$$I_3 = [3.47\ \mu A(s\ mV^{-1}) \times (\text{scan rate}[mV\ s^{-1}])] + 542.9\ \mu A,\ R^2 = 0.989 \quad (3)$$

$$I_4 = -[2.52\ \mu A(s\ mV^{-1}) \times (\text{scan rate}[mV\ s^{-1}])] - 465.64\ \mu A,\ R^2 = 0.987 \quad (4)$$

where I_1 and I_3 are the cathodic peak current of APTES/nY_2O_3/ITO and BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO; I_2 and I_4 are the anodic peak current of APTES/nY_2O_3/ITO and BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO electrode, respectively. Furthermore, it was observed that the difference between the peak potential of the cathode and anode of the APTES/nY_2O_3/ITO and BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO electrodes exhibited a linear relationship with respect to the scan rate (Inset ii, Figure 5 iii,iv) and followed Equations (5) and (6).

$$\Delta E_1 = [0.006\ V(s/mV) \times (\text{scan rate}[mV/s])] + 1.08\ V,\ R^2 = 0.987 \quad (5)$$

$$\Delta E_2 = [0.006\ V(s/mV) \times (\text{scan rate}[mV/s])] + 1.01\ V,\ R^2 = 0.988 \quad (6)$$

where ΔE_1 and ΔE_2 reveal differences in peak potential of APTES/nY_2O_3/ITO and BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO, respectively.

The diffusion coefficient "D" of the electrode was calculated via RandlesSevick equation[17,47]:

$$I_p = (2.69\times10^5)\ n^{3/2}AD^{1/2}C\upsilon^{1/2} \quad (7)$$

where I_p is the peak current of the electrode, n is the number of electrons transferred in the redox reaction (n = 1), A is the surface area of the electrode (0.25 cm^2), C is the concentration of the redox species (5×10^{-3} $mol\cdot cm^{-2}$), and υ is the scanning rate (50 $mV\cdot s^{-1}$). The "D" value at each step of electrode fabrication was determined to be 9.68×10^{-11} $cm^2\cdot s^{-1}$(APTES/nY_2O_3/ITO), 9.18×10^{-11} $cm^2\cdot s^{-1}$ (anti-CYFRA-21-1/APTES/nY_2O_3/ITO), and 8.55×10^{-11} $cm^2\cdot s^{-1}$(BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO), respectively. The diffusion coefficient of the fabricated immunoelectrode was found to be lower as compared to the value of diffusion coefficient as 1.12×10^{-3} cm^2/s [13], 2.12×10^{-3} cm^2/s [15], and 0.62×10^{-3} cm^2/s [16]) for reported immunoelectrodes, indicating that the nY_2O_3 was a more efficient nanomaterial for the fabrication of a cancer biosensor.

Quantification of antibodies onto the fabricated biosensing electrode surface are known to provide a better insight into the functioning of a biosensor, and it can be calculated by using Brown-Anson equation (Equation(8))[16].

$$I_p = n^2F^2\gamma A\upsilon(4RT)^{-1} \quad (8)$$

where I_p represents the peak current; A is the surface area of the electrode (we have taken 0.25 cm^2 active surface area for performing the electrochemical characterization and response studies); ν is the scan rate (V/s); γ is the surface concentration of the absorbed electro-active species; F is the Faraday constant; R is the gas constant; and T is room temperature.The number of antibodies immobilized onto the fabricated immunoelectrode was found to be 5.91×10^{-8} mol/cm^2, which was higher as compared to the reported biosensing platforms for CYFRA-21-1 detection [16]. Kumar et al. [13] indicating that nY_2O_3 provided better conjugation of biomolecules, leading to enhanced biosensing characteristics.

3.5. Electrochemical Impedometric Response, Control, Interferent, Reproducibility,and Shelf Life Studies

EIS is an important electrochemical technique that can be used to investigate the dynamics of antigen–antibodies interactions. We conducted EIS measurements using a three-electrode system in the frequency range from 0.01 to 10^5 Hz at a potential of 0.10V. The impedometric response obtained between fabricated immunoelectrode and CYFRA-21-1 can be understood by using Nyquist plot. In this plot, impedance can be expressed by the real part (Z') or the modules |Z| and the phase shift φ

with respect to the imaginary part (Z″) determined by the Randles circuit[48]. In the Randles circuit, R_{ct} is the charge transfer resistance, C_{dl} is capacitance of the dielectric layer, and Z_w is Warburg impedance[48]. As shown in Figure 6a, we plotted the Nyquist as a function of CYFRA-21-1 concentration. It was observed that the diameter of the semicircle increased with increasing concentration of CYFRA-21-1 (Figure 6a). We also drew the linearity graph between the R_{ct} value obtained from Randles circuit and the concentration of CYFRA-21-1 (Figure 6b). With increasing concentration, the R_{ct} value increased linearly from the log of 0.01 ng·mL^{-1} to 50 ng·mL^{-1}. This increase in R_{ct} was attributed to the insulating characteristics of theCYFRA-21-1 layer formed onto the BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO immunoelectrode[49,50]. As the number of CYFRA-21-1 molecules increased, the F_{ab} region of anti-CYFRA-21-1 present on the immunoelectrodes became saturated due to binding of the paratope of CYFRA-21-1 on the antibodies. When the F_{ab} region of the immunoelectrode was saturated, CYFRA-21-1 did not bind with the immunoelectrode, hence we did not find any change in the R_{ct} value[49,50]. The linearity graph (Equation 9) indicated that the fabricated immunoelectrode could be utilized to measure the concentration of CYFRA-21-1 from 0.01 to 50 ng·mL^{-1}.

$$I_p = [(0.23\ K\Omega \cdot mL \cdot ng^{-1}) \times \log \{\text{conc. of CYFRA-21-1 (ng mL}^{-1})\}] + 1.59,\ R^2 = 0.94 \quad (9)$$

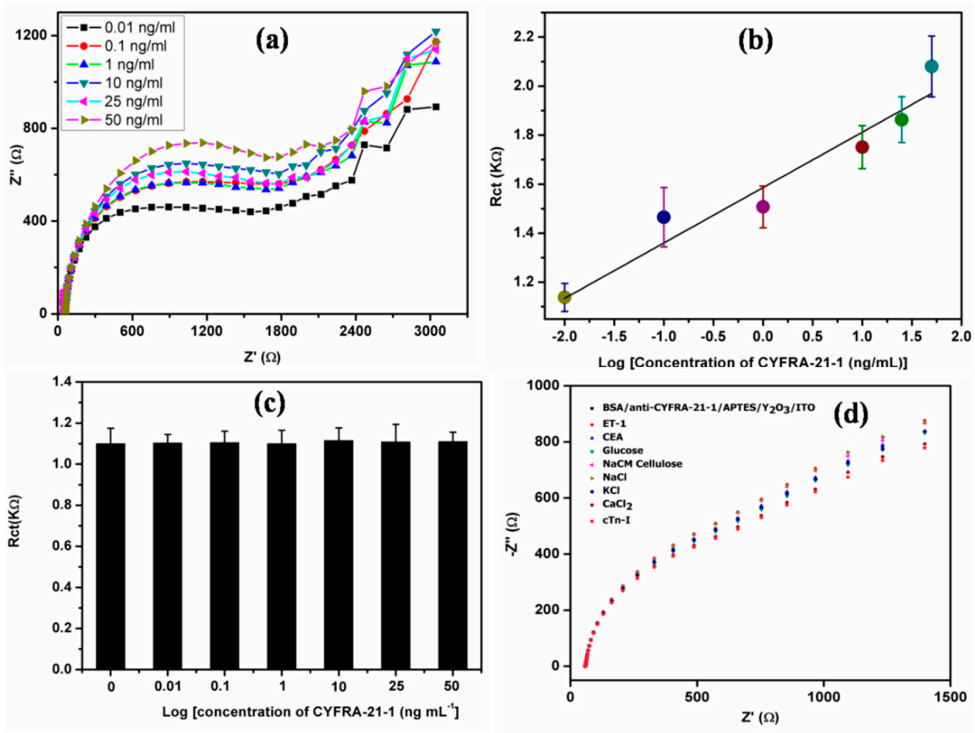


Figure 6. (**a**) Nyquist plot for electrochemical impedometric response of the BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO immunoelectrode on different concentrations of CYFRA-21-1 antigen. (**b**) Linearity graph obtained in between concentration of log [CYFRA-21-1 (ng mL^{-1})] and charge transfer resistance (R_{ct}). (**c**) Bar chart graph between different concentrations of CYFRA-21-1 and R_{ct} value obtained from electrochemical impedance response of APTES/nY_2O_3/ITO electrode with respect to various concentrations of CYFRA-21-1 by using Randles circuit. (**d**) Electrochemical impedometric response of the BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO immunoelectrode with respect to various interferents present in saliva samples of cancer patients.

A control study was conducted via EIS to investigate the interaction of theAPTES/nY_2O_3/ITO electrode with increasing CYFRA-21-1 concentration from 0.01 ng·mL^{-1} to 50 ng·mL^{-1}. The obtained R_{ct} values through Randles circuit were plotted as a function of CYFRA-21-1 concentration (Figure

6c). We did not observe any significant change in the R_{ct} value due to the APTES/nY_2O_3/ITO electrode as a function of different concentrations of CYFRA-21-1 biomolecules. The observed electrochemical response revealed that the CYFRA-21-1 biomolecules did not interact with the APTES/nY_2O_3/ITO electrode. The observed sensing characteristics were due to the interaction between antibodies immobilized onto the biosensing platform (BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO) and CYFRA-21-1 biomolecules.

We investigated the effect of various interferents present in real saliva samples of OC patients onto the EIS response of the fabricated biosensing electrode. In saliva samples, many analytes such as ET-1, CEA, glucose, NaCM cellulose, NaCl, Kcl, $CaCl_2$, cTn-I, etc. are known to be present[14]. We added these analytes into the PBS solution one-by-one, after which the EIS studies were conducted (Figure 6d). We did not find any significant changes on addition of analytes. The obtained results indicate that the fabricated immunoelectrode exhibited the EIS response towards CYFRA-21-1 biomolecules only. We checked the reproducibility of the BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO biosensing electrode. In this study, we took eight different fabricated biosensing electrodes and measured both CV and EIS response. We found that deviation in the electrochemical response. The average percentage relative standard deviation (%RSD) was 5.4% and 2.1% in the CV and EIS response, respectively. The obtained results are shown in Figure S2. The shelf life of the developed biosensing electrode was measured by EIS in ferro-ferricynide containing PBS buffer at regular intervals of one week to five weeks (Figure S3). We did not find any significant changes in the Nyquist plot, Randles circuit,orin R_{ct} value, indicating that the immunoelectrode can be safely used for upto five weeks.

3.6. Oral Cancer Patient Samples Analysis

Concentration of CYFRA-21-1 biomarker present in salivary samples of seven different OC patients was estimated using a double sandwich ELISA technique. Thirty microliters of these biological saliva samples were used to estimate the CYFRA-21-1 concentration using the biosensing electrode (BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO), and results were correlated with ELISA (Table S2). The good correlation (average %RSD = 3.92%) between CYFRA-21-1 concentration obtained via the electrochemical impedance spectroscopy and ELISA techniques revealed that the developed biosensing platform could be efficiently used for the detection of salivary CYFRA-21-1 biomarker in OC patients. We also compared the biosensing characteristics of this biosensor with those reported in literature to date (Table 1). It can be seen that the fabricated biosensor is more efficient as compared to the reported biosensors for CYFRA-21-1 detection.

Table 1. Characteristics of the BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO-based biosensor, along with those reported in literature for oral cancer detection.

Biomarker	Labeled	Detection Technique	Materials Used	Linear Detection Range	Sensitivity	Lower Detection Limit	Ref.
has-miR-200a (mi-RNA)	Yes	CV	Gold	1 aM–10 fM	-----	-----	[51]
CYFRA-21-1 (Protein)	No	CV	APTES/nZrO_2	2–16 ng mL^{-1}	2.2 µA mL ng^{-1}	0.08ng mL^{-1}	[14]
			APTES/nHfO_2	2–18 ng mL^{-1}	9.28 µA mL ng^{-1} cm^{-2}	0.21ng mL^{-1}	[16]
		DPV	APTES/nZrO_2-RGO	2–22 ng mL^{-1}	0.756 µA mL ng^{-1}	0.122ng mA^{-1}	[13]
			Serine/nZrO_2	0.01–29 ng mL^{-1}	0.295 µA mL ng^{-1}	0.01ng mL^{-1}	[15]
			Cys-$La(OH)_3$	0.001–10.2 ng mL^{-1}	12.044 µA mL ng^{-1} cm^{-2}	0.001ng mL^{-1}	[18]
			APTES/nHfO_2@RGO	0–30 ng mL^{-1}	18.24 µA mL ng^{-1}	0.164ng mL^{-1}	[17]

	EIS	APTES/nY_2O_3	0.01–50 ng mL^{-1}	226.0 Ω mL ng^{-1}	0.33 ng mL^{-1}	Present Work

4. Conclusions

The biocompatible and low dielectric constant nY_2O_3 has been synthesized via solvothermal process and has been used for the fabrication of a label-free impedometric biosensor for saliva-based cancer detection. The fabricated BSA/anti-CYFRA-21-1/APTES/Y_2O_3/ITO immunoelectrode can be used to detect salivary CYFRA-21-1 in the linear range from 0.01 to 50 ng·mL^{-1}, sensitivity of 226.0 Ω·mL·ng^{-1}, lower detection limit of 0.33 ng·mL^{-1}, and shelf life of five weeks. It should be interesting to utilize the biocompatible nY_2O_3 for the development of point-of-care devices, both for detection of communicable and non-communicable diseases.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, **Figure S1**: (a) CV response of APTES/nY_2O_3/ITO electrode as a function of scan rate (50-150 mV/s), (b) magnitude of anodic (I_{pa}) and cathodic (I_{pc}) peak current as a function of scan rate (mV/s) (c) difference potential $\Delta E_p = E_{pa}-E_{pc}$ as function of scan rate, (d) CV response of BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO electrode as a function of scan rate (50-150 mV/s), (e) magnitude of anodic (I_{pa}) and cathodic (I_{pc}) peak current as a function of scan rate (mV/s) and (f) difference potential $\Delta E_p = E_{pa}-E_{pc}$ as function of scan rate. **Figure S2:** Reproducibility studies of BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO immunoelectrode by using (a) CV and (b) EIS studies. **Figure S3:** Shelf life studies of BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO immunoelectrode via EIS studies. **Table S1:** % Cell viability of HEK 295 Cells with respect to various concentration of nY_2O_3 along with % cell viability error bar (n=3). **Table S2:** Determination of CYFRA-21-1 concentration in saliva samples of oral cancer patient using BSA/anti-CYFRA-21-1/APTES/nY_2O_3/ITO immunoelectrodes.

Author Contributions: S.K. and B.D.M conceived and designed the experiments. S.K. and S.P. performed the experiments and analyzed the data. S.P wrote the initial draft. S.K., S.A., S.K., and B.D.M took part in discussions towards the review,editing and finalisations, of the manuscript.

Funding: This research received no external funding

Acknowledgments: We thank Yogesh Singh, Vice Chancellor, Delhi Technological University (DTU), Delhi, India for providing the facilities. Suveen Kumar and Saurabh Kumar are thankful to DST for the financial support (DST/INSPIRE/04/2017/001336 and DST/INSPIRE/04/2017/002750, respectively). Bansi D. Malhotra thanks the Science & Engineering Research Board (Govt. of India) for the award of a Distinguished Fellowship (SB/DF/011/2019).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Warburg,O.Ontheoriginofcancercells. *Science* **1956**, *123*,309–314.
2. Ecsedy,J.;Hunter,D. *The Origin of Cancer*;OxfordUniversityPress:NewYork,NY,USA,2002.
3. world health organization, Available online:http://www.who.int/news-room/fact-sheets/detail/cancer, (accessedon 25/11/2018).
4. Silverman,S. *Oral Cancer*;PMPH-USA:Raleigh,NC,USA,2003;Volume1.
5. Winn,D.M.;Blot,W.J.;Shy,C.M.;Pickle,L.W.;Toledo,A.;Fraumeni,J.F., Jr.Snuffdippingandoralcanceramongwomeninthesouthernunitedstates. *New Engl. J. Med.* **1981**, *304*,745–749.
6. Patton,L.L.;Epstein,J.B.;Kerr,A.R.Adjunctivetechniquesfororalcancerexaminationandlesiondiagnosis:Asystematicreviewoftheliterature. *J. Am. Dent. Assoc.* **2008**, *139*,896–905.
7. Rosenberg,D.;Cretin,S.Useofmeta-analysistoevaluatetoloniumchlorideinoralcancerscreening. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **1989**, *67*,621–627.
8. Neville,B.W.;Day,T.A.Oralcancerandprecancerouslesions. *Ca A Cancer J. Clin.* **2002**, *52*,195–215.
9. Alevizos,I.;Mahadevappa,M.;Zhang,X.;Ohyama,H.;Kohno,Y.;Posner,M.;Gallagher,G.T.;Varvares,M.;Cohen,D.;Kim,D.Oralcancerinvivogeneexpressionprofilingassistedbylasercapturemicrodissectionandmicroarrayanalysis. *Oncogene* **2001**, *20*,6196.

10. Yu,L.Q.;Wang,L.Y.;Su,F.H.;Hao,P.Y.;Wang,H.;Lv,Y.K.Agate-openingcontrolledmetal-organicframeworkforselectivesolid-phasemicroextractionofaldehydesfromexhaledbreathoflungcancerpatients. *Microchim. Acta* **2018**, *185*,307.
11. Malhotra,R.;Urs,A.B.;Chakravarti,A.;Kumar,S.;Gupta,V.;Mahajan,B.Correlationofcyfra21-1levelsinsalivaandserumwithck19mrnaexpressioninoralsquamouscellcarcinoma. *Tumor Biol.* **2016**, *37*,9263–9271.
12. Rajkumar,K.;Ramya,R.;Nandhini,G.;Rajashree,P.;RameshKumar,A.;NirmalaAnandan,S.Salivaryandserumlevelofcyfra21–1inoralprecancerandoralsquamouscellcarcinoma. *Oral Dis.* **2015**, *21*,90–96.
13. Kumar,S.;Sharma,J.G.;Maji,S.;Malhotra,B.D.Nanostructuredzirconiadecoratedreducedgrapheneoxidebasedefficientbiosensingplatformfornon-invasiveoralcancerdetection. *Biosens. Bioelectron.* **2016**, *78*,497–504.
14. Kumar,S.;Kumar,S.;Tiwari,S.;Srivastava,S.;Srivastava,M.;Yadav,B.K.;Kumar,S.;Tran,T.T.;Dewan,A.K.;Mulchandani,A.Biofunctionalizednanostructuredzirconiaforbiomedicalapplication:Asmartapproachfororalcancerdetection. *Adv. Sci.* **2015**, *2*, 1500048.
15. Kumar,S.;Sharma,J.G.;Maji,S.;Malhotra,B.D.Abiocompatibleserinefunctionalizednanostructuredzirconiabasedbiosensingplatformfornon-invasiveoralcancerdetection. *RSC Adv.* **2016**, *6*,77037–77046.
16. Kumar,S.;Kumar,S.;Tiwari,S.;Augustine,S.;Srivastava,S.;Yadav,B.K.;Malhotra,B.D.Highlysensitiveproteinfunctionalizednanostructuredhafniumoxidebasedbiosensingplatformfornon-invasiveoralcancerdetection. *Sens. Actuators B Chem.* **2016**, *235*,1–10.
17. Kumar,S.;Kumar,S.;Augustine,S.;Yadav,S.;Yadav,B.K.;Chauhan,R.P.;Dewan,A.K.;Malhotra,B.D.Effectofbrownianmotiononreducedagglomerationofnanostructuredmetaloxidetowardsdevelopmentofefficientcancerbiosensor. *Biosens. Bioelectron.* **2018**, *102*,247–255.
18. Tiwari,S.;Gupta,P.K.;Bagbi,Y.;Sarkar,T.;Solanki,P.R.L-cysteinecappedlanthanumhydroxidenanostructuresfornon-invasivedetectionoforalcancerbiomarker. *Biosens. Bioelectron.* **2016**, *89*, 1042–1052.
19. Solanki,P.R.;Kaushik,A.;Agrawal,V.V.;Malhotra,B.D.Nanostructuredmetaloxide-basedbiosensors. *Npg Asia Mater.* **2011**, *3*,17–24.
20. Tan,C.;Cao,X.;Wu,X.J.;He,Q.;Yang,J.;Zhang,X.;Chen,J.;Zhao,W.;Han,S.;Nam,G.H.Recentadvancesinultrathintwo-dimensionalnanomaterials. *Chem. Rev.* **2017**, *117*,6225–6331.
21. Kim,J.;Park,S.J.;Min,D.H.Emergingapproachesforgrapheneoxidebiosensor. *Anal. Chem.* **2016**, *89*,232–248.
22. Thapa,A.;Soares,A.C.;Soares,J.C.;Awan,I.T.;Volpati,D.;Melendez,M.E.;Fregnani,J.H.T.G.;Carvalho,A.L.;OliveiraJr,O.N.Carbonnanotubematrixforhighlysensitivebiosensorstodetectpancreaticcancerbiomarkerca19-9. *ACS Appl. Mater. Interfaces* **2017**, *9*,25878–25886.
23. Othman,A.;Karimi,A.;Andreescu,S.Functionalnanostructuresforenzymebasedbiosensors:Properties,fabricationandapplications. *J. Mater. Chem. B* **2016**, *4*,7178–7203.
24. Zhao,J.;Qin,L.;Hao,Y.;Guo,Q.;Mu,F.;Yan,Z.Applicationoftubulartetrapodmagnesiumoxideinabiosensorforhydrogenperoxide. *Microchim. Acta* **2012**, *178*,439–445.
25. Shi,X.;Gu,W.;Li,B.;Chen,N.;Zhao,K.;Xian,Y.Enzymaticbiosensorsbasedontheuseofmetaloxidenanoparticles. *Microchim. Acta* **2014**, *181*,1–22.
26. Xu,J.;Wang,Y.;Hu,S.Nanocompositesofgrapheneandgrapheneoxides:Synthesis,molecularfunctionalizationandapplicationinelectrochemicalsensorsandbiosensors.Areview. *Microchim. Acta* **2017**, *184*,1–44.
27. Kumar,S.;Ahlawat,W.;Kumar,R.;Dilbaghi,N.Graphene,carbonnanotubes,zincoxideandgoldaselitenanomaterialsforfabricationofbiosensorsforhealthcare. *Biosens. Bioelectron.* **2015**, *70*,498–503.
28. Hasanzadeh,M.;Shadjou,N.;delaGuardia,M.Ironandiron-oxidemagneticnanoparticlesassignal-amplificationelementsinelectrochemicalbiosensing. *Trac. Trends Anal. Chem.* **2015**, *72*,1–9.
29. Luo,X.;Morrin,A.;Killard,A.J.;Smyth,M.R.Applicationofnanoparticlesinelectrochemicalsensorsandbiosensors. *Electroanalysis* **2006**, *18*,319–326.
30. Vasudev,A.;Kaushik,A.;Bhansali,S.Electrochemicalimmunosensorforlabelfreeepidermalgrowthfactorreceptor(egfr)detection. *Biosens. Bioelectron.* **2013**, *39*,300–305.
31. Campuzano,S.;Pedrero,M.;Nikoleli,G.P.;Pingarrón,J.;Nikolelis,D.Hybrid2d-nanomaterials-basedelectrochemicalimmunosensingstrategiesforclinicalbiomarkersdetermination. *Biosens. Bioelectron.* **2017**, *89*,269–279.
32. Solanki,S.;Pandey,C.M.;Soni,A.;Sumana,G.;Biradar,A.M.Anamperometricbienzymaticbiosensorforthetriglyceridetributyrinusinganindiumtinoxideelectrodecoatedwithelectrophoreticallydepositedchitosan-wrappednanozirconia. *Microchim. Acta* **2016**, *183*,167–176.
33. Casero,E.;Alonso,C.;Petit-Domínguez,M.D.;Vázquez,L.;Parra-Alfambra,A.M.;Merino,P.;Álvarez-García,S.;deAndrés,A.;Suárez,E.;Pariente,F.Lactatebiosensorbasedonabionanocompositecomposedoftitaniumoxidenanoparticles,photocatalyticallyreducedgraphene,andlactateoxidase. *Microchim. Acta* **2014**, *181*,79–87.

34. Huang,G.;Zhanglian,H.;Shizhu,Z.;Pengyue,Z.;Xianping,F.Synthesisofyttriumoxidenanocrystalviasolvothermalprocess. *J. Rare Earths* **2006**, *24*,47–50.
35. Li,N.;Yanagisawa,K.Yttriumoxidenanowires.In *Nanowires Science and Technology*;InTech:London, UK, 2010.
36. Tang,Q.;Liu,Z.;Li,S.;Zhang,S.;Liu,X.;Qian,Y.Synthesisofyttriumhydroxideandoxidenanotubes. *J. Cryst. Growth* **2003**, *259*,208–214.
37. Curtis,C.Propertiesofyttriumoxideceramics. *J. Am. Ceram. Soc.* **1957**, *40*,274–278.
38. Razavi,R.S.;Ahsanzadeh-Vadeqani,M.;Barekat,M.;Naderi,M.;Hashemi,S.H.;Mishra,A.K.Effectofsinteringtemperatureonmicrostructuralandopticalpropertiesoftransparentyttriaceramicsfabricatedbysparkplasmasintering. *Ceram. Int.* **2016**, *42*,7819–7823.
39. Mori,T.;Lee,J.H.;Li,J.;Ikegami,T.;Auchterlonie,G.;Drennan,J.Improvementoftheelectrolyticpropertiesofy2o3-basedmaterialsusingacrystallographicindex. *Solid State Ion.* **2001**, *138*,277–291.
40. Rasheed,P.A.;Radhakrishnan,T.;Shihabudeen,P.;Sandhyarani,N.Reducedgrapheneoxide-yttriananocompositemodifiedelectrodeforenhancingthesensitivityofelectrochemicalgenosensor. *Biosens. Bioelectron.* **2016**, *83*,361–367.
41. Korotcenkov,G.Nanocompositesinelectrochemicalsensors.In *Handbook of Gas Sensor Materials*;Springer:Berlin/Heidelberg,Germany,2014;pp.223–235.
42. Tsutsumi,T.Dielectricpropertiesofy2o3thinfilmspreparedbyvacuumevaporation. *Jpn. J. Appl. Phys.* **1970**, *9*,735.
43. Dutta,P.K.;Szabo,N.F.;Du,H.;Akbar,S.A.Potentiometricnoxsensorsbasedonyttria-stabilizedzirconiawithzeolitemodifiedelectrode.InGooglePatentsNo.US6843900B2,18January2005.
44. Pavia,D.L.;Lampman,G.M.;Kriz,G.S.;Vyvyan,J.A. *Introduction to Spectroscopy*;CengageLearning:Boston, MA, USA, 2008.
45. Liu,G.;Lin,Y.Electrochemicalsensorfororganophosphatepesticidesandnerveagentsusingzirconiananoparticlesasselectivesorbents. *Anal. Chem.* **2005**, *77*,5894–5901.
46. Kumar,S.;Kumar,S.;Augustine,S.;Malhotra,B.D.Proteinfunctionalizednanostructuredzirconiabasedelectrochemicalimmunosensorforcardiactroponinidetection. *J. Mater. Res.* **2017**, *32*,2966–2972.
47. Xu,F.;Deng,M.;Li,G.;Chen,S.;Wang,L.Electrochemicalbehaviorofcuprousoxide–reducedgrapheneoxidenanocompositesandtheirapplicationinnonenzymatichydrogenperoxidesensing. *Electrochim. Acta* **2013**, *88*,59–65.
48. Orazem,M.E.;Tribollet,B. *Electrochemical Impedance Spectroscopy*;JohnWiley&Sons:Hoboken,NJ,USA,2011;Volume48.
49. Srivastava,S.;Kumar,V.;Ali,M.A.;Solanki,P.R.;Srivastava,A.;Sumana,G.;Saxena,P.S.;Joshi,A.G.;Malhotra,B. Electrophoreticallydepositedreducedgrapheneoxideplatformforfoodtoxindetection. *Nanoscale* **2013**, *5*,3043–3051.
50. Ali,M.A.;Singh,C.;Mondal,K.;Srivastava,S.;Sharma,A.;Malhotra,B.D.Mesoporousfew-layergrapheneplatformforaffinitybiosensingapplication. *ACS Appl. Mater. Interfaces* **2016**, *8*,7646–7656.
51. Wang,Z.;Zhang,J.;Guo,Y.;Wu,X.;Yang,W.;Xu,L.;Chen,J.;Fu,F.Anovelelectricallymagnetic-controllableelectrochemicalbiosensorfortheultrasensitiveandspecificdetectionofattomolarleveloralcancer-relatedmicrorna. *Biosens. Bioelectron.* **2013**, *45*,108–113.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

AIP Conference Proceedings

Controlling the radiation pattern of a microstrip patch antenna using a checkerboard patterned metasurface

Cite as: AIP Conference Proceedings **2136**, 030003 (2019); https://doi.org/10.1063/1.5120911
Published Online: 09 August 2019

Nishant Shankhwar, Ritika Ranga, Yogita Kalra, and Ravindra Kumar Sinha


View Online

Export Citation



© 2019 Author(s).

Controlling the radiation pattern of a microstrip patch antenna using a checkerboard patterned metasurface

Nishant Shankhwar[1], Ritika Ranga[1], Yogita Kalra[1,a)] and Ravindra Kumar Sinha[1,2]

[1]*TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication, Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi-110042, India*
[2]*CSIR-Central Scientific Instruments Organisation (CSIO), Chandigarh-160030, India.*

[a)]dryogitakalra@gmail.com

Abstract. In this work, radiation manipulation capabiltites of patterned metasurfaces has been discussed. It has been shown that the radiation pattern can be modified as desired by using a metasurface patch, instead of a simple patch. The type of modification achived depends upon the type of pattern used. In this article a checkerboard type metasurface patch has been employed which generates four beam of radition as compared to a single beam in case of the conventional microstrip patch antenna (MPA). Specifically designed metasurfaces can not only facilitate the splitting of radiation into desired number of beams but also provide control over the orientation of the beams and direction of radiation by suppressing one particular lobe in favour of another.

INTRODUCTION

Metasurfaces are planar metamaterials spread in two dimensions. They are developed by patterning of thin films. Compared to bulk metamaterials [1, 2, 3, 4], metasurfaces exhibit greater compatibility to on-chip or substrate based devices. So far, metasurfaces and metascreen have found a broad spectrum of applications ranging from perfect reflection to wavefront enginnering, optical switching, frequency selective filtering, cavity resonators, superlensing, etc [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. Metasurfaces have also been employed to a great extent, in enhancing the performance of antennas [17, 18, 19, 20, 21, 22, 23]. In this article, beam manipulation using metasurfaces has been discussed. Is has been shown that, by modification of the radiating element of the antenna one can attain control over its radiation pattern. This idea has been verified by means of a microstrip patch antenna (MPA) equipped with a checkerboard style metasurface. The results of the numerical computations for the proposed design have been compared with those for a conventional patch antenna.

DESIGN

The proposed antenna structure consist of metal (Cu) patches printed in a checkerboard like pattern on a dielectric substrate of permittivity ($\varepsilon = 3.38$). It is a 5×5 grid in which size of each square is 10 mm. Size of the substrate is 90 mm and thickness is 0.5 mm. Total span of the checkerboard pattern is 50 mm $\times$ 50 mm. It also has uniform metal coating at its bottom to act as a ground plane. For comparative purposes, a simple MPA having similar geometry has also been designed which has a uniform metallic patch of size 50 mm $\times$ 50 mm. The schematic illustration of both the designs is shown in figure 1.

NUMERICAL ANALYSIS

The structure has been analysed in frequency domain using finite element method based commercial software package COMSOL Multiphsics. The electric field has been calculated at the surface of the antenna, as well as, in the far field zone for both the designs. The devices have been designed to operate in microwave region of spectrum. The results of the computation have been discussed in the next section.

International Conference on Photonics, Metamaterials & Plasmonics
AIP Conf. Proc. 2136, 030003-1–030003-4; https://doi.org/10.1063/1.5120911
Published by AIP Publishing. 978-0-7354-1879-0/$30.00

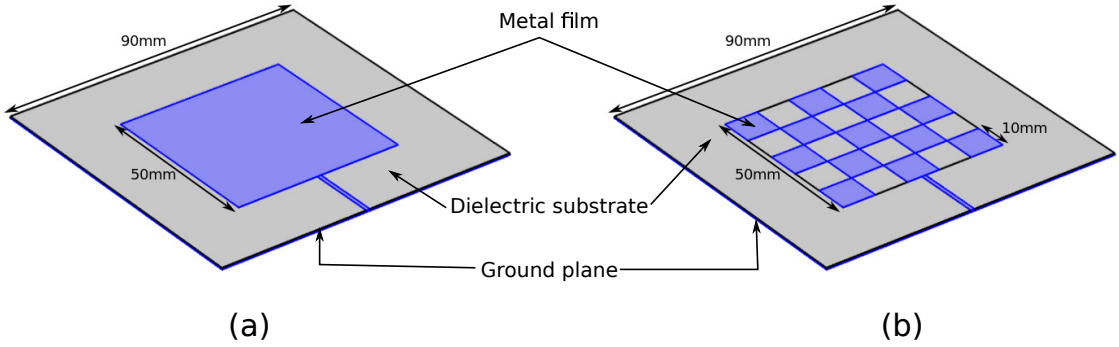


FIGURE 1. (a) Conventional MPA, (b) Checkerboard MPA. Blue region indicates metal film.

RESULTS

Figure 2 shows the radiation pattern obtained in the two cases and it can be obeserved that patterning has drastically modified the radiation profile. Compared to a single lobe of radiation in case of conventional MPA, four symmetrically oriented lobes are obtained for the proposed design, indicating flow of power in four different directions. It should be noted that pattering results in change in resonant frequency of the antenna. The resonant frequency for the checkerboard MPA is 5 times greater than simple MPA, which seems obvious, as the patterned design comprises a 5×5 grid in which the size of each sqaure patch is 1/5th of that of the simple MPA. Hence, while the resonant frequency of the simple MPA is 1.63 GHz, it is 8.16 GHz for the proposed antenna.

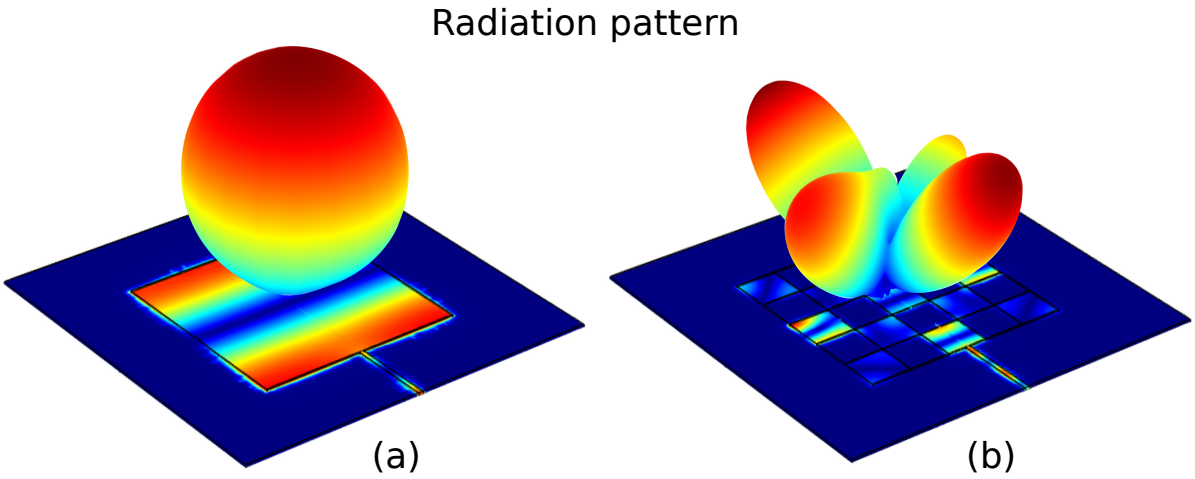


FIGURE 2. Radiation pattern for (a) Conventional MPA, (b) Checkerboard MPA

Figure 3 shows the distribution of z component of electric field on the patch (radiating element) in both the designs. One can notice that, the mode field profile in case of proposed design is similar to that of the conventional design, but it is confined in 1/5 times smaller size patches, resulting in 5 times increase in the resonant frequency.

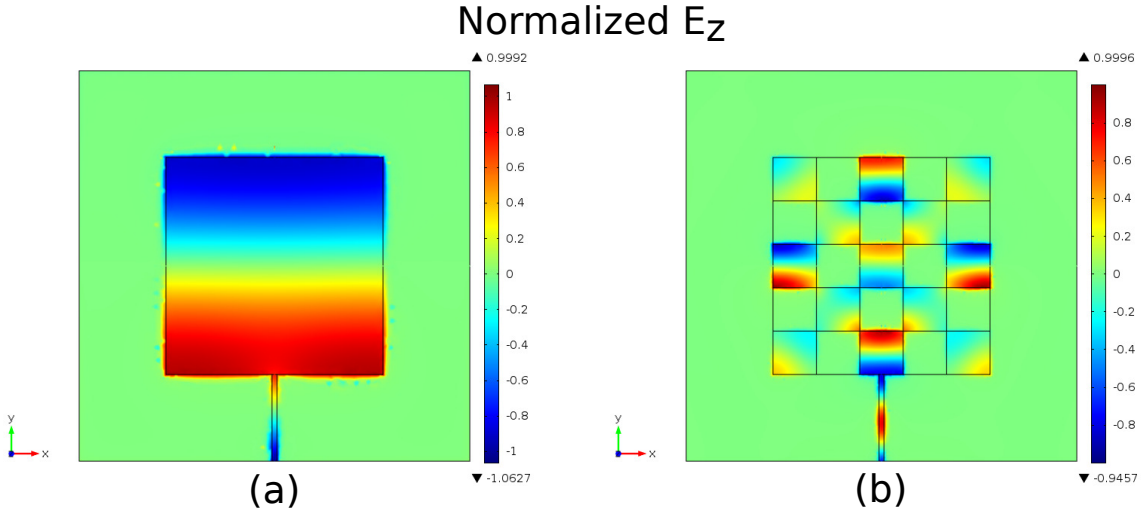


FIGURE 3. Distribution of z component of electric field for (a) Conventional MPA, (b) Checkerboard MPA

CONCLUSION

Capability of checkerboard metasurface in beam manipulation has been succesfully demonstrated using a microwave patch antenna. The MPAs are lightweight and their fabrication is simple and low-cost. The metasurfaces are inherently compatible with printed antennas due to similarity in design and fabrication process.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Initiatives and support for the establishment of TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication at Delhi Technological University (Formerly Delhi College of Engineering) Delhi, through the Mission REACH program of Technology Vision-2020 of the Government of India. Nishant Shankhwar acknowledges the support through Junior Research Fellowship awarded by University Grants Commission (India) under UGC-NET-JRF scheme.

REFERENCES

[1] C. L. Holloway, E. F. Kuester, J. A. Gordon, J. O'Hara, J. Booth, and D. R. Smith, IEEE Antennas and Propagation Magazine **54**, 10–35 (2012).
[2] H.-T. Chen, A. J. Taylor, and N. Yu, Reports on Progress in Physics **79**, p. 076401 (2016).
[3] H.-H. Hsiao, C. H. Chu, and D. P. Tsai, Small Methods **1**, p. 1600064 (2017).
[4] N. Meinzer, W. L. Barnes, and I. R. Hooper, Nature Photonics **8**, p. 889 (2014).
[5] N. Yu and F. Capasso, Journal of Lightwave Technology **33**, 2344–2358 (2015).
[6] M. Decker, I. Staude, M. Falkner, J. Dominguez, D. N. Neshev, I. Brener, T. Pertsch, and Y. S. Kivshar, Advanced Optical Materials **3**, 813–820 (2015).
[7] N. Shankhwar, R. K. Sinha, Y. Kalra, S. Makarov, A. Krasnok, and P. Belov, Photonics and Nanostructures-Fundamentals and Applications **24**, 18–23 (2017).
[8] L. Xu, C. A. Curwen, P. W. Hon, Q.-S. Chen, T. Itoh, and B. S. Williams, Applied Physics Letters **107**, p. 221105 (2015).
[9] F. Aieta, P. Genevet, N. Yu, M. A. Kats, Z. Gaburro, and F. Capasso, Nano letters **12**, 1702–1706 (2012).
[10] F. Aieta, P. Genevet, M. A. Kats, N. Yu, R. Blanchard, Z. Gaburro, and F. Capasso, Nano letters **12**, 4932–4936 (2012).
[11] Y. F. Yu, A. Y. Zhu, R. Paniagua-Domínguez, Y. H. Fu, B. Luk'yanchuk, and A. I. Kuznetsov, Laser & Photonics Reviews **9**, 412–418 (2015).

[12] A. V. Kildishev, A. Boltasseva, and V. M. Shalaev, Science **339**, p. 1232009 (2013).
[13] B. Gholipour, J. Zhang, K. F. MacDonald, D. W. Hewak, and N. I. Zheludev, Advanced materials **25**, 3050–3054 (2013).
[14] Y. Yao, R. Shankar, M. A. Kats, Y. Song, J. Kong, M. Loncar, and F. Capasso, Nano letters **14**, 6526–6532 (2014).
[15] M. Ghasemi and P. Choudhury, Optik-International Journal for Light and Electron Optics **127**, 9932–9936 (2016).
[16] J. Ortiz, J. Baena, V. Losada, F. Medina, R. Marques, and J. A. Quijano, IEEE Microwave and Wireless Components Letters **23**, 291–293 (2013).
[17] J. D. Kraus, R. J. Marhefka, and A. S. Khan, *Antennas and wave propagation* (Tata McGraw-Hill Education, 2006).
[18] G. Minatti, M. Faenzi, E. Martini, F. Caminita, P. De Vita, D. González-Ovejero, M. Sabbadini, and S. Maci, IEEE Trans. Antennas Propag **63**, 1288–1300 (2015).
[19] A. Epstein, J. P. Wong, and G. V. Eleftheriades, Nature communications **7**, p. 10360 (2016).
[20] M. Faenzi, F. Caminita, E. Martini, P. De Vita, G. Minatti, M. Sabbadini, and S. Maci, IEEE Antennas Wireless Propag. Lett. **15**, 610–613 (2016).
[21] G. Minatti, F. Caminita, E. Martini, M. Sabbadini, and S. Maci, IEEE Transactions on Antennas and Propagation **64**, 3907–3919 (2016).
[22] F. H. Lin and Z. N. Chen, IEEE Transactions on Antennas and Propagation **65**, 1706–1713 (2017).
[23] D. González-Ovejero, G. Minatti, G. Chattopadhyay, and S. Maci, IEEE Transactions on Antennas and Propagation **65**, 2923–2930 (2017).

Proceedings of the 4th World Congress on Civil, Structural, and Environmental Engineering (CSEE'19)
Rome, Italy – April, 2019
Paper No. ICGRE 130
DOI: 10.11159/icgre19.130

Deformation Behaviour of Tunnels under Different Loading Conditions: State of Art

Parvesh Kumar[1], Amit Kumar Shrivastava[1]
[1]Civil Engineering Department, Delhi Technological University
Bhawana Road, Delhi, India
parveshkumar_2k16phdce06@dtu.ac.in; aksrivastava@dce.ac.in

Abstract- Construction of tunnels as transport medium has become a very important part in various civil engineering practises. Tunnels are constructed and commonly being used for different purposes such as for underground mining, underground parking facilities, military purposes, underground gas pipelines, transportation purposes in metro cities and many more fields. In the present study, the deformation behaviour of tunnels under various loading conditions that may be static load acting on the tunnels due to the overburden pressure, dynamic load due to the earthquakes and impact load causes due to terrorist activities is studied. Various factors which may be responsible for causing deformation cracks in tunnels are being studied. Various theories on stress distribution in tunnels are given by a lot of researchers were also studied.

***Keywords*:** Deformation, Underground Structures, Cover depth, In-situ stress.

1. Introduction

Tunnels are classified as underground artificial passages which are constructed without making any disturbance in the ground surface. Tunnels are generally used for various purposes in our day to day life. In4000 B.C underground caves were made by human being for their shelter. Therefore, contraction of tunnels and underground passages is very old activity which is commonly performed by human being for their different purposes. Tunnels can be constructed through rock masses, hills, rivers etc. In the present day, tunnels are used for various purposes. Apart from this, tunnels can be used for different application work which is related to various civil and military fields. Tunnels are used as a means of transportation in hilly areas where modes of transportations are very less or negligible in some places. Underground structures are used for construction of railways lines, roads etc. Also, they are used for various military practices. In difficult conditions, such as soft ground, large jointed rocks, massive hills, underground structures are very effective. When we talk about the underground structures it is very important to understand the distributions of various stresses on tunnels. There are various types of loads which are subjected on the tunnels structures. Due to these loads subjected over tunnels, deformation occurs which further effects the stability of the structures. Generally, the underground structures are subjected to static load, dynamic loads and impact loads. Static load is caused due to the overburden pressure acting on the structure whereas dynamic and impact load is caused in earthquake and blast conditions. Stability of underground structures is very important aspect which should be kept in mind while designing the structure. In unstable conditions, the structure may deform in a very short span of time. When the tunnels are subjected to static loading conditions, the maximum load occurs at the roof and the side walls of the tunnels structures. If the structure is not reinforced properly then the structure may subjected to tensile and compression failure. Various theories were given by a lot of researchers to study the stress distribution in both single and twin tunnels. From these theories it may be noted that the tunnel structures are basically subjected to three types of stresses i.e. radial stress, tangential stress and shear stress. Different equations of stresses are derived by various researchers which are discussed briefly in this paper.

2. Deformations of tunnels under different conditions

The influence of the various factors on the deformation behavior of tunnels has been studied by many researchers. From the review it can be noticed that there are a lot of factors which affect the stability of underground structures. Some of them are summarized here. Stability of tunnels mainly depends upon the relative position of tunnel and its construction procedure. Chehade and Shahrour, (2008) studied the interaction between the twin tunnels with the help of numerical software. The

studied was done to know how the position of the twin tunnels and their construction procedure may affect the settlement behavior of the soil structures. For the study, three different alignments of twin tunnels were selected i.e. horizontally, vertically and inclined. From the results, it can be concluded that when the upper tunnel is constructed first, it may lead to higher settlement as compare to that obtained by construction of lower tunnel at first. It was also noted that when the tunnels are vertically aligned, they may cause the highest settlement of soil. Idris et al., (2008), conducted a study to determine the deformation behavior of ancient tunnels structure by using UDEC. The studied was done to determine the effect of various factors such as masonry block cohesion, tensile strength and friction angle. From the study, it was notified that there is a significant effect of cohesion and friction angle on the masonry behavior. Jia and Tang, (2008) conducted a study on the failure mechanism of tunnel in jointed rock mass using numerical code RFPA. The study deals with the influence of different dip angles of layered joint on the stability of the tunnels. The study indicates that both the dip angle and lateral pressure have significant impact on the failure mode. As the dip angle of the joint increases the displacement of the sidewall also increases. Nunes and Meguid, (2009) conducted an experimental study to determine the effect of stresses caused due to the overlying soil on the stability behavior of tunnel. Experimental tests were performed to know the influence of the overlying layer on tunnels deformation behavior. Effect of stiffness ratio and effect of layer thickness is also studied. Numerical investigation was also done by using Plaxis V8 software to compare the results obtained from experimental tests. From results, it can be notified that there is a significant effect on the stresses acting on tunnel due to the presence of the soil layer above the tunnel. It was concluded that if the thickness and location of the soil layer which is present above the tunnel changes, then the stress distribution may also vary. Idris et al., (2009) conducted a study on the numerical modeling of masonry joints degradation in built tunnels. In this study an attempt is made to study the influence of the mechanical properties of the masonry joints on the tunnel behavior by the help of experimental design method distinct elements numerical modeling. The factors selected for this study were masonry joint, cohesion, joint tensile strength and joint friction angle. It was concluded that the joint cohesion is the most important factor followed by joint tensile strength and finally the joint friction angle. Elshamy et al., (2013) conducted a study to determine the deformation behavior of different shapes of twin tunnels in soft clay. A numerical model is constructed as a plane strain problem taking into consideration the soil-structure interaction. The study was done to examine the effect of thickness-radius ratio on the tunnel deformation behavior, forces acting on the tunnel and the stress distribution. The settlement of the surface was noticed for various shapes and dimension of different twin tunnels. Three different tunnel shapes (circular, elliptical and rectangle with circular arch) were selected for this study. From the results, it was concluded that the circular shape tunnel is the best twin tunnel as vertical displacement and internal forces acting on the tunnel structure were less in circular tunnels as compare to the other two shapes of twin tunnels. It was also noticed that there is an increase in the internal forces up to 40%, when the diameter of the tunnel varies from 8 to 12m. Also, there is an increase in soil stresses and surface displacement when the diameter is increased from 8 to 12 meters.

Zhu et al., (2015) made an attempt to study the effect of principle stress orientation on the stability of regular tunnel and crack tunnel. For that, three typical tunnel models were prepared and the materials used for tunnel model formation were cement mortar and sandstone. The first one was a regular tunnel model loaded by various orientation principle stresses. The second one was a tunnel with various inclination cracks. And the third one is a tunnel with a fixed angle crack loaded by various orientation principle stresses. A 5cm radial crack with different inclination angles β at tunnel spandrels is made. The results show that the critical compressive stress varies with the angle of inclination of the crack. When β increases from 0° to 135°, the critical stress decreases and when β is about $135^{\circ}$the critical stress is lowest. As an increase in the angle of inclination β from 135° to 180°, the critical stress increases. Khan et al., (2015) conducted a study to analyze the interactive behavior between tunnel lining and rock with the variation of stresses and deformation. In this study different tunnels section i.e. circular; semicircular and horse shoe were considered in both fractured and unfractured rock. Stress and deformation are analyzed in all the three section and the most suitable section is determined. Numerical modeling using FEM (Abacus Software) is done for modeling and meshing of tunnel in 2D. From results it was concluded that for unfractured rock the most suitable section from vertical stress point of view is horse shoe and from deformation point of view is semicircular section. For fractured rock the most suitable section from both points of view is semicircular section. Mishra et al., (2016) conducted a study to determine the deformation behavior of shallow tunnels under static and dynamic loading. For this study ABACUS is used as modeling software which is based on FEM numerical tool. In this study natural soft rock conditions are

simulated in laboratory by using Plaster of Paris and mix material as tunnel model material. Then the results which are obtained from numerical modeling are then compared with experimental results. From the results, it was noticed that in case of static loading the length of the crack is large as compare to the crack obtained in case of dynamic loading. It was noticed that the zone is totally crushed in case of dynamic loading. Tiwari et al., (2016) conducted a study on analyze the behavior of twin tunnel under dynamic conditions. The twin tunnel consists of tunnel reinforced concrete lining which is subjected to internal blast loading. Numerical modeling is done with the help of ABACUS. Pressure acting in the Reinforced concrete lining and the surrounding soil mass, which is caused by the explosive induced shock wave, was determined for both tunnels. And from the results, it was concluded that the damage and deformation of reinforced concrete lining and soil mass depend upon charge weight and the clearance between the tunnels. Ahmad and Bahador, (2016) conducted a study to analyze the effect of different factors on settlement behavior of ground due to excavation of twin tunnels. Comparison of Flac2D and Flac3D modeling was done. The results concluded that the 3D model show a more realistic profile than 2D model. The maximum settlement was occurred along the center line of the tunnel at ground level and the highest rate of displacement in the model occurred in the crown of the tunnel. It was also observed that the impact area of tunnel excavation is extended up to 4 times the diameter from the center of the tunnel after that there is no change. It was also noticed that there was larger displacement in case of simultaneous excavation this may be due to the interaction between the tunnel faces. Kumar and Shrivastava, (2017) reviewed the various factors which effect the stability of underground structures. It was concluded that the stability of tunnels depends upon many factors such as types of tunnel section, crack inclination angle and stress caused on tunnels due to overlying strata. Shalabi, (2017) conducted a study on the interaction behavior of shallow twin tunnels of circular shape by using FE analysis. The study was conducted to determine the effect on the stability behavior of tunnel when the spacing between the tunnels varies. From the results it was noticed that deformation effect will be more when the spacing between the tunnels is less. Mishra, (2017) conducted a study to determine the effects of static and cyclic loading conditions on lined and unlined tunnels samples. From the results it was concluded that in case of unlined tunnels the number of cycles required to damage the tunnel is less from lined samples this is due to the resistance offered by the lining material in case of lined samples. Muhammad, (2017) conducted a study on circular tunnels under static and dynamic loads at different depths. Numerical investigation is conducted with the help of a finite element software tool MIDAS GTS NX. It was noticed that the depth of the tunnel has a significant effect on the stability behavior.

3. Various Theories on stress distribution in Tunnels

During the designing of the tunnel, evaluation of stresses and displacement around the circular opening is primary requirement. There are various methods are available which are helpful in determing the stress distribution and displacement behaviour of tunnels when they are subjected to different loading conditions. Numerous solutions have been given for the distribution of stresses around the circular opening. A lot of researchers had giver their theories and solutions for the stresses acting around the underground structures. Kirsch was the first who had given the analytical solution for the stress and displacement around the circular opening in plain strain conditions in 1898. According to Kirsch, for a circular opening of radius “a” and radial distance “r”, the expression for radial, tangential and shear stress are given.

$$\sigma_r = 1/2.(\sigma_x+\sigma_z).(1-a^2/r^2)+1/2.(\sigma_x-\sigma_z).(1+3a^4/r^4-4a^2/r^2).\cos 2\theta \quad (1)$$

$$\sigma_\theta = 1/2.(\sigma_x+\sigma_z).(1+a^2/r^2)-1/2.(\sigma_x-\sigma_z).(1+3a^4/r^4).\cos 2\theta \quad (2)$$

$$\tau_{r\theta} = -1/2.(\sigma_x-\sigma_z).(1-3a^4/r^4+2a^2/r^2).\sin 2\theta \quad (3)$$

Where, σ_r = radial stress
σ_θ = tangential stress
$\tau_{r\theta}$ =shear stress
$\sigma_z = \gamma.h$ = overburden pressure
$\sigma_x = K.\sigma_z$ = horizontal pressure
a= radius of the opening
r = radial distance from the centre of opening

From the above equations given by Kirsch (1898) following conclusions can be drawn:

(i) In case $\sigma_x = 0$, the maximum tangential stresses will be equal to three times the applied stresses.

(ii) For $\sigma_x = 0$, tangential stress will be equal to the applied stress with opposite sign at the boundary of the opening when $\theta = 90^\circ$ or 270°.

(iii) For $\sigma_x = \sigma_z$, the maximum tangential stress will be equal to the twice the applied stress. Also the maximum tangential stress will occur at the boundary of circular opening.

Terzaghi and Richart (1952) also gave some equation to study the stress distribution behaviour of circular opening by using Kirsch solution. The equations given by Terzaghi and Richart (1952) are as follows:

$$\sigma_h = 1/2.(\sigma_\theta + \sigma_r) + 1/2\,(\sigma_\theta - \sigma_r).\cos 2\theta + \tau_{r\theta} \sin 2\theta \quad (4)$$

$$\sigma_v = 1/2.(\sigma_\theta + \sigma_r) - 1/2\,(\sigma_\theta - \sigma_r).\cos 2\theta - \tau_{r\theta} \sin 2\theta \quad (5)$$

$$\tau_{vh} = -1/2.(\sigma_\theta - \sigma_r)\sin 2\theta + \tau_{r\theta}\cos 2\theta \quad (6)$$

Where, σ_r = radial stress
σ_θ = tangential stress
$\tau_{r\theta}$ =shear stress
$\sigma_z = \gamma.h$ = overburden pressure
$\sigma_x = K.\sigma_z$ = horizontal pressure
a= radius of the opening
r = radial distance from the centre of opening

4. Conclusion

From the above review following conclusions can be made:

(i) There are many factors which affect stability of underground structures such as type of soil, overburden pressure, rock profile, discontinuities or joints present in rock, relative position of tunnels, construction procedure and alignment of the tunnel etc.

(ii) There is a significant effect of the shape of the tunnel section on its stability behaviour. It can be noticed from the above review that circular section is best shape tunnel section as the deformation experienced in the circular section is least as compare to other tunnel sections.

(iii) Various stress distribution theories are derived by different researchers and it can be noticed that stress distribution behaviour of tunnel is mainly affected by radial stresses, tangential stresses and shear stresses. Also, with the help of these theories the boundary conditions of the tunnel model can also be decided.

Acknowledgements

We would like to convey our deep gratitude to Head of the Civil engineering Department and all the faculty members for their cooperation during the study. We would also like to thank Delhi Technological University authorities for their sponsorship and financial support. The helpful and supportive behavior of laboratory assistants of rock mechanics lab and geotechnical engineering lab is appreciatively acknowledged.

References

[1] E. A. Elshamy, G. Attai, H. Fawzy and K. Abdel, "Behaviour of different shap of twin tunnels in soft clay soil," *Int. J. Engineering and Innovative Technology (IJEIT),* vol. 2, 2013.

[2] F. Ahmad and Y. Bahodor, "Analysis of the impact of various factors on the ground settlement profile due to excavation twin tunnels," *International Conference in New Research on Civil Architectural and Urban management*,2016.
[3] F. Chehade and I. Shahrour, "Numerical analysis of the interaction between twin-tunnels: Influence of the relative position and construction procedure," *Tunnelling and Underground Space Technology*, vol. 23, pp. 210-214, 2008.
[4] F. Shalabi, "Interaction of Twin Circular shallow tunnels in soil-Parametric Study," *Open Journal of Civil Engineering*, vol. 7, pp.100-115, 2017.
[5] I. A. Khan, K. Venkatesh, R. K. Srivastava, "Fractured and Un-fractured analysis of a rock tunnel by finite element method," *Indian Geotechnical Conference*, 2015.
[6] J. Idris, T. Verdel and M. Al-Heib, "Numerical modelling and mechanical behaviour analysis of ancient tunnel masonry structures," *Tunnelling and Underground Space Technology*, vol. 23, pp. 251-263, 2008.
[7] J. Idris, T. Verdel and M. Al-Heib, "Numerical modelling of masonry joints degradation in built tunnels," *Tunnelling and Underground Space Technology*, vol. 24, pp. 617–626, 2009.
[8] J. Mohammed, "Numerical modelling for circle tunnel under static and dynamic loads for different depth," *Research Journal of Mining*, vol. 1, 2017.
[9] M. A. Nunes and M. A. Meguid, "A study on the effect of overlying soil strata on the stresses developing in a tunnel lining," *Tunnelling and Underground Space Technology*, vol. 24, pp. 716-722, 2009.
[10] P. Jia and C. A. Tang, "Numerical study on failure mechanism of tunnel in jointed rock mass," *Tunnelling and Underground Space Technology*, vol. 23, pp. 500-507, 2008.
[11] P. Kumar and A. K. Shrivastava, "Various Factors Effecting Stability of Underground Structures: State of Art," *7th Indian Rock Conference*, 2017, pp. 576-580.
[12] R. Tiwari, T. Chakraborty, and V. Matsagar, "Dynamic Analysis of a Twin Tunnel in Soil Subjected to Internal Blast Loading," *Indian Geotech J.*, vol. 46, no. 4, pp. 369-380, 2016.
[13] S. Mishra, K.S. Rao, N.K.Gupta and A.Kumar, "Damage to Shallow Tunnels under Static and Dynamic Loading," *Procedia Engineering 173, 1322-1329*, 2016.
[14] S. Mishra, A. Challa, A. Singh, A. Kumar, K. S. Rao and N. K. Gupta," Parametric study of lined and unlined tunnels at shallow depths under coupled static and cyclic loading conditions," *Tunnelling in Himalayan Geology*, vol. 340, 2017.
[15] Z. Zhu, Y. Li, J. Xie and B. Liu, "The effect of principle stress orientation on tunnel stability," *Tunnelling and Underground Space Technology*, vol. 49, pp. 279-286, 2015.

AIP Conference Proceedings

Design and analysis of microstructured optical fiber for supercontinuum generation

Cite as: AIP Conference Proceedings **2136**, 050002 (2019); https://doi.org/10.1063/1.5120935
Published Online: 09 August 2019

Ankit Sharma, Pooja Chauhan, Ajeet Kumar, Yogita Kalra, and Than Singh

View Online Export Citation

ARTICLES YOU MAY BE INTERESTED IN

Effects of polycrystalline birefringent grains on the morphology dependent resonance modes of a spherical resonator
Journal of Applied Physics **126**, 053102 (2019); https://doi.org/10.1063/1.5094333

Investigation on nonlinear optical responses of different pyrrole derivatives: A computational study
AIP Conference Proceedings **2136**, 040005 (2019); https://doi.org/10.1063/1.5120919

Tuning of band gap energy of semiconducting Zn-Cu nanoferrite by varying ions (Cu^{+2}, Zn^{+2}) concentration
AIP Conference Proceedings **2136**, 040018 (2019); https://doi.org/10.1063/1.5120932

AIP | Conference Proceedings
Get 30% off all print proceedings!
Enter Promotion Code PDF30 at checkout

AIP Publishing

© 2019 Author(s).

Design and Analysis of Microstructured Optical Fiber for Supercontinuum Generation

Ankit Sharma[1], Pooja Chauhan[1, a)], Ajeet Kumar[1, b)], Yogita Kalra[2] and Than Singh[3]

[1]*Advanced Photonics Simulation Research Laboratory, Department of Applied Physics, Delhi Technological University, Delhi, INDIA*
[2]*TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication, Department of Applied Physics, Delhi Technological University, Delhi, INDIA*
[3]*Optical Functional Materials Laboratory, Toyota Technological Institute, Nagoya, JAPAN*

[a)]Corresponding author: pooja_ch21@yahoo.com
[b)] ajeetdph@gmail.com

Abstract.A microstructured optical fiber in $Ge_{11.5}As_{24}Se_{64.5}$ chalcogenide glass has been numerically modeled for supercontinuum generation. The proposed design holds all-normal dispersion property and offers very high nonlinear coefficient at a pump wavelength of 2.8 μm.

INTRODUCTION

Chalcogenide-based microstructured optical fibers (ch-MOF) are becoming superior in comparison of conventional optical fibers due to their flexibility in engineering dispersion profile, high nonlinear refractive index, and broad optical transparency window in the mid-infrared region [1,2]. Several chalcogenide glasses such as $Ge_{11.5}As_{24}Se_{64.5}$, As_2S_3, As_2Se_3, and $Ge_{11.5}As_{24}S_{64.5}$ are widely in practice for improving the output efficiency of the infrared photonic devices because of their high nonlinearity and need of low threshold power [3].

In our study, we have tailored the geometrical parameter of the proposed ch-MOF in order to achieve an all-normal-dispersion profile and high nonlinearity to get ultra-broadband supercontinuum generation (SCG).

PROPOSED CH-MOF DESIGN

The air-holes are arranged symmetrically in a rectangular lattice pattern. We have considered the diameter of air-holes as d in the cladding region made up of $Ge_{11.5}As_{24}Se_{64.5}$ chalcogenide glass and the center to center distance of air-holes is denoted by Λ and taken as 1μm.



FIGURE 1. Electric field distribution of fundamental mode at a pump wavelength of 2.8 μm.

International Conference on Photonics, Metamaterials & Plasmonics
AIP Conf. Proc. 2136, 050002-1–050002-4; https://doi.org/10.1063/1.5120935
Published by AIP Publishing. 978-0-7354-1879-0/$30.00

RESULTS AND DISCUSSION

The linear and nonlinear characteristics for the SCG are studied using the finite element method. We have followed the process of optimization of the geometrical parameter i.e., *d* in order to obtain the all-normal dispersion profile.

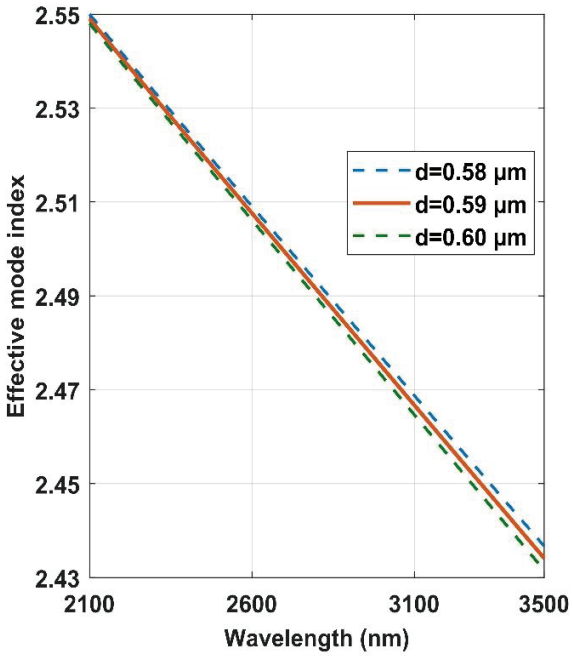


FIGURE 2. Variation of mode index along wavelength.

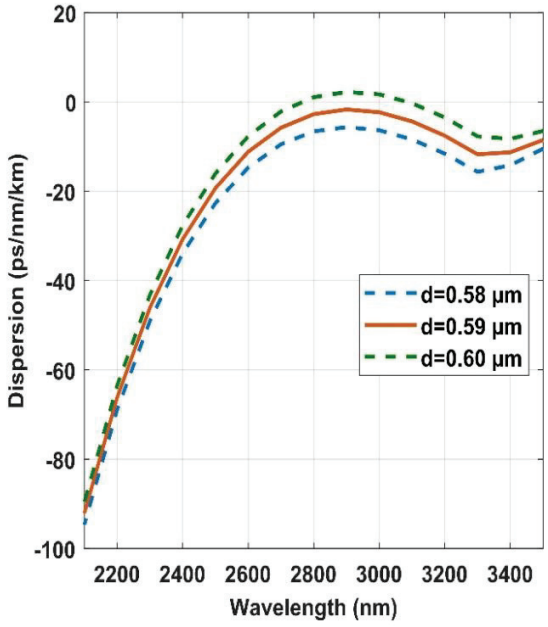


FIGURE 3. Effect of *d* on chromatic dispersion.

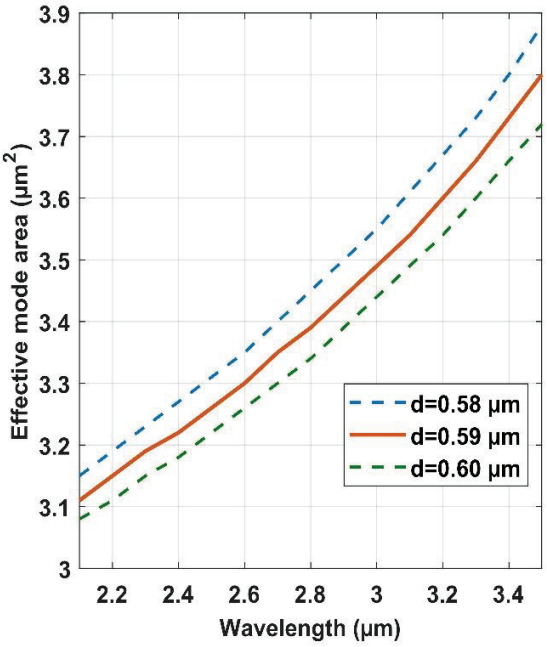


FIGURE 4. Effect of *d* on mode area.

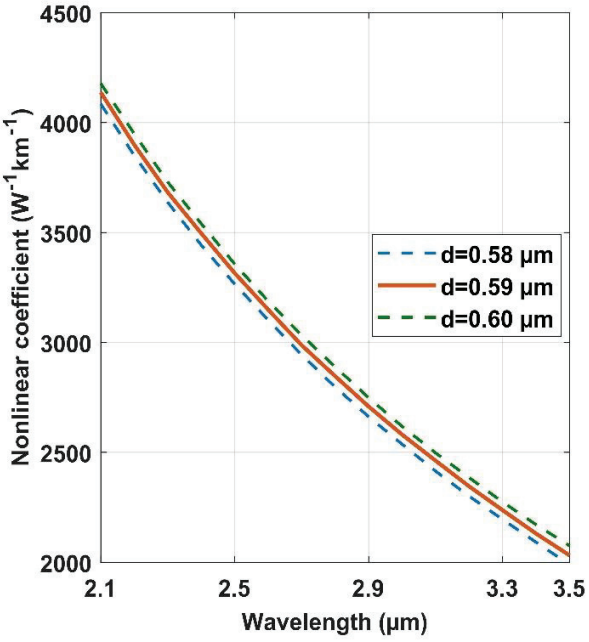


FIGURE 5. Effect of *d* on nonlinearity.

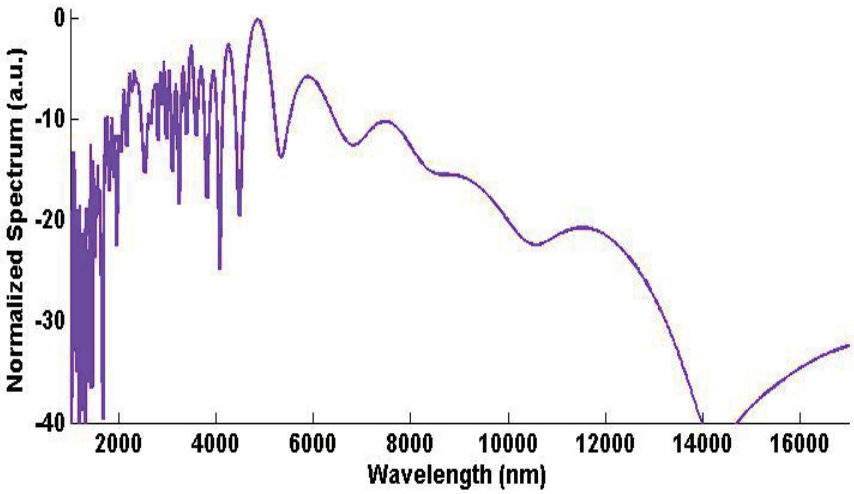


FIGURE 6.SCG spectra

CONCLUSIONS

In our study, we have numerically proposed a $Ge_{11.5}As_{24}Se_{64.5}$ chalcogenide based microstructured optical fiber for the supercontinuum generation in the mid-infrared region. We have achieved an ultra-broadband supercontinuum spanning, 2-15 µm using a femtosecond laser source in a few millimeters fiber length.

TABLE 1. Optimized parameters with numerically reported results

Optimized dimensions	Pump Wavelength	Dispersion	Nonlinear Coefficient	Effective Mode Area	SCG
d =0.59 µm; Λ=1 µm	2.8 µm	-2.7 ps/nm/km	2889 $W^{-1}.km^{-1}$	3.39 μm^2	2-14 µm

.

REFERENCES

1. P. Russell, "Photonic crystal fibers", *Science* **299**, 358-362 (2003).
2. J. M. Dudley, G. Genty, and S. Coen, "Supercontinuum generation in photonic crystal fiber", *Rev. Mod. Phys*. **78**, 1153-1184 (2006).
3. M. R. Karim, B. M. A. Rahman and G. P. Agrawal, "Dispersion engineered Ge-As-Se nanowire for supercontinuum generation: a parametric study", *Opt. Expres* **22**. 312029-31040 (2014).

Silicon
https://doi.org/10.1007/s12633-019-00246-x

ORIGINAL PAPER

Design Considerations and Capacitance Dependent Parametric Assessment of Gate Metal Engineered SiNW MOSFET for ULSI Switching Applications

Neha Gupta[1] · Ajay Kumar[2] · Rishu Chaujar[3]

Received: 5 April 2019 / Accepted: 12 August 2019
© Springer Nature B.V. 2019

Abstract
This work investigates design consideration of Gate Electrode Workfunction Engineered (GEWE) silicon nanowire MOSFET at room temperature. It is perceived from the results that the parasitic capacitances are higher in the deep inversion region, however; the coupling of parasitic capacitance is weak in GEWE-SiNW due to lower gate metal workfunction at the drain end. Moreover, the influence of drain voltage on C_{gs} is very less owing to constant depletion charges at the source side which results in almost constant value of C_{gs} with the change in V_{ds}. With tuning of channel length (L_g) and oxide thickness (t_{ox}), the effect of capacitances in GEWE-SiNW further reduces significantly due to quantization effect at such scaled dimensions. Furthermore, Transconductance Frequency Product (TFP), Energy Delay Product (EDP) and Gain Bandwidth Product (GBP) have also been calculated with an aim to analyze the device DC and switching performance as they are directly or indirectly linked with parasitic capacitance. Calibrated 3D simulations validate that the GEWE-SiNW MOSFET exhibit 2.57, 34.5 and 2.75 times improvement in TFP, EDP and GBP respectively in comparison to SiNW in the linear region.

Keywords Quantum capacitance · Gate electrode Workfunction engineering · Silicon nanowire

1 Introduction

To achieve high package density, low power dissipation, and higher cut-off frequencies, the integrated circuit's device dimensions are continuously scaled down to meet the demand of ULSI industry. Scaling of the device to such a small dimensions causes CMOS device to show shortcomings such as V_{TH} roll-off, drain induced barrier lowering (DIBL), gate leakage current, etc. [1, 2]. From past few years, numerous device structures such as Double gate MOSFET [3], Recessed Channel MOSFET [4–6], cylindrical gate MOSFET [7, 8], Fin-FETs [9, 10], Nanowire MOSFET [11, 12] have been described in literature to overcome the drawbacks associated with scaling of device dimensions. Amongst all, silicon nanowire (SiNW) MOSFET has appeared as a most capable contender for future based silicon nano-device due to ideal symmetry, high on-current due to their quantum nature and because of small volume, low leakage current and their adaptability for high-density integration including that of 3D [13].

Inversion layer capacitance [14] is major concern in nano-devices with thin gate oxides as it greatly affects the performance of scaled devices owing to its influence on transconductance and thus hamper the driving capability of a device. It is jointly determined by both centroid and quantum capacitance as given by Eq. 1 [14]. In inversion layer, the shape of charge distribution whether cylindrical or planar determines the centroid capacitance, it is dependent on the geometry of device structure and quantum capacitance is due to the density of states (DOS), which causes diffusion of fermi

✉ Rishu Chaujar
chaujar.rishu@dtu.ac.in

Neha Gupta
gupta.neha@niecdelhi.ac.in

Ajay Kumar
ajaykumar@dtu.ac.in

[1] Department of Applied Science and Humanities, ADGITM (Formerly Northern India Engineering College), New Delhi, India

[2] Electronics and Communication Engineering Department, JIIT, Noida, India

[3] Applied Physics Department, Delhi Technological University, Delhi, India

Published online: 21 August 2019

level inside the sub-bands of quantum well as given by Eq. 2 and 3 [15].

$$C_{inv} = \left(\frac{1}{C_{qs}} + \frac{1}{C_{es}}\right)^{-1} \tag{1}$$

Where,

$$C_{qs} = \frac{\left(\frac{m^* q^2}{\pi\hbar^2}\right)}{1 + \left[\exp\left(E_i - E_f\right)/kT\right]} \tag{2}$$

$$C_{es} = C_{q,n}\frac{\partial(E_F - E_i)}{\partial(E_i - E_c)} \tag{3}$$

The total gate capacitance of scaled MOSFET, C_{gg} is defined by Eq. 4. In scaled devices (t_{ox} = 1-2 nm), C_{ox} becomes very large and thus leads to degradation in C_{gg} owing to high C_{inv}.

$$C_{gg} = \left(\frac{1}{C_{ox}} + \frac{1}{C_{inv}}\right)^{-1} \tag{4}$$

Where,

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \tag{5}$$

Moreover, a significant key parameter which describes the speed and performance of CMOS devices is the gate delay (τ), which is defined as $C_{gg}V/I$. Since it is dependent on the total gate capacitance, therefore, for the first time, we have scrutinized the gate capacitance of GEWE-SiNW MOSFET at the sub-nm range by means of numerical simulations. We focus on the GEWE-SiNW [16–19] scheme as it is most favorable for scaled CMOS technology owing to higher immunity against SCEs and also it enhances the overall current driving capability of device. In Section II, we present the 3D pictorial view of GEWE-SiNW MOSFET along with its physical parameters. The simulation methodology that we invoked during this analysis and calibration graph is explained in section III. Section IV discussed the results so obtained by examining the influence of operating regions on the gate capacitance and also the effectiveness of gate length, oxide thickness scaling and gate workfunction engineering scheme is also observed. Furthermore, TFP, GBP, and EDP are also studied for the same. In the end, Section V summarizes the main conclusion of the paper.

Figure 1 (a) shows the 3D view of GEWE SiNW MOSFET. It contains two gates of material gold (Φ_{m1} = 4.8 eV) and aluminum (Φ_{m2} = 4.4 eV) of equal length (15 nm) each as shown in Fig.1 (a). The gate length (L_g) and radius of silicon nanowire (R) are taken 30 nm and 5 nm respectively. The thickness of gate oxide is 1.5 nm, the source and drain regions are heavily doped with uniform doping profiles of 5×10^{19} cm^{-3}. In this investigation, results so obtained are simultaneously compared with the SiNW MOSFET and metal gate used for SiNW is gold (Φ_m = 4.8 eV) throughout this analysis unless stated otherwise. Figure 1(b) displays the intrinsic (or bias dependent) parasitic capacitance associated with GEWE-SiNW and Fig. 1(c) indicates the gate capacitance (C_{gg}) of nanowire MOSFET under strong inversion conditions and its components. C_{gg} consists of oxide capacitance (C_{ox}) and inversion layer capacitance (C_{inv}) [20]. The C_{inv} includes the influence of electrostatic capacitance which is due to channel electron's average distance from the Si/SiO_2 interface, and C_{qc} is due to the density of states in inversion layer [15].

2 Simulation Methodology and Calibration

In sub-nm devices, the effect of quantum confinement (QC) may not be overlooked as it plays an important role in device physics [21]. As the inversion charge layer is equivalent to the nanowire dimension, hence the effect of QC in the sub-nm device may not be ignored. To incorporate QC in the channel region, Bohm Quantum Potential (BQP) model is used in this analysis. In this model, the classical potential is reformed by the quantum potential in the same manner as for the continuity equation [22]. From Figs. 2(a-b), it is perceived that with quantum models, GEWE-SiNW MOSFET shows significant enhancement in on-current performance in comparison to classical Drift Diffusion Model (DDM) and semi-classical Model (Hydrodynamic Model) since quantum models considers the quantum confinement in 1-D, enhances the mobility and thus device performance. On the contrary, DDM neglects the velocity overshoot and all non-local effects which arise due to SCEs [23] and thus leads to the sinking of drain current as is shown in Fig. 2(a-b) whereas HDM models takes velocity overshoot in consideration which has a strong influence on the terminal currents and thus has a higher drain current in comparison to DDM model. Therefore, to analyze the better understanding of GEWE-SiNW MOSFET, Quantum mechanical model is required.

BQP is always used either with HDM or Energy Balance (EB) transport model. Thus, hydrodynamic model (HDM) is used with BQP to take into account the velocity overshoot and diffusion associated with the carrier temperature. The model introduces a position dependent quantum potential, Q, which is added to the Potential Energy of a given carrier type. This quantum potential is derived using the Bohm interpretation of quantum mechanics and is described below [22]:

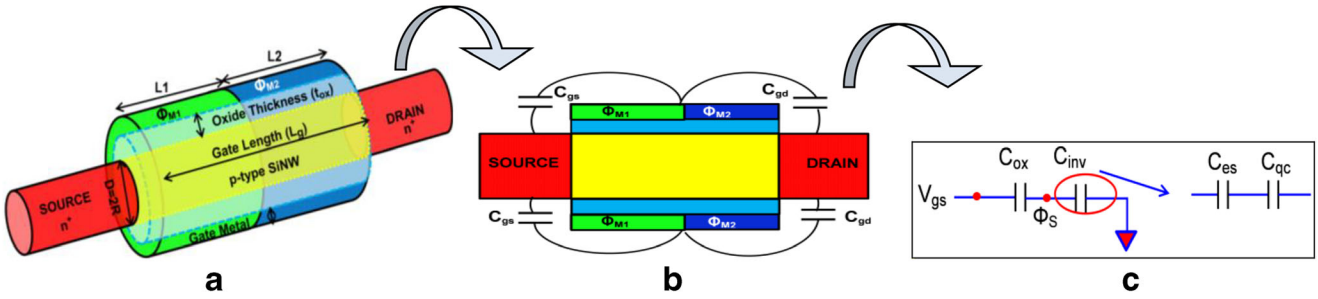


Fig. 1 (**a**) Schematic view of GEWE-SiNW MOSFET, (**b**) Intrinsic Capacitances associated with GEWE-SiNW MOSFET and (**c**) Components of total gate capacitances of GEWE-SiNW MOSFET

Time independent Schrodinger equation can be written as:

$$\left(\frac{-\hbar}{2m^*}\nabla^2 - V\varphi\right)\psi_j = E_j\psi_j \tag{6}$$

According to Bohm's interpretation of quantum mechanics, the wave function in a polar representation can be written as:

$$\psi = R.\exp\left(\frac{iS}{\hbar}\right) \tag{7}$$

Where R is probability density per unit volume, S has dimensions of action (energy×time).

By substituting Eq. 7 in 6, the Schrodinger equation can be written as:

$$\frac{\hbar^2}{2}\nabla\left\{M^{-1}\nabla\left[R.\exp\left(\frac{iS}{\hbar}\right)\right]\right\} + V.\mathrm{R}\exp\left(\frac{iS}{\hbar}\right) = E.R.\exp\left(\frac{iS}{\hbar}\right) \tag{8}$$

Where M^{-1} is the inverse of effective mass tensor and $M^{-1}\nabla S$ is local velocity of particle associated to wave function, E is total energy and equal to sum of potential (V) and kinetic energy ($0.5\ M^{-1}\nabla S$). Therefore, Quantum potential is defined as:

$$V = \frac{-\hbar^2}{2}\frac{\nabla\left(M^{-1}\nabla R\right)}{R} \tag{9}$$

Potential V is obtained by the single particle Schrodinger equation and use "*mean field approximation*". Effective Quantum Potential is written as:

$$Q_{eff} = \frac{-\hbar^2}{2}\gamma\nabla\frac{\left[M^{-1}\nabla(n^{\alpha})\right]}{n^{\alpha}} \tag{10}$$

where α and γ are two adjustable parameters, M^{-1} is the inverse effective mass tensor, n is the electron (or hole) density and h is a Planck's constant.

Bohr Quantum Potential (BQP) Model with alpha = 0.5 and gamma = 1.2 was used to take quantum mechanical effects. Additionally, bi-conjugate stabilized (BICGST) numerical method has also been called in simulation to obtain better convergence in 3D structure [24]. To validate the simulated data, results are first calibrated with the experimental results of SiNW MOSFET. Figure 3(a) demonstrates the simulated and the experimental results of I_{ds}-V_{gs} of SiNW MOSFET. The good agreement between simulated and the experimental data supports the validity of the simulation models and its parameters used in the simulation. Figure 3(b) shows the performance comparison of GEWE-SiNW with SiNW at V_{ds} =

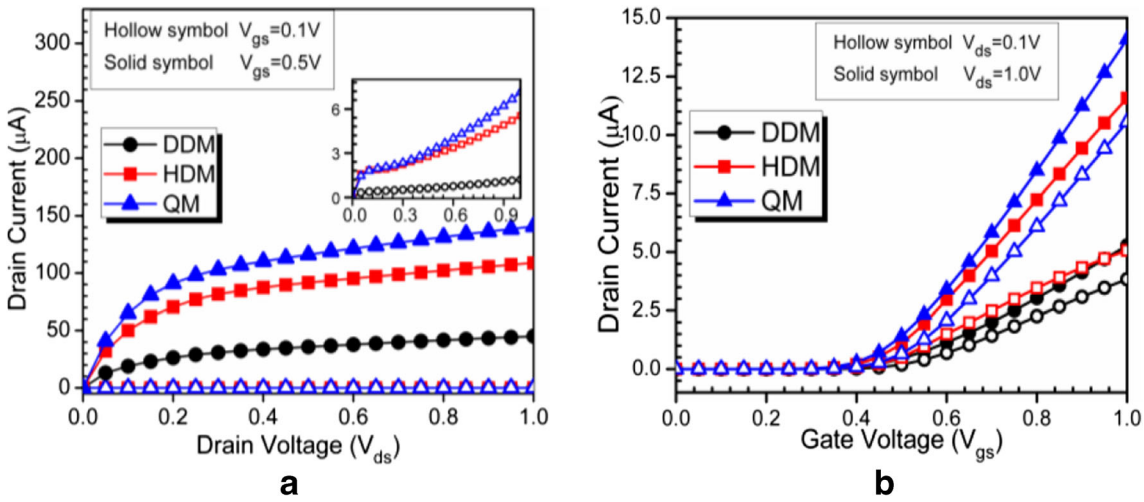


Fig. 2 (**a**-**b**) Transfer and Output Characteristics of GEWE-SiNW MOSFET at different transport models

0.2 V, and it is found that subthreshold swing (SS) and threshold voltage (V_{th}) of GEWE-SiNW is comparatively lower in comparison to SiNW due to improved gate controllability and improved SCEs immunity owing to GEWE scheme. Moreover, DIBL which is an important indicator of SCEs also improved considerably as evident from the inset of Fig. 3(b). DIBL is significantly reduced in GEWE-SiNW MOSFT in comparison to SiNW MOSFET owing to gate workfunction engineering scheme in which $V_{th1} > V_{th2}$ due to high workfunction at source side which results in screening of drain bias and improves DIBL. In addition, to validate the simulation models, the device is calibrated with the experimental result obtained from [26] in terms of output characteristics as shown in Fig. 3(c).

3 Results and Discussion

In sub-nm regime, it is very challenging to achieve lowest possible delay for low power digital applications due to aggressive scaling of MOSFETs which causes degradation in capacitances [27] and thus causes hindrance in device switching performance. In this section, bias dependent total gate capacitance (C_{gg}) of GEWE-SiNW is examined in terms of the gate to source (C_{gs}), gate to drain (C_{gd}) capacitance in different regions of operation to analyze the behavior of the device. Since C_{gg} is symmetrically distributed amid both the source and drain regions as both are coupled to the inversion region formed beneath the gate dielectric region. The influence of device parameters variation (L_g, t_{ox} and gate metal workfunction) of GEWE-SiNW MOSFET on the same has also been considered.

3.1 C-V Analysis

Figure 4(a-b) shows the variation of bias-dependent source-gate capacitance (C_{sg}) and drain-gate capacitance (C_{dg}) for GEWE-SiNW MOSFET and SiNW MOSFET for different drain bias voltage. With the increase in drain bias from 0.2 V to 0.8 V, C_{gs} shows very small variation with an increase in gate voltage since with increase in drain voltage, the charges in the depletion region at the source end remains almost constant. But, C_{gd} decreases with increase in drain voltage at higher gate voltage due to the widening of reverse biased drain depletion region which reduces C_{gd}. In addition, at a higher gate voltage (i.e., in deep inversion region) the inversion layer shifts towards the source side which leads to the enhanced gate to source capacitance in comparison to gate to drain capacitance as clear from Fig.4 (a-b). Moreover, GEWE-SiNW exhibits smaller C_{gs} and C_{gd} in comparison to its conventional SiNW MOSFET due to metal gate workfunction difference which results in the improved screening of conducting channel from drain variations. C_{gs} and C_{gd} of GEWE-SiNW and SiNW as a function of drain voltage for different gate bias variation are depicted in Fig. 5(a-b).

In MOS devices, when V_{gs} is less than V_{th} (threshold voltage), there is no inversion layer present beneath the surface, and no current flows even for $V_{ds} > 0$, and thus the inversion layer capacitance is ideally zero. But in reality, C_{gs} and C_{gd} still exist due to depleted charges present in the region below the gate which results in capacitances as is depicted from Fig. 5(a). With further increase in gate voltage to 0.5 V (>V_{th}), a conductive channel is formed under the gate region, i.e. a non-zero transverse field is present. At this point, the formation of inversion layer starts from the source side to drain side, and an increase in drain bias pinches the inversion layer towards the drain side, leads to accumulation of charge carriers at the source side which results in an increment of C_{gs}. To conserve the charge neutrality, the charges proximate to the drain region decreases hence the C_{gd} due to the presence of the higher potential barriers at the drain end. This is evident from Fig. 5(a-b). The resultant value of C_{gs} and C_{gd} is very less in GEWE-SiNW MOSFET as compared to SiNW MOSFET due to the incorporation of GEWE scheme, hence making

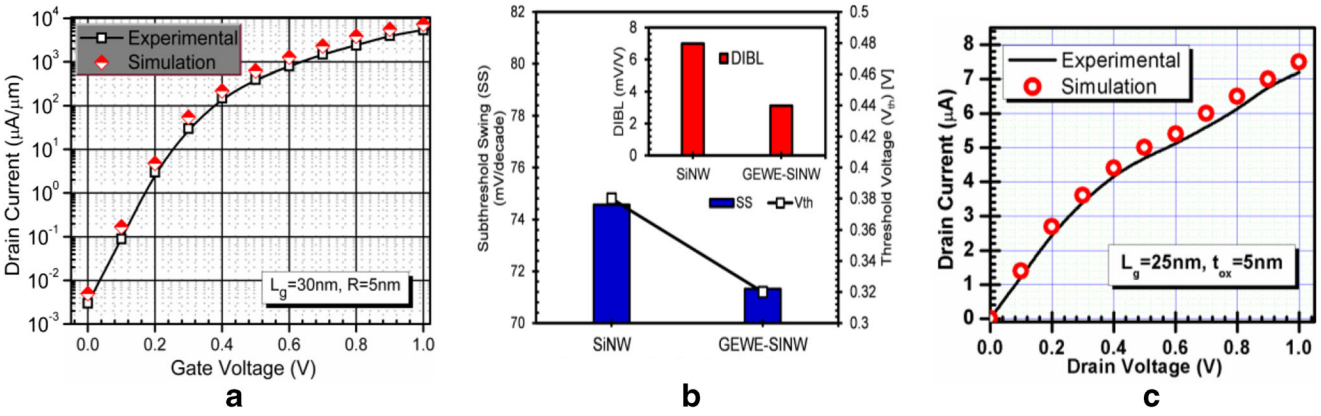


Fig. 3 Calibration with the experimental results, (**a**) I_{ds}-V_{gs} characteristics of the 5 nm radius with 30 nm short channel cylindrical gate silicon nanowire MOSFET with $t_{ox} = 2$ nm and TiN as the metal gate at $V_{ds} =$ 1.0 V, (**b**) Subthreshold Swing (SS), Threshold Voltage and DIBL of GEWE-Silicon Nanowire MOSFET at $V_{ds} = 0.2$ V [25], (**c**) Calibration of I_{ds}-V_{ds} with the experimental results obtained from [26]

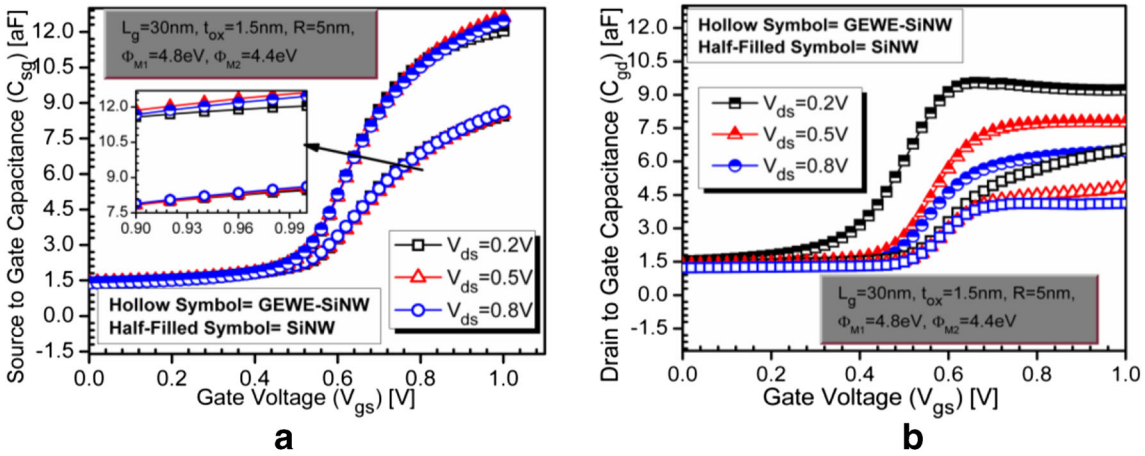


Fig. 4 (**a**) Source to the gate (C_{gs}) and (**b**) Drain to Gate (C_{dg}) capacitance of SiNW and GEWE-SiNW MOSFET as a function of gate voltage at different values of drain voltage

GEWE-SiNW MOSFET as a promising device for analog switching applications.

Figure 6(a-d) shows the dependence of capacitances on the channel length of GEWE-SiNW MOSFET at different drain bias voltage as a function of gate voltage. It is evident from figs as the gate length reduces to 10 nm, total gate capacitance C_{gg} ($C_{gs} + C_{gd}$) improves (or decrease) due to a reduction in inversion charge density in the channel region. Moreover, with rising in drain bias from 0.2 V to 0.8 V, both C_{gs} and C_{gd} degrades and this effect is more prominent as device bias in deep inversion region (0.7 V–1.0 V) because with change in V_{gs} from subthreshold to deep inversion region charges below the gate increases but as gate length reduces, the number of charges reduces and thus C_{gd} and C_{gs}. With the change in drain voltage from 0.2 to 0.8 V, depletion region near the drain end decrease. Moreover, with an increase in gate voltage by keeping drain bias constant at 0.2 V ($<V_{Dsat}$), both C_{gs} and C_{gd} increases as shown in Fig. 6(a).

As gate voltage increases further beyond 0.4 V (i.e., V_{th} of GEWE-SiNW MOSFET), conducting channel is formed beneath the gate, but due to the very small lateral field, less amount of charge carriers are pulled towards drain side thus, C_{gd} is reduced in comparison to C_{gs} as clearly evident from Fig. 6(a-d). By a further rise in gate voltage, charges are now increased at the source side (C_{gs} increases) and reduced at the drain side which also leads to reduced drain current. As drain bias increases to 0.7 V ($=V_{Dsat}$), the inversion charge density at the drain end is very small (essentially zero) and charge carriers are now easily move from source to drain side due to sufficient lateral field, and if V_{gs} is increased beyond V_{th}, more charge carriers are now responsible for the flow of current thus both C_{gs} and C_{gd} increases with increase in V_{gs}. In addition, the efficacy of workfunction engineering scheme (GEWE) is also studied in terms of gate capacitance (both C_{gs} and C_{gd}). It is observed from Figs. 7(a-d) that by modulating the workfunction difference (ΔW), capacitance effect can be reduced in working region, i.e., strong inversion

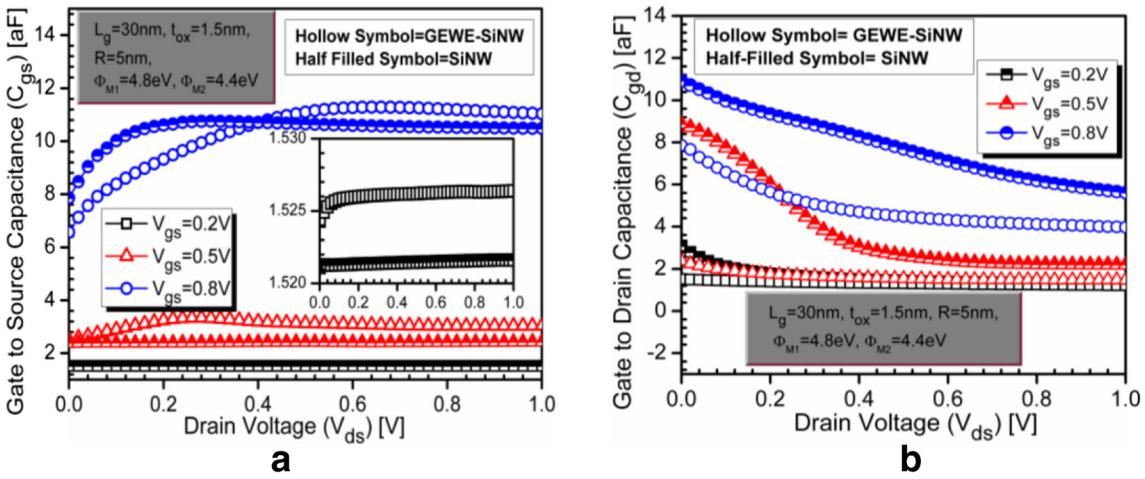


Fig. 5 (**a**) Gate to Source (C_{gs}) and (**b**) Gate to Drain (C_{dg}) capacitance of SiNW and GEWE-SiNW MOSFET as a function of drain voltage at different gate bias values

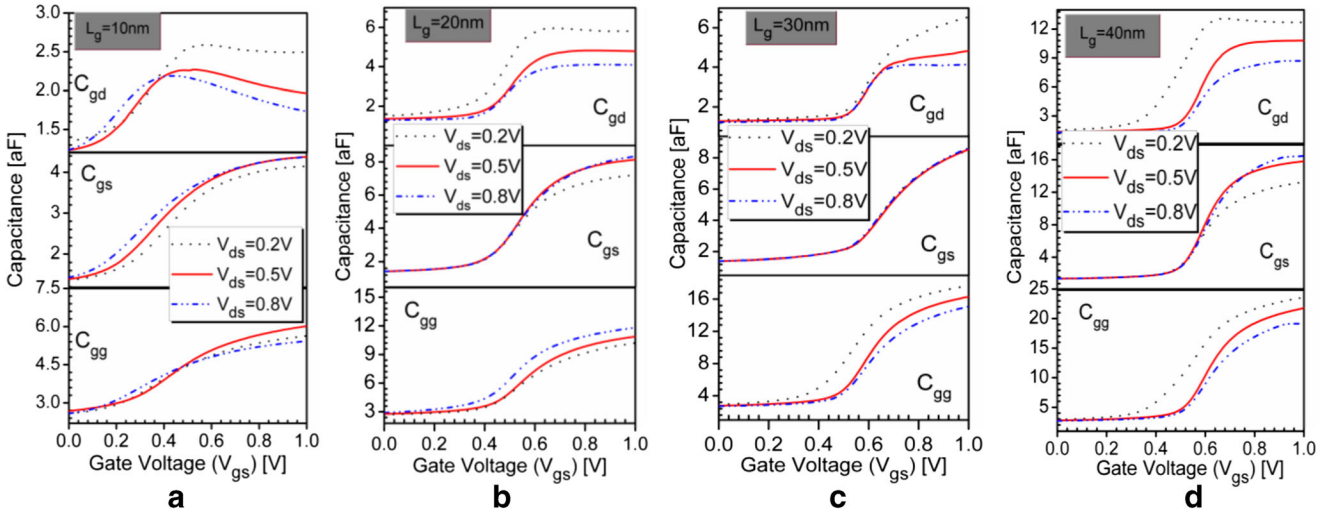


Fig. 6 (**a**-**d**) Parasitic Gate to the source (C_{gs}), Gate to Drain (C_{gd}) and total Gate Capacitance (C_{gg}) of GEWE-SiNW MOSFET at gate length (**a**) 10 nm, (**b**) 20 nm, (**c**) 30 nm and (**d**) 40 nm for different drain voltages

region. This improvement is due to the fact that with an increase in ΔW, the bias voltage is screened at the drain end which reduces the C_{gd} and thus C_{gg}. This ultimately results in reduced off current and thus improves the switching ratio which is an utmost requirement of ULSI technology.

3.2 Transconductance Frequency Product (TFP)

TFP is a very important parameter for determining the high-frequency performance of MOS devices. It is mathematically given in Eq. (11) [28]. It mainly depends upon the device efficiency and cut-off frequency. TFP is largely used for high-speed designs as it shows the compromise between power and bandwidth. It is found that in subthreshold region (0 V–0.4 V), TFP increases significantly in SiNW and GEWE-SiNW and then reaches a maximum peak value in inversion region (0.4 V–0.8 V). This peak value is higher in case of GEWE-SiNW MOSFET (≈ 100 GHz) as compared to SiNW MOSFET (≈ 25GHz) as evidently shown in Fig. 8. This peak value is higher in GEWE-SiNW due to enhancement in transconductance [29] and reduction in C_{gs} and C_{gd} ($C_{gg} = C_{gs} + C_{gd}$). With further increase in V_{gs} (0.6 V–1.0 V), TFP starts declining owing to increment in C_{gg} as evidently shown in Fig. 4(a-b).

$$TFP = \left(\frac{g_m}{I_d}\right) \times \left(\frac{g_m}{2\pi C_{gg}}\right) \tag{11}$$

Further, the effect of drain bias variation is also observed on the same parameter, and it is revealed that with $V_{ds} = 0.2$ V and

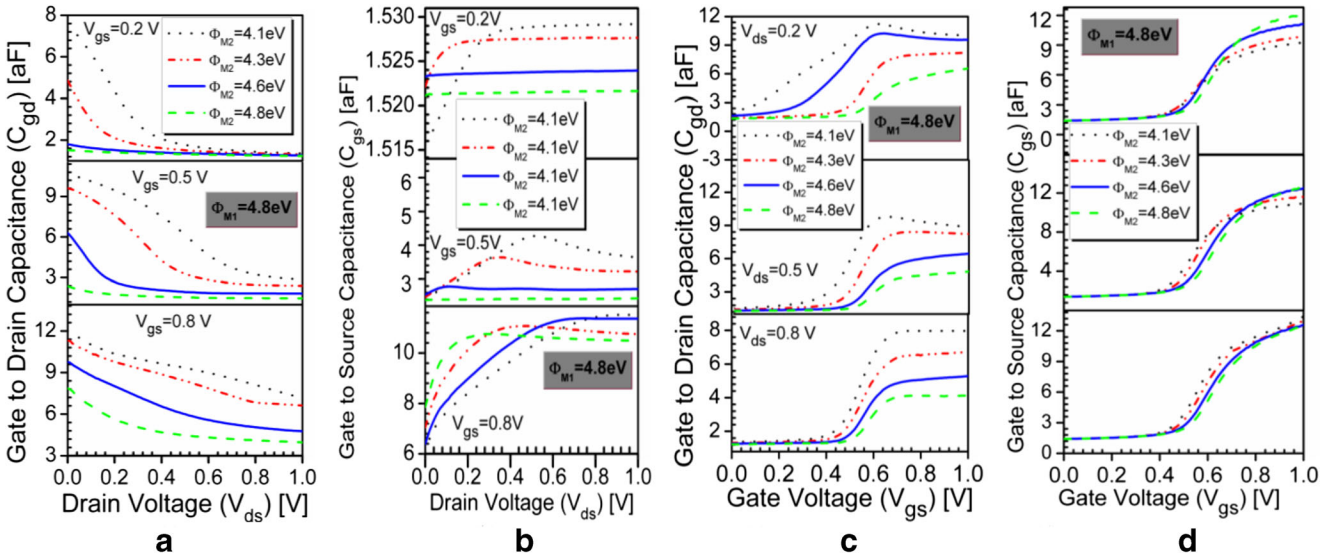


Fig. 7 (**a**-**d**) Parasitics Gate to the source (C_{gs}), Gate to Drain (C_{gd}) of GEWE-SiNW MOSFET for different gate metal workfunction at the different regime of operation

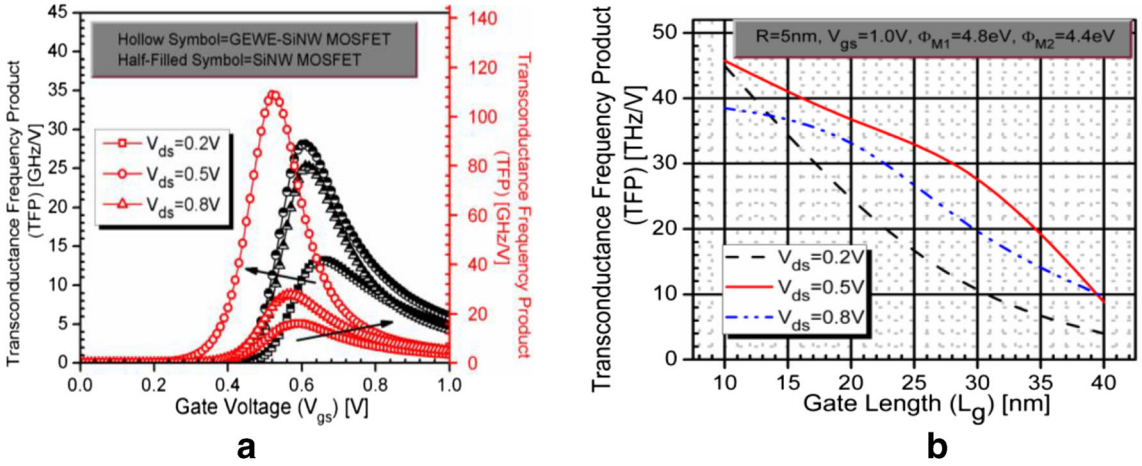


Fig. 8 (**a**) Transconductance Frequency Product (TFP) V/s gate voltage for GEWE-SiNW, SiNW MOSFET at different drain bias, (**b**) TFP as a function of gate length for different drain bias

0.5 V, TFP increases but with $V_{ds} = 0.8$ V, its value decreases. Thus we need to optimize the value drain bias as per the requirements. Figure 8 (b) shows that peak value of TFP rises considerably with scaling of the gate length of GEWE-SiNW MOSFET to 10 nm. This rise is due to improvement in transconductance due to quantum effects and workfunction engineering scheme. Moreover, it is also found that shrinking of gate length and oxide thickness simultaneously leads to better TFP performance in GEWE-SiNW MOSFET as reflected in Fig. 9(a). Figure 9(b) shows the variation of TFP as a function of drain end workfunction. It is also observed that with an increase in workfunction difference (ΔW), TFP rises significantly as transconductance increases with increase in ΔW.

3.3 Gain Bandwidth Product (GBP) and Energy Delay Product (EDP)

GBP is another important capacitance dependent parameter which is a collective measure of both bandwidth and gain of an amplifier. Mathematically, it is given by Eq. (12) [28].

$$GBP = \frac{g_m}{2\pi \times 10 \times C_{gd}} \tag{12}$$

GEWE-SiNW MOSFET exhibit higher GBP in comparison to conventional SiNW as depicted in Fig. 10(a). It is observed that GBP improves as drain voltage increases due to a reduction in C_{gd} with an increase in V_{ds} [see Fig. 5(b)]. Also, by shrinking the gate length to 10 nm in GEWE-SiNW MOSFET, GBP improves prominently as clearly indicated in Fig. 10(b).

Moreover, Energy Delay Product (EDP) calculates how much energy is spent per switching cycle and it is most important parameter especially for circuit applications comprises logic gates/families. Lower its value more efficient the device, and thus circuit is for ULSI applications. Mathematically, it is given by Eq. 13 [28]; it is the product of average energy and delay. Figure 11(a) shows the peak value of EDP at different

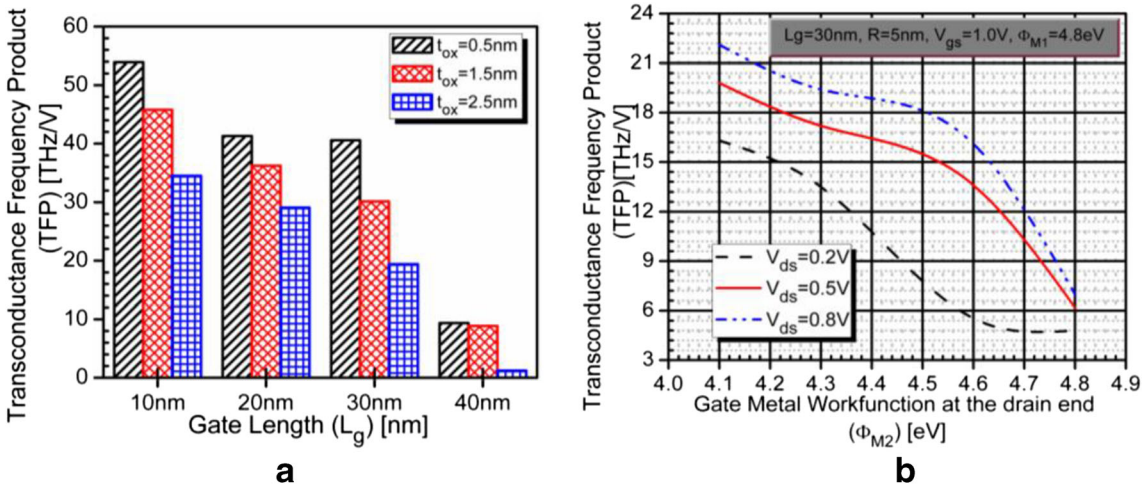


Fig. 9 (**a**) TFP as a function of gate length for different oxide thickness (t_{ox}) and (**b**) TFP as a function of gate metal at the drain end at different drain bias

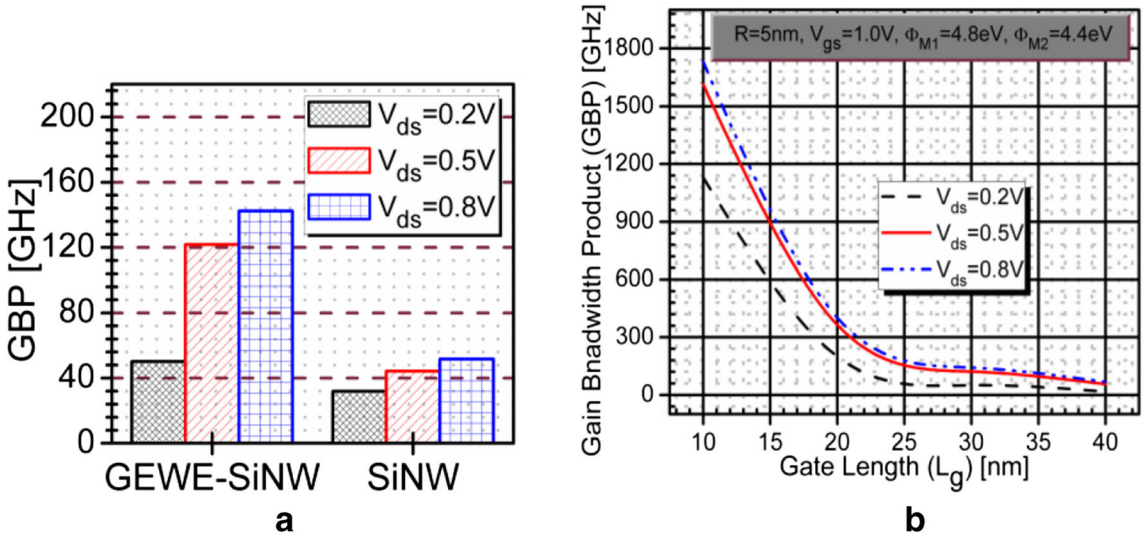


Fig. 10 (**a**) Variation of GBP of SiNW and GEWE-SiNW MOSFET for different drain bias, (**b**) GBP of GEWE-SiNW MOSFET as a function of gate length (L_g) at a different drain bias

drain bias as its value depends upon drain bias as given by Eq. 6. In GEWE- SiNW,

EDP enhances at lower drain bias in contrast to SiNW MOSFET as apparent from Fig. 11(a), due to considerable improvement in on-current and gate capacitance.

$$EDP = \left(C_{gg}V_d^2\right) \times \left(\frac{C_{gg}V_d}{I_{on}}\right) \tag{13}$$

Likewise, scaling down the gate length to 10 nm, EDP reduces appreciably by an order of 1×10^{-2} as is clear from Fig. 11(b) owing improvement in I_{on} and capacitance [see Fig.6 (a-d)]. Also, it is observed that if GEWE-SiNW is biased in the linear region, EDP improves significantly owing to reduced drain bias and capacitance coupling as shown in Fig. 11(b). Thus, provides its efficacy for low power digital applications. Gate oxide thickness in Nanodevices also plays an important role in determining switching ratio and thus delay and performance. Therefore, the effect of oxide thickness by modulating gate length is also examined. It is found that with modulation in oxide thickness along with gate length results in improvement in EDP as evidently shown in Fig.12 (b). Thus there is need to optimize the value of both gate length and oxide thickness for the appreciable value of EDP. From Fig. 12(c-d), it is perceived that influence of GEWE scheme significantly alter the EDP and GBP of GEWE-SiNW MOSFET owing to improved transconductance and reduce gate capacitance. As workfunction difference of gate

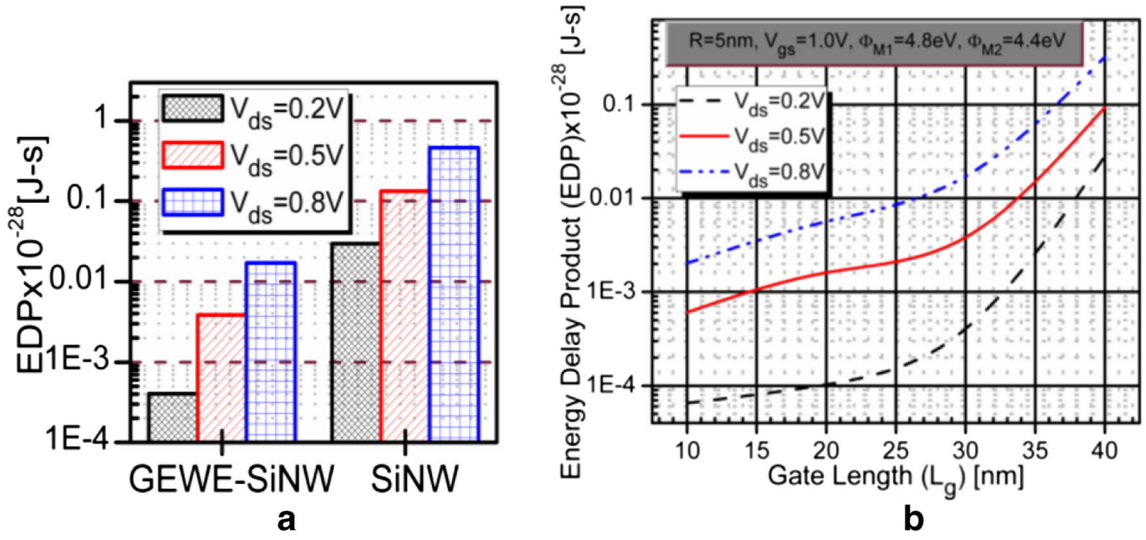


Fig. 11 (**a**) Variation of EDP of SiNW and GEWE-SiNW MOSFET for different drain bias, (**b**) EDP of GEWE-SiNW MOSFET as a function of gate length (L_g) at different drain bias

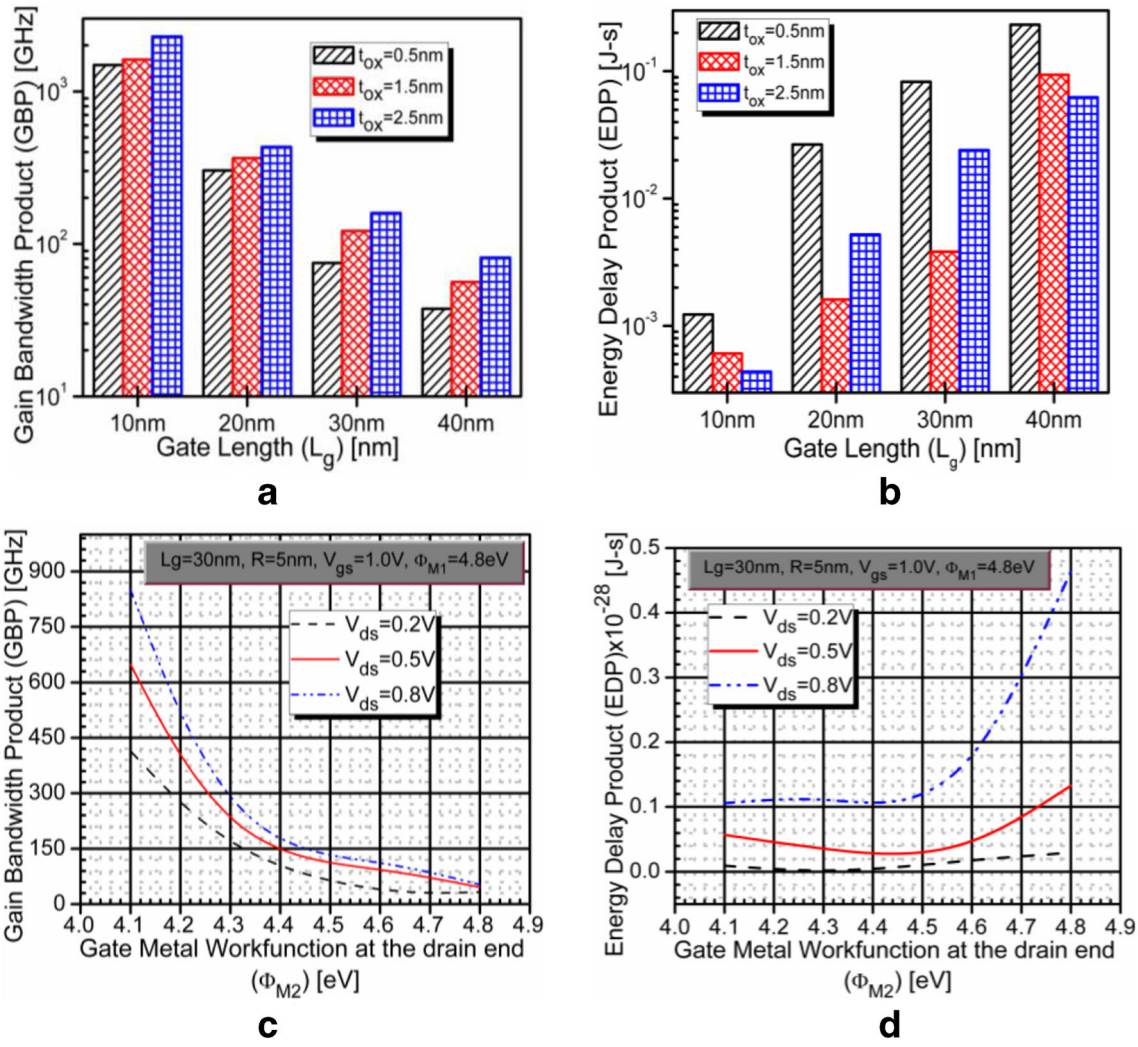


Fig. 12 (**a**) GBP and (**b**) EDP as a function of gate length (L_g) at different oxide thickness respectively and (**c**) GBP and (**d**) EDP as a function of gate metal workfunction at the drain end respectively

metal increases (by reducing the gate metal workfunction at the drain end), EDP reduces considerably due to improved current driving capability and reduced gate capacitance. Also, Gain Bandwidth product (GBP) also improved by modulating the gate length and oxide thickness as revealed in Fig. 12(a). In addition, it is observed that GBP attains higher value in the saturation region, since C_{gd} is lesser in that region (see Fig. 6).

4 Conclusion

Through this work we have explored the switching performance of gate engineered silicon nanowire MOSFET at room temperature considering the behaviour of total gate capacitance. Simulation results exhibit that influence of total gate capacitance is observed in deep inversion region i.e. beyond $V_{gs} = 0.4$ V at $V_{ds} = 0.2$ to 0.4 V. This is mainly due to density of inversion charge carriers in the channel. Moreover, variations in device design parameters are also observed to get better insight of C_{gg} in gate engineered SiNW MOSFET. In terms of magnitude, it is observed that with $L_g = 20$ nm and optimum gate workfunction difference gate engineered SiNW MOSFET exhibit improved performance due to reduced parasitic capacitance. In addition, merits that greatly affect switching performance are also calculated and found that TFP, EDP and GBP advances considerably with modulation in design parameters of gate engineered SiNW MOSFET. Thus, this study also provides the design guidelines for aptly choosing the work functions of the gate metals, gate length and the oxide thickness for low power switching applications.

Acknowledgments The authors would be grateful to Microelectronics Research Lab, Department of Applied Physics, Delhi Technological University and ADGITM (Formerly Northern India Engineering College) for providing the required aid to carry out this research work.

References

1. Borkar S (1999) Design challenges of technology scaling. IEEE micro 19(4):23–29
2. Kuhn KJ (2011) Moore's crystal ball: device physics and technology past the 15nm generation. Microelectron Eng 88(7):1044–1049
3. Kranti A, Armstrong GA (2006) Engineering source/drain extension regions in nanoscale double gate (DG) SOI MOSFETs: analytical model and design considerations. Solid State Electron 50(3): 437–447
4. Kumar A, Tripathi M, Chaujar R (2018) Comprehensive analysis of sub-20 nm black phosphorus based junctionless-recessed channel MOSFET for analog/RF applications. Superlattice Microst 116: 171–180
5. Kumar A (2017) Effect of trench depth and gate length shrinking assessment on the analog and linearity performance of TGRC-MOSFET. Superlattice Microst 109:626–640
6. Kumar A, Tripathi M, Chaujar R (2018) Reliability issues of in 2 O 5 Sn gate electrode Recessed Channel MOSFET: impact of Interface trap charges and temperature. IEEE Transactions on Electron Devices 65(3):860–866
7. Auth CP, Plummer JD (1997) Scaling theory for cylindrical, fully-depleted, surrounding-gate MOSFET's. Electron Device Letters, IEEE 18(2):74–76
8. Sarkar A, De S, Dey A, Sarkar CK (2012) Analog and RF performance investigation of cylindrical surrounding-gate MOSFET with an analytical pseudo-2D model. J Comput Electron 11(2):182–195
9. C.-Y. Chang, C.-H. Chang, C.-H. Hou, K.-L. Lin, K.-Y. Lee, X.-F. Yu, C.-O. Chui, Semiconductor devices, Finfet devices and methods of forming the same, US Patent App 15/876,223, 2019
10. K. Cheng, D.D. Lu, X. Miao, T. Yamashita, FinFET with reduced parasitic capacitance, Google patents, 2019
11. A. Abdelmoneam, B. Iñiguez, M. Fedawy, Compact modelling of quantum confinement in III-V gate all around nanowire MOSFET, 2018 Spanish Conference on Electron Devices (CDE), IEEE, 2018, pp. 1–4
12. Muscato O, Castiglione T, Coco A (2019) Hydrodynamic modeling of electron transport in gated silicon nanowires transistors, Atti della Accademia Peloritana Dei Pericolanti-Classe di Scienze Fisiche. Matematiche e Naturali 97(S1):18
13. Yang B, Buddharaju K, Teo S, Singh N, Lo G, Kwong D (2008) Vertical silicon-nanowire formation and gate-all-around MOSFET. IEEE Electron Device Letters 29(7):791–794
14. Pal HS, Cantley KD, Ahmed SS, Lundstrom MS (2008) Influence of bandstructure and channel structure on the inversion layer capacitance of silicon and GaAs MOSFETs. IEEE Transactions on Electron Devices 55(3):904–908
15. Luryi S (1988) Quantum capacitance devices. Appl Phys Lett 52(6):501–503
16. Kumar MJ, Chaudhry A (2004) Two-dimensional analytical modeling of fully depleted DMG SOI MOSFET and evidence for diminished SCEs. IEEE Transactions on Electron Devices 51(4):569–574
17. Long W, Chin KK (1997) Dual material gate field effect transistor (DMGFET), Electron devices meeting, 1997. IEDM'97. Technical digest. International, IEEE:549–552
18. Gupta N, Chaujar R (2016) Optimization of high-k and gate metal workfunction for improved analog and intermodulation performance of gate stack (GS)-GEWE-SiNW MOSFET. Superlattice Microst 97:630–641
19. Gupta N, Chaujar R (2016) Influence of gate metal engineering on small-signal and noise behaviour of silicon nanowire MOSFET for low-noise amplifiers. Applied Physics A 122(8):1–9
20. Takagi S-i, Toriumi A (1995) Quantitative understanding of inversion-layer capacitance in Si MOSFET's. IEEE Transactions on Electron Devices 42(12):2125–2130
21. Chaudhry A, Roy J (2010) MOSFET models, quantum mechanical effects and modeling approaches: a review. JSTS: J Semiconductor Tech Sci 10(1):20–27
22. Iannaccone G, Curatola G, Fiori G Effective Bohm quantum potential for device simulators based on drift-diffusion and energy transport, simulation of Semiconductor processes and devices 2004, Springer2004, pp 275–278
23. Gupta N, Chaujar R Implications of transport models on the analog performance of gate electrode Workfunction engineered (GEWE) silicon nanowire MOSFET, devices. Circuits and systems (ICDCS), 2014 2nd international conference on, IEEE, 2014, pp 1–5
24. I. SILVACO, ATLAS User's manual, Santa Clara, CA, Ver 5 (2016)
25. Suk SD, Lee S-Y, Kim S-M, Yoon E-J, Kim M-S, Li M, Oh CW, Yeo KH, Kim SH, Shin D-S (2005) High performance 5 nm radius twin silicon nanowire MOSFET (TSNWFET): fabrication on bulk Si wafer, characteristics, and reliability. IEDM Tech Dig:717–720
26. Moon D-I, Choi S-J, Kim C-J, Kim J-Y, Lee J-S, Oh J-S, Lee G-S, Park Y-C, Hong D-W, Lee D-W (2011) Silicon nanowire all-around gate MOSFETs built on a bulk substrate by all plasma-etching routes. IEEE Electron Device Letters 32(4):452–454
27. Kumar A, Tripathi MM, Chaujar R (2018) In2O5Sn based transparent gate recessed channel MOSFET: RF small-signal model for microwave applications. AEU Int J Electron Commun 93:233–241
28. Kumar A, Tripathi M, Chaujar R (2017) Investigation of parasitic capacitances of In2O5Sn gate electrode recessed channel MOSFET for ULSI switching applications. Microsyst Technol 23(12):5867–5874
29. Gupta N, Kumar A, Chaujar R (2015) Impact of device parameter variation on RF performance of gate electrode workfunction engineered (GEWE)-silicon nanowire (SiNW) MOSFET. J Comput Electron 14:798–810

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

Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Effect of Mast Modification in Ingress Problem in Gas Turbine of Naval Ship

To cite this article: P S Yadav *et al* 2019 *J. Phys.: Conf. Ser.* **1240** 012076

View the article online for updates and enhancements.



IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 14.139.251.100 on 13/08/2019 at 11:30

IOP Conf. Series: Journal of Physics: Conf. Series **1240** (2019) 012076 doi:10.1088/1742-6596/1240/1/012076

Effect of Mast Modification in Ingress Problem in Gas Turbine of Naval Ship

P S Yadav[1], V Saxena[1], H S Pali[2], N Kumar[3] and S N Singh[4]

[1] CASRAE, Delhi Technological University, Delhi.
[2] Mechanical Engineering Department, National Institute of Technology, Srinagar
[3] Mechanical Engineering Department, Delhi Technological University, Delhi
[4] Applied Mechanics Department Indian Institute of Technology, Delhi

Abstract: Over the years, these funnels which were tallest structure on ships are no more the tallest structures and are now dominated by masts which have various electronics and antenna's. This change has resulted in to smoke ingress problem. The gas turbine intake is located close to the funnels. The smoke entrapped in the wake of masts gets sucked in the intake of the turbine thereby raising the temperature at the intake resulting in lowering the efficiency of gas turbine. The flow and performance characteristics for a simplified model are carried out by providing opening in mast structure of different shapes to see the effect on smoke ingress problem.

1. Introduction

Stacks discharging the products of combustion to the atmosphere have long been the most common industrial method of disposing hot waste gases. These are also used in to discharge the exhaust gases from the gas turbines (GT). The smokestacks on ships are often referred as '**FUNNEL**'. These are located at the top side at the ship. The engineering function of funnel of ship is to discharge the products of combustion in such a manner that exhaust stays clear of the ship. Hot gases of funnel affect the deck due the smoke ingress. Due to the tall masts and funnel structure creates a low pressure zone that allows the suction of smoke in wake region. The trapping of exhaust gases into the recirculation zones results in the exhaust gases being sucked by GT intakes. Fig. 1a shows the superstructure and Fig. 1b shows model of the superstructure. The entrance of smoke in the GT affects the performance of GT engine in two ways. One is the heat contamination and other is the supply of oxygen which blended by plume. Since the GT efficiency depends on inlet temperature that showed noticeable effect on turbine efficiency. Hot gases also influence the results occurring from the topside devices like antennas, electronic communication instruments, and radar. Hot air in the vicinity of antennas can interfere with the electromagnetic communication through ionization of air.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

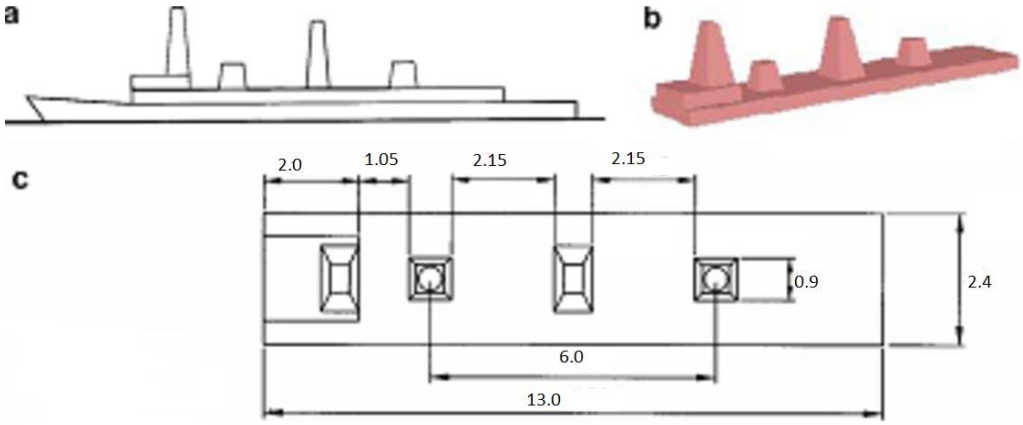


Figure 1. Generic Frigate (a) Superstructure (b) Model of superstructure (c) Graphical representation of superstructure model at 1:50 scale. (All dimensions are in meter)

Similarly the down wash can result in exposing the top deck weapons to high temperature. In order to minimize the above problems, it is important to predict the path of plume. The computational fluid dynamics (CFD) models are able to handle the most of the complexities of the superstructure. It defines the information of the overall distribution of plume on the ship and dispersion of gases from funnel. It will also provide forthcoming solution to solve the smoke ingress problem in the wake of masts. Smoke ingress problem in the naval ship continues to be a subject of research. An investigation at I.I.T Delhi has done to resolve the issue of smoke ingression in to GT intakes of the modern ship by simulating the phenomenon in a wind tunnel. A wooden 1:50 scale model Figure1(c) was used. Various locations are chosen to visualize the flow condition of the model. The flow visualization techniques of smoke injection and tuff probes ware used. Study of exhaust smoke interaction with the rest of superstructure of the ship has been through wind tunnel modeling flow behavior [1-5]. An investigation has been done and reported the flow visualization and numerical studies on generic frigate model with two funnels and progressively introducing the top side obstacle like mast and super structure [6]. To understand the interaction between bluff body air wake and the ships exhausts on naval ships, visualization has done and concluded that higher velocity from the funnel is necessary to clear the plume from the wake of superstructure and prevent the downwash.[7]. Velocity ratio of 2 at least is required for satisfactory operation when the exhaust descends sufficiently low, the exhaust smoke can get sucked in to the GT intake but to avoid infrared signature the velocity ratio should be less [8]. The study reveals that if the distance between mast and the funnel is larger than one mast height, then its effect on the downwash of the plume exhaust is not very significant [9-11]. It has been suggested the result of the flow visualization studies and several modifications including the geometry of funnel outlets and to overcome the problem of GT ingress [12-13]. An experimental study has done over the simplified superstructure and that provides information of the near field behavior of hot flumes [14-15].

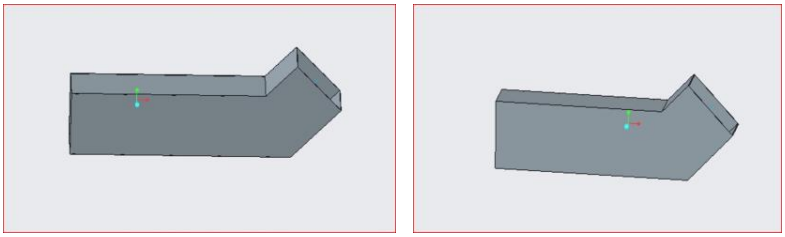
These results can also take as the reference data for direct correlation with the results for validation of the CFD code in the numerical simulation. The smoke nuisance is a very acute problem for naval ship of the world. Besides reducing the efficiency of the gas turbine, high temperature makes the ships susceptible to infrared signature. One of the methods which could be used to achieve these objectives is to blow the mast wake. The aim of present study is same.

2. Modification in geometry of mast

It is proposed to provide rectangular openings with different shapes in the mast of the prototype ship geometry and simulate flow interaction between masts and funnels to reduce the smoke ingress problem on ships. The computational code **FLUENT** is used to conduct the parametric investigation. The following two cases are investigated.

Case: 1 Mast without opening

Case: 2 Opening provided in the Mast as shown in Figure 2



(a) **(b)**

Figure 2. Openings in mast with 45 degree bend with horizontal

Table.1: Specifications of opening

S.NO	Opening	fig.	Cross-section of opening (m^2)	Height from the deck (cm)	Vertical height of opening (m)	Length of opening before bend (m)	Length of opening after bend (m)
1	**Opening 1**	**2a**	**.2x.3**	**.6**	**.295**	**.983**	**.270**
2	**Opening 2**	**2b**	**.2x.35**	**.5**	**.344**	**.9923**	**.277**

2.1 Mathematical formulation

In this, Computational fluid dynamics fluent version 6.3.26 is used. It provides a wide range of modeling capabilities for fluid flow problem. Fluent can be classified in to two parts "GAMBIT' and "FLUENT SOLVER". To create geometry and meshing, GAMBIT can be used. Other modeling software can use for modeling then import to FLUENT. The boundary condition is also applied in GAMBIT. FLUENT SOLVER solves the governing conservation equation of fluid flow by a finite volume formulation on a structured and unstructured, non-orthogonal, curvilinear co-ordinate grid system using a collocated variable arrangement. The meshed geometry in GAMBIT is exported to FLUENT solver. Since there are a number of turbulence models and solver options which can be used to solve the problem. Previous studies have shown that k–ε Turbulence model provides the best solution for such flows [14-15]. So it is used for present study. The basic governing equations used in the present study involve conservation of mass and momentum equation.

2.2 Governing Equations

Continuity Equations: The mass conservation or continuity equation (1) can be written in Cartesian tensor form for steady flow as follows: $\frac{\partial}{\partial x_i}(u_i, \rho) = s_m$ ……..(1) Where ρ=density, s_m = mass added to continuous phase from the dispersed phase u_i= velocity in i[th] direction. This is generalized equation for the conservation of mass and it is valid for both the compressible and incompressible flows.

Momentum Equation: The momentum of conservation equation (2) can be written in partial differential form in the i[th] direction as:

$$\frac{\partial}{\partial x_j}(u_i u_j \rho) = -\frac{\partial p}{\partial x_j} + \frac{\partial T_{IJ}}{\partial x_j} + g_i \rho + f_i \quad \text{…… (2)}$$

And Stress tensor equation (3) for Newtonian fluid can be written as:

$$T_{ij} = [\mu(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i})] - 2/3(\mu \frac{\partial u_i}{\partial x_i}\delta_{ij}) \quad \text{……..…….. (3)}$$

In equation (2), f_i is the external body force, $g_i\rho$ is the gravitational body force, p is the static pressure whereas μ is the molecular viscosity as shown in equation (3).

K-ε turbulence model: For closure of the above set two transport equations (4) and equation (5) are applied for k- **ε** turbulence model. These two equations of turbulence model have been recommended from earlier studies. In the rivat deviation of k- ε model, it is considered in assumption that the effect of molecular viscosity is negligible as well as the flow is fully turbulent. Therefore standard k- ε model is used for fully turbulent flow.

$$\rho \frac{\partial k}{\partial x_i} = \frac{\partial}{\partial x_i}[(\mu + \frac{\mu_t}{\sigma_k})\frac{\partial k}{\partial x_i}] + G_k + G_b - \rho\xi - Y_m + S_K \quad \text{……... (4)}$$

$$\rho \frac{\partial \varepsilon}{\partial x_i} = \frac{\partial}{\partial x_i}[(\mu + \frac{\mu_t}{\sigma_k})\frac{\partial \varepsilon}{\partial x_i}] + \frac{c_{1\varepsilon}\varepsilon}{k}(G_k + C_{3\varepsilon}G_b) + C_{2\varepsilon}(\frac{\varepsilon^2}{k})\rho + S_\varepsilon \quad \text{……... (5)}$$

Where G_b shows the generation of turbulent kinetic energy cause of buoyancy. G_k represents the generation of turbulent kinetic energy due to the mean velocity gradients. Y_m Represents the contribution in the fluctuation of dilatation in compressible turbulence to the overall dissipation rate.

2.3 Boundary Condition

The superstructure model is analyzed in computational domain as illustrated in Figure 4 and shown in **T**able 2.

1. V_w, Represents the inlet velocity of air at the entry of computational domain.
2. Since the interest of the results is inside the computational domain. So a no-slip adiabatic wall boundary condition is applicable on side walls on the domain.
3. Since there is a domain, so at the exit of it outlet boundary condition is applied.
4. V_e Exit velocity of air from the funnel. 5. (-)V_i a negative inlet velocity is imposed on the gas turbine intakes

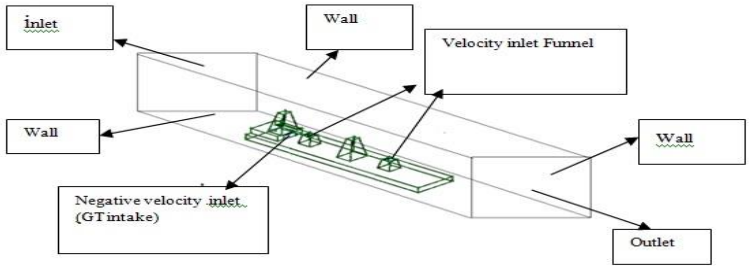


Figure 3. Computational domain with boundary conditions.V_i

Table 2: Specifications of boundary conditions

S.NO	Condition	Velocity Ratio	Boundary conditions		
			V_w	V_e	V_i
1	With air injection through exhaust and suction through intake	$K = V_e/V_w$	Inlet velocity of air(m/s)	Exit velocity of plume(m/s)	Inlet velocity of GT
		0.5	10	5	-7.5

2.4 Meshing and computational fluid analysis

The computational domain was discretized in volumes and then meshing was done as shown in Figure 5. The meshing of computational domain was done in such a way so that its quality criteria specified in FLUENT is satisfied. Coarse grid was used in the computational domain to solve the plume path and flow pattern. Solution –adaptive refinement technique was undertaken and based on velocity gradient, mesh was adapted. It is very easy to identify the flow pattern near the superstructure because of its flow feature. Total pressure deficit show the wake region, and high velocity regions are identify by the jets. The grid adaptation feature of FLUENT through pressure and velocity parameters was used for the adaption of grid. Using volume adaption, the grid was enhanced and the refining improved using the change in volume between the cell and its neighbor. To get the initial solution, a coarse mesh with 341856 tetrahedral cells was used. Thereafter, the wake region was identified by total pressure plot on different

planes and the increased velocity and its direction in the region of jet were identified by velocity vector plot in horizontal and vertical planes. Now the adapted mesh had 750467 in the region of interest. Simulation of coarse mesh took more time to converge the solution. The converged solution provides the results at subsequent section.

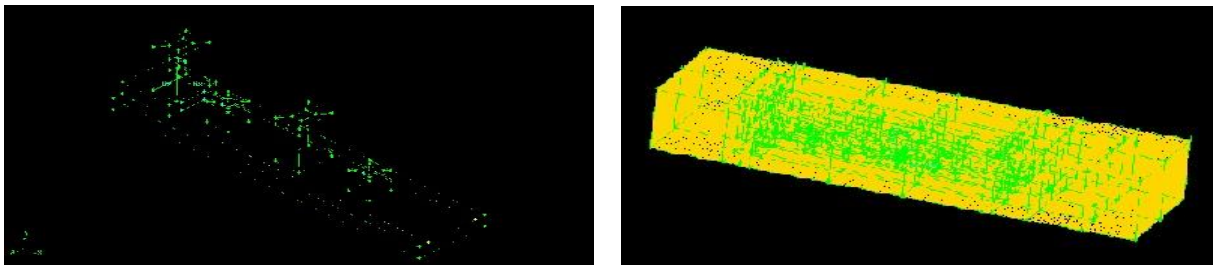
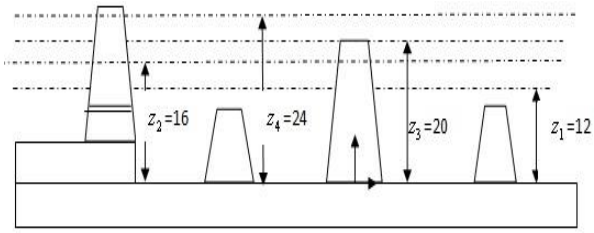


Figure 4. Meshing of computational domain

3. Results and Discussion

It is very easy to understand the behavior of plume path through flow visualization of the smoke from CFD simulation. Velocity vector provides wide information of Mass less particles, which were released from the funnel exit. The converged solution of the numerical simulation enabled the magnitude of velocity and its direction for different configuration of opening. The flow was analyzed at four horizontal and four vertical planes as shown in Figure 6 and Figure 7.



Cl0=Y= 0, Cl1=Y= -2, Cl2=Y= -4, Cl3=Y= -6
Figure 5. location of XZ planes at different Y

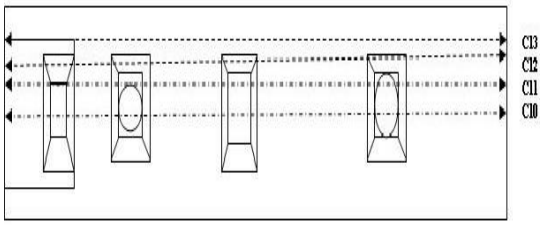


Figure 6. location of XY planes at different height (z)

3.1 Case.1: Mast without opening

To understand the effect of mast without opening and gas turbine intake interaction on the plume path, a simplified superstructure has been studied using CFD. The velocity vector of air (v=10 m/s) in X-direction, plume (v=5m/s) in z-direction in XZ plane at CL0 (central plane) with forward mast and funnel, velocity magnitude are shown in Fig.8. The intake of plume in the gas turbine decreases the efficiency. Since the momentum of plume from forward funnel is less that's why the negative pressure generated due to forward mast and gas turbine are dominant to bends the plume towards forward mast and gas turbine prone to do.

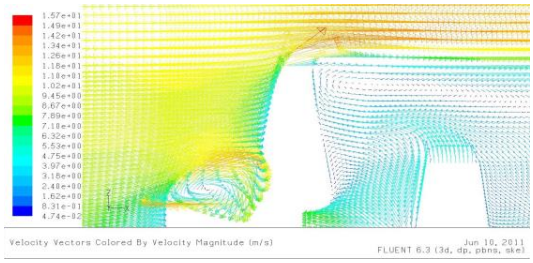


Figure7. Velocity vector plot on CL0 (Mast without opening)

Case 2: Mast with opening (figure 2 a)

The analysis of velocity vector plot on XZ plane has done for case 2(a) i.e. straight opening at 45˚. As the depth of opening size increases in downward direction the mass flow rate increases. The results showed in Figure 8 (CL0-CL1). The momentum of air pushes the plume in forward direction to the air velocity. But as the distance increases from the vertical central plane (CL0) the effect of air velocity reduces and low pressure region dominated as shown in Figure 8 (CL2). Similar results also seen in Figure 8 (CL3).Since the opening shifted towards the gas turbine intake so the impact of dispersion of plume nearby the turbine is more and the recirculation zone of plume shifted towards the edge of model superstructure as shown in Figure 9 (Z1-Z2). As the plane move away from the deck, effect of air momentum reduces and plume tends to bend towards the mast as shown in Figure 9 (Z3-Z4)

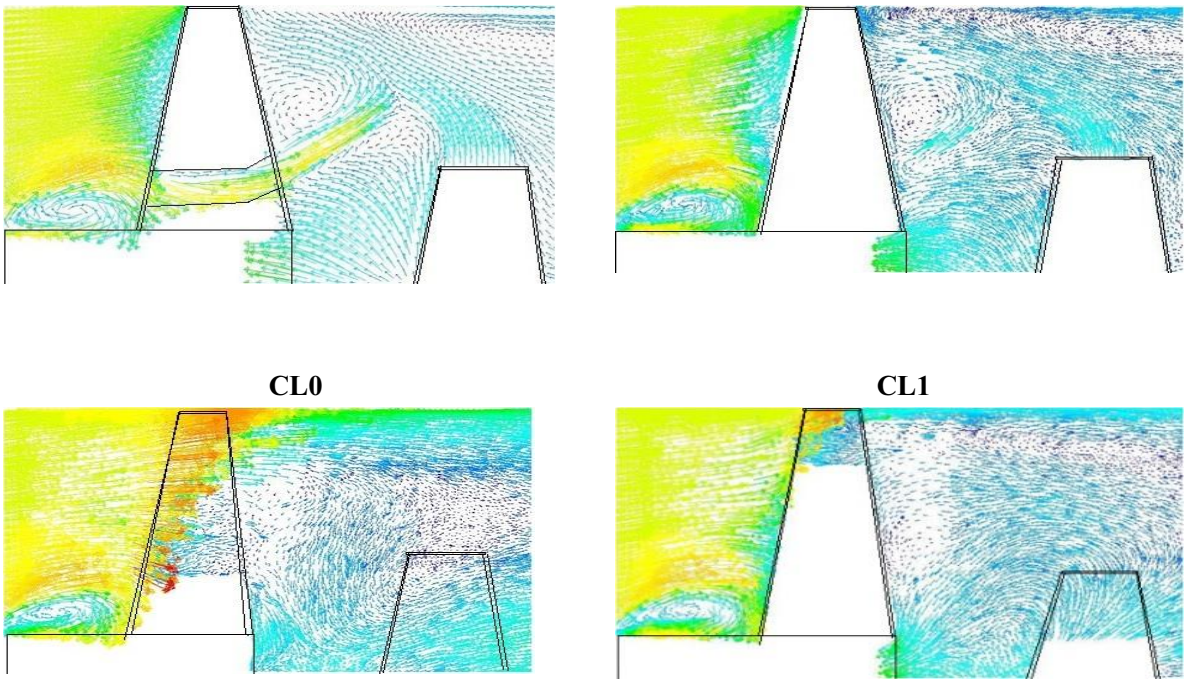


Figure 8. Velocity vector plot of Case.2(a) in XZ plane at different Y(CL0,CL1,CL2,CL3)

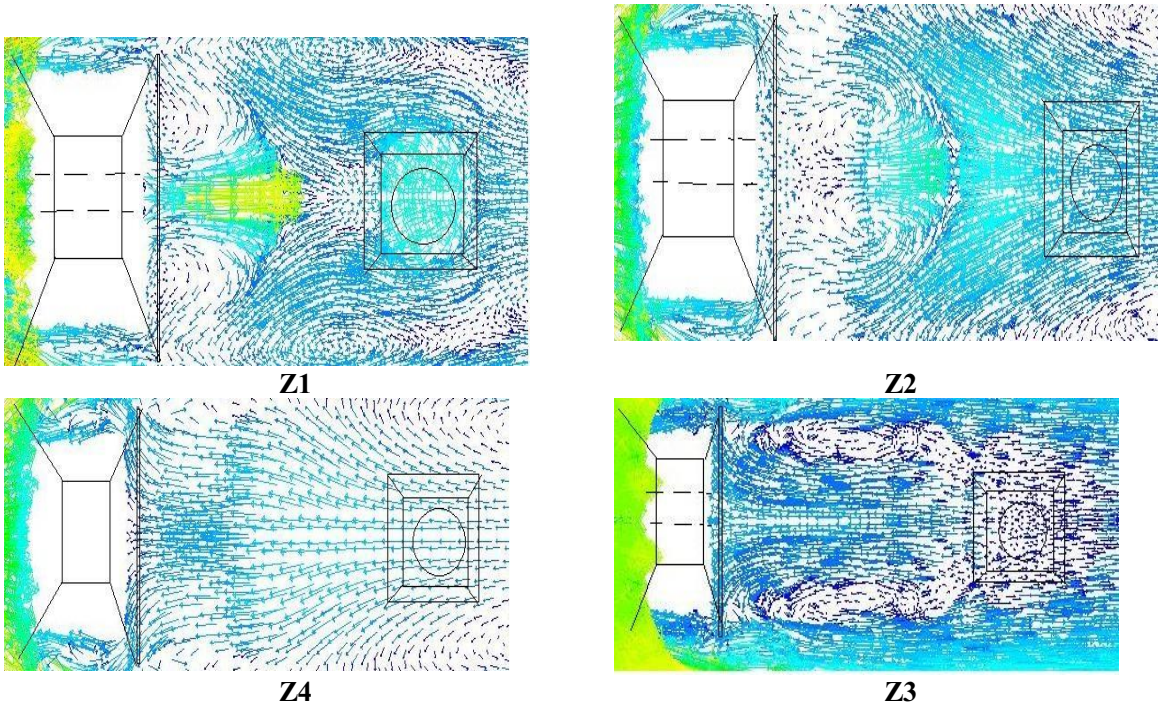
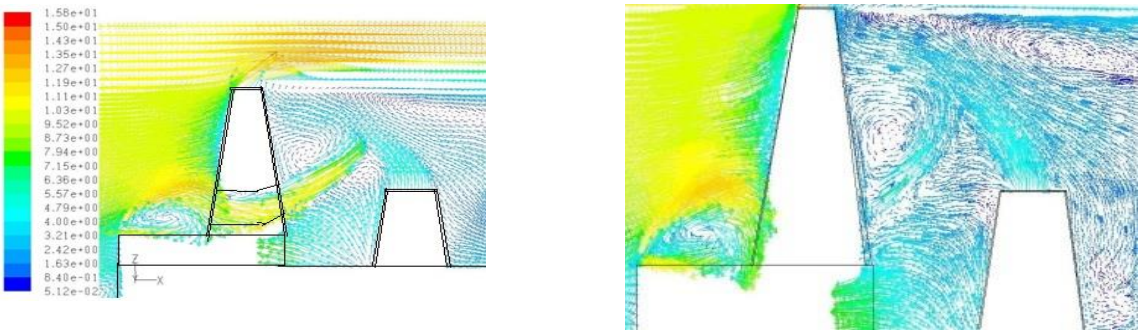


Figure 9. Velocity vector plot of Case.2(a) in XY plane at different Z i.e. (Z1,Z2,Z3,Z4)

Case 2: Mast with opening (figure 2 b)

The study reveals as the depth of opening increases, Cross-section reaches near the deck, mass flow rate got choked. The leading edge of deck form a boundary layer near the cross-section which prevents the entrance of air through opening as shown Figure 10 (CL0 to CL3). The influence of air shifted towards the deck so it has better impact near the gas turbine. But as the plane shifted away from central position velocity vector shows the more clearance near the gas turbine as shown in Figure 10 (CL2-CL3).From the velocity vector at the exit of opening, it is seen that the momentum of air shifted towards deck so the dispersion of plume is significant as shown in Figure 11 (Z1-Z2). This air momentum clears the plume near gas turbine region. It is more than above mentioned cases. As the XY plane shifted towards top side of the mast, it is observed that the impact of velocity reduces so the wake region dominates on the top side of mast. Due to this the plume bends towards the mast as shown in Figure 11 (Z3-Z4).



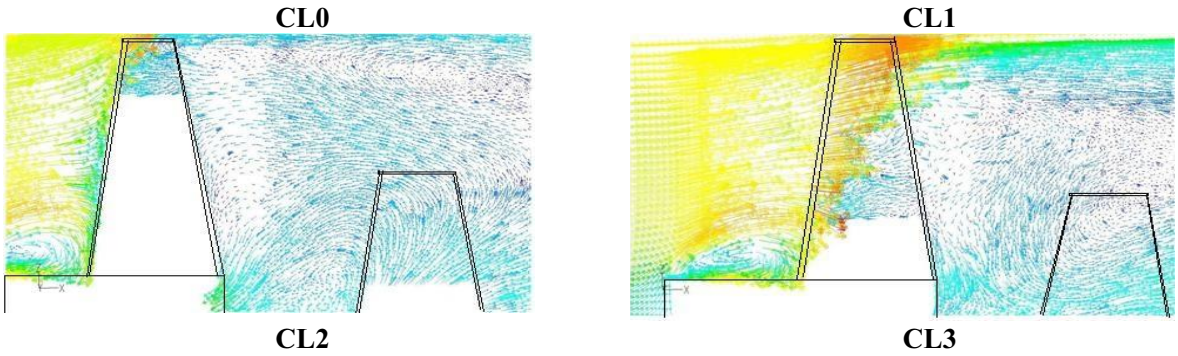


Figure 10. Velocity vector plot of Case.2(b) in XZ plane at different Y(CL0,CL1,CL2,CL3)

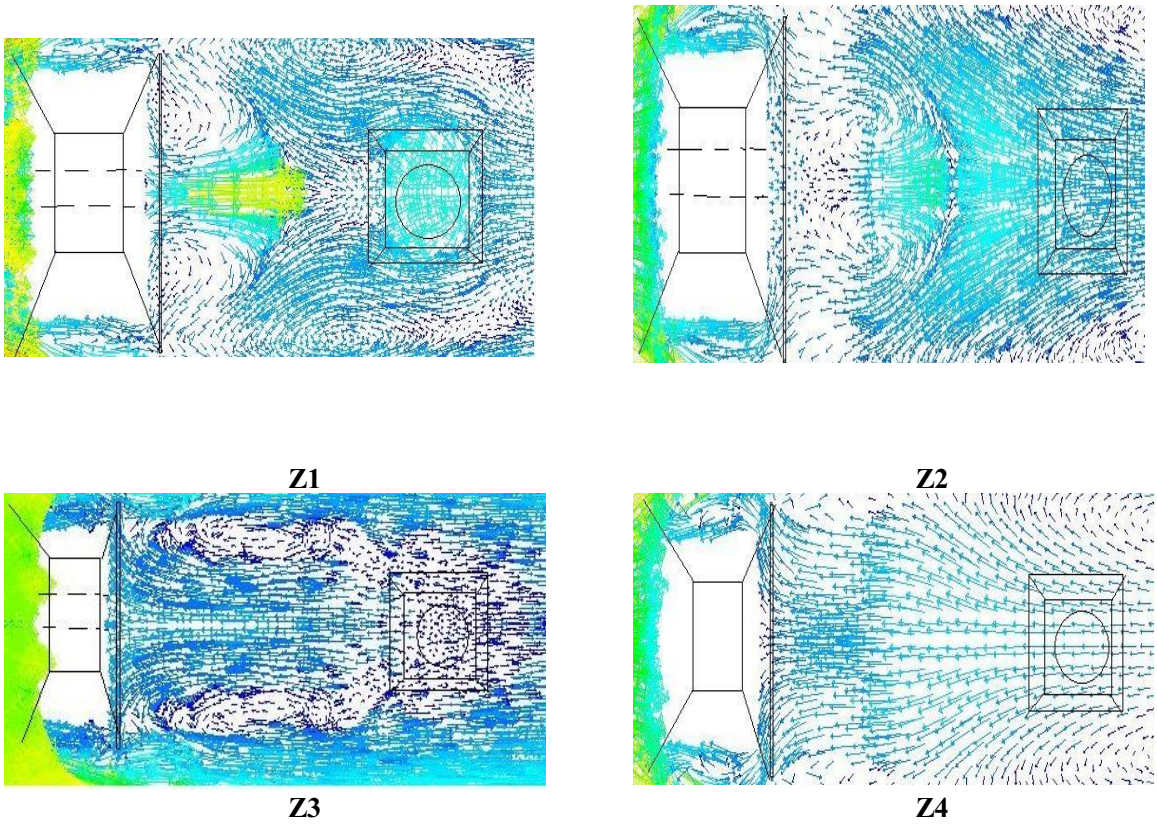


Figure 11. Velocity vector plot of Case.2(b) in XY plane at different Z i.e. (Z1,Z2,Z3,Z4)

4. Conclusion

Present work depicts that different size of opening in the mast has taken at constant velocity ratio (K= 0.5) of smoke at funnel exit and gas turbine inlet velocity (V_i = 7.5 m/s). The following conclusions are explored. Negative velocity at the inlet of gas turbine has significant effect on the plume path. Case 2(a) provides the better solution than other cases by reducing oxygen depleted air and clear the deck. This solution reduced the wake region which is generated due to mast. The recirculation of plume is also reduced, subsequently disperse the plume sideways. Since the velocity ratio is less so ascending height of plume reduced cause of less chance of infrared signature.

Acknowledgement: The present study was sponsored by Naval Research Board (NRB). The authors are very grateful to them

REFERENCES:

[1] DW Brayer and RC Pankhurst (1971*) pressure probe methods for determining wind speed and flow direction* (national physical laboratory) 126pp
[2] Davies ME Cole LR and O'Neill PGG 1979 *Wind tunnel investigation of the temperature field due to the hot exhaust of power generation plants on offshore platforms* (UK: National Maritime Institute) NMI R-58(OT-R-7935)
[3]Taylor K Smith AG (1997) *CFD prediction of exhaust plumes and interaction with superstructures. Application of fluid dynamics in the safe design of topsides and superstructures* London: Institute of Marine Engineers; 56–61.
[4] Micheal KJohns Val Healy J (1989) The air wake of a DD 963 class destroyer *Naval Eng J*, ASNE : 36–42
[5] Jin E Yoon J Kim Y (2001) *A CFD based parametric study on the smoke behavior of a typical merchant ship* PRADS'01, Shanghai, 459–65.
[6] Kulkarni PR Singh SN and Seshadri V (2005) *Behaviour of a ship funnel in the wake of a bluff body MARINE CFD* In: RINA's 4th int conference on marine hydrodynamics, Southampton,UK,30–31 . 115– 24.
[7] Kulkarni PR Singh SN and Seshadri V (2005) *Comparison of CFD simulation of exhaust smoke–superstructure interaction on a ship with experimental data*. In: Proc of naval platform technology seminar – (NPTS-05), Singapore, 17–18, 150–72.
[8] Kulkarni PR Singh SN and Seshadri V (2005) Flow visualization studies of exhaust smoke–superstructure interaction On naval ships. *Naval Eng J ASNE*; **117(1):**41–56
[9] Kulkarni PR Singh SN Seshadri V (2005) Experimental study of the flow field over simplified superstructure of a ship. *Int J Maritime Eng* IJME **Part A3**:19–42.
[10]Kulkarni PR Singh SN and Seshadri V (2005) Study of smoke nuisance problem on ships – a review. *Int J Maritime Eng, IJME* Part **A2**:27–50
[11] Kulkarni P.R Singh S.N Seshadri V and Nagababu G. (2006) *Study of Exhaust Smoke Ingress into GT Intake in Naval Ships* International Conference in Maritime Hydrodynamics, Visakhaptnam, 5-7 , 847-858
[12] Seshadri V and Singh S.N. (2005) *Wind tunnel studies to obviate the problem of unwarranted rise in air intake* internal report applied mechanics IIT Delhi.
[13] Seshadri V Singh S.N and Kulkarni PR(2006) A study of the problem of ingress of exhaust smoke into the GT intakes in naval ships *Journal of Ship Technol*; **2(1)**:22–35
[14] Vijayakumar R Seshadri V Singh S.N and Kulkarni P.R (2008) A Wind Tunnel Study on the Interaction of Hot Exhaust from the Funnel with the Superstructure of a Naval Ship *Oceans Kobe*, Japan 18-21,
[15] Vijayakumar R Singh SN and V Seshadri (2014) CFD prediction of Hot exhaust from the funnel of a Naval Ships in presence of Ship Superstructure *International Journal of Marine Engineers, Proceedings of Royal Institute Naval Architecture* . Vol. **156**, pp 1-24.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/327052241

Dielectric studies of 0.1 La and Pb doped SrBi2Nb2O9 ferroelectric ceramic

Conference Paper · May 2006

CITATIONS
0

READS
11

3 authors:



Vaibhav Shrivastava
Shiv Nadar University
15 PUBLICATIONS **99** CITATIONS
SEE PROFILE



A.K. Jha
Ambedkar Institute of Advanced Communication Techn…
95 PUBLICATIONS **880** CITATIONS
SEE PROFILE



R. G. Mendiratta
ABES Engineering College
112 PUBLICATIONS **1,975** CITATIONS
SEE PROFILE

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

Microwave synthesis augmenting dielectric response of materials and assisting in nano particle formation View project

PhD Thesis View project

All content following this page was uploaded by Vaibhav Shrivastava on 15 October 2018.

The user has requested enhancement of the downloaded file.

National Seminar on Advancement in Electroceramics, DRDO Pashan Pune on 5-6th May'2006

Effects of La and Pb doping on structural and dielectric properties of $Sr_{1-x}A_xBi_2Nb_2O_9$ (A=La, Pb and x=0.1) ferroelectric ceramic

Vaibhav Shrivastava, A.K.Jha[#] and R.G.Mendiratta[*]

Department of Applied Physics, Delhi College of Engineering, Delhi-110042
**Department of Physics, ABES Engineering College, Ghaziabad-201001*

In this paper, structural and dielectric studies of $Sr_{1-x}A_xBi_2Nb_2O_9$ (A=La, Pb and x=0.1) ceramics are reported. Tetragonal strain decreases on introduction of both lanthanum and lead. Addition of lead decreases Curie temperature by 5 ^oC while for sample with lanthanum content it decreases by 41 ^oC. The sample with x=0.0 shows a rapid fall in dielectric constant values up to 1 KHz and above 100 KHz. The lanthanum and lead doped samples show reduced dielectric loss values.

INTRODUCTION

Aurivillius ceramics are the potential materials for many applications (ref. 1-3) due to their excellent fatigue characteristics and high Curie temperature. These ceramics have been reported to have orthorhombic unit cell (ref. 4), which is highly useful for manufacturing ferroelectric memories with high remnant/ spontaneous polarization and piezoelectric coefficients (ref. 5, 6). These ceramics suffer from high dielectric loss due to the evaporation of bismuth oxide during sample preparation (ref. 7, 8), which limits its use for the above said applications. Among Aurivilllius ceramics $SrBi_2Nb_2O_9$ (SBN) is a material with less distorted octahedron structure. In the present work, the same structure have been studied for its structural and dielectric properties with doping of lanthanum and lead. In the present work, the studied compositions are $Sr_{1-x}A_xBi_2Nb_2O_9$; A= La, Pb and x=0.0 & 0.1.

EXPERIMENTAL

Samples were prepared using solid-state reaction method taking $SrCO_3$, La_2O_3, PbO, Bi_2O_3 and Nb_2O_5 (all from Aldrich) in stoichiometric proportions. Mixtures were calcined at 900 ^oC in air for 2 hours. Powdered samples were mixed with two weight-percent solution of polyvinylalcohol (Aldrich) and then molded into disc shape pellets by applying a pressure of 270 MPa. Pellets were sintered at 1150 ^oC for 2 hours. X-ray diffractograms were recorded for all the calcined and sintered samples on Philips X-ray diffractometer PW 1710 using Cu$K\alpha$ radiation of wavelength 1.54439 Å in the range $10^o \leq 2\theta \leq 70^o$ at a scanning rate of 0.05^o / second. All pellets were coated using silver paste and cured at 600 ^oC for half an hour. Dielectric measurements were taken on an HP 4192A Impedance Analyzer

operating at oscillation level of 1 V and a frequency of 100 KHz was chosen for Curie temperature measurements.

RESULTS AND DISCUSSION

Fig.1 shows the X-ray diffractograms of all the sintered samples. The formation of desired layered perovskite phase is confirmed in all the samples. From the observed intensity ratio of different peaks, it is concluded that almost similar phase is formed in all samples. In an earlier report published the maximum intensity peak (105) is reported to be the set of pseudo tetragonal phase and the peak (115) is reported to be orthorhombic (ref. 10). Moret *et al* (ref. 11) have reported about the coexistence of orthorhombic and tetragonal phases in these ceramics. The lattice parameters a and c were calculated from the obtained d-values using standard deviation method. The introduction of smaller lanthanum yields a decrease of 0.23 % in lattice parameter c and of 0.16 % in lattice parameter a. The tetragonal lattice strain (c/a) (ref. 12, 13) has been determined for all the three samples (Table 1). The effect of this parameter on Curie temperature is discussed later.

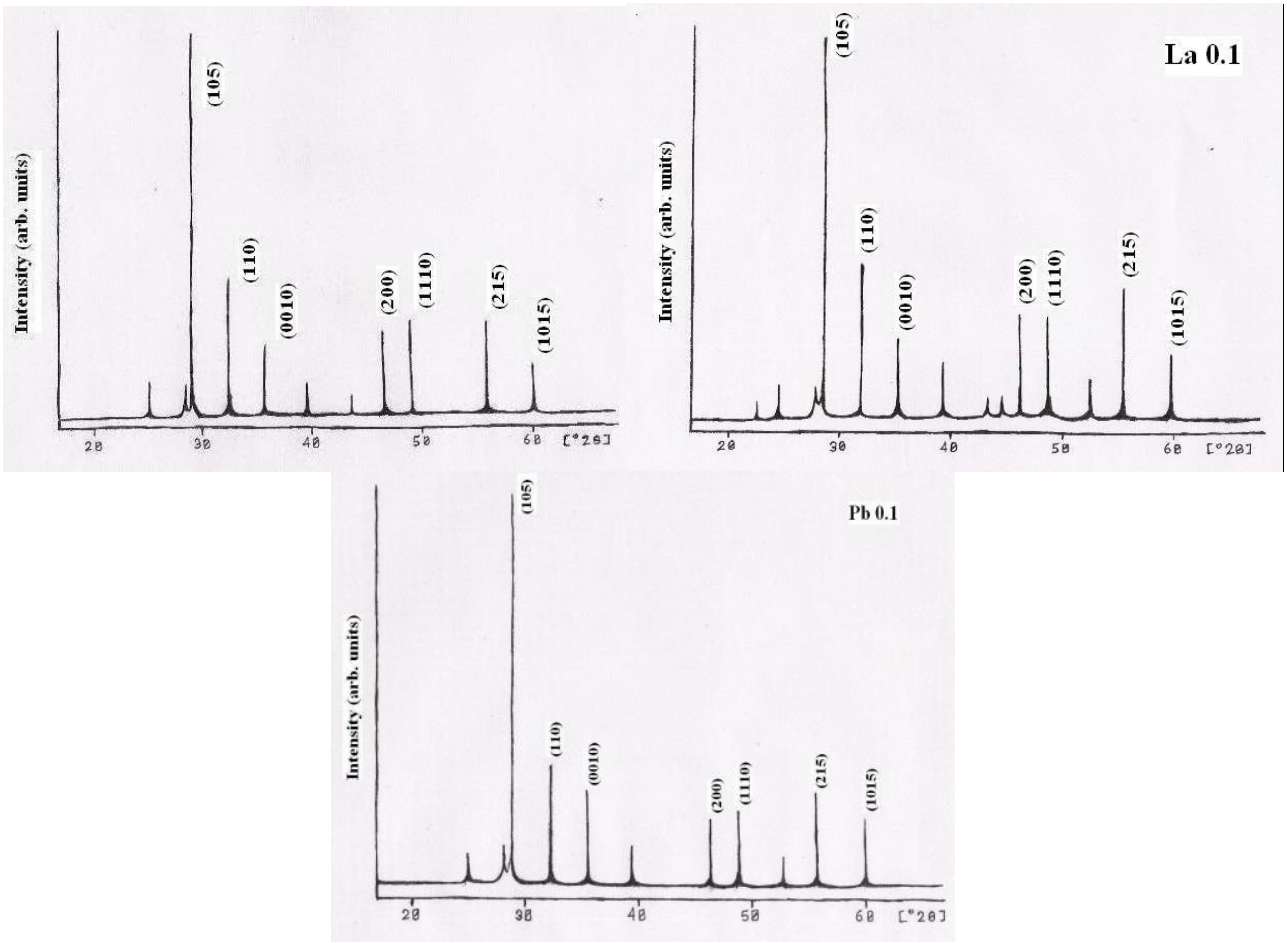


Fig.1 X-ray diffractograms of undoped SBN and 0.1 lanthanum and lead doped samples.

TABLE 1 Tetragonal strain (c/a), ionic radius and bond strengths of elements

S.N.	Composition	c / a	Ionic radius with valence state[14]	Bond strengths with oxygen (KJ/mole) (ref. 14)
1	$SrBi_2Nb_2O_9$	6.46	Sr^{2+} 1.44 Å	Sr^{2+} 426
2	$Sr_{0.9}La_{0.1}Bi_2Nb_2O_9$	6.43	La^{3+} 1.36 Å	La^{3+} 799
3	$Sr_{0.9}Pb_{0.1}Bi_2Nb_2O_9$	6.45	Pb^{2+} 1.49 Å	Pb^{2+} 386

Fig.2 shows the variation of dielectric constant (ε) values as a function of frequency ranging from 50 Hz to 1 MHz at room temperature for all the studied samples. The lead doped sample shows almost constant value of dielectric constant over the frequency range. The undoped sample and lanthanum doped samples show a rapid decrease in dielectric constant values near 1 KHz and relatively slower decrease at higher frequencies. In the undoped sample, steep decrease in dielectric constant values with increase in frequency up to 100 KHz is attributed to the relaxation and resonance absorption phenomena (ref. 15).

The range of frequency near 1 KHz is known to show relaxation absorption in dielectric constant values because of space charge polarization. The space charge is known to saturate up to 1 KHz; however, in a few ceramics its complete saturation may extend up to 100 KHz. The introduction of lanthanum is known to increase dielectric constant (ref. 16) compared to that of the undoped sample because of its high ionic polarazability. The doping of two lanthanum ions onto strontium sites would introduce two units of free positive charge in the structure, i.e., one strontium vacancy. Therefore, doping of lanthanum enhances the creation of strontium vacancies in the structure to maintain charge neutrality (ref. 9). The absence of pronounced space charge polarization in the doped samples is possibly because of charge neutrality in lanthanum doped sample and similar valence state of lead with that of strontium respectively (ref. 17).

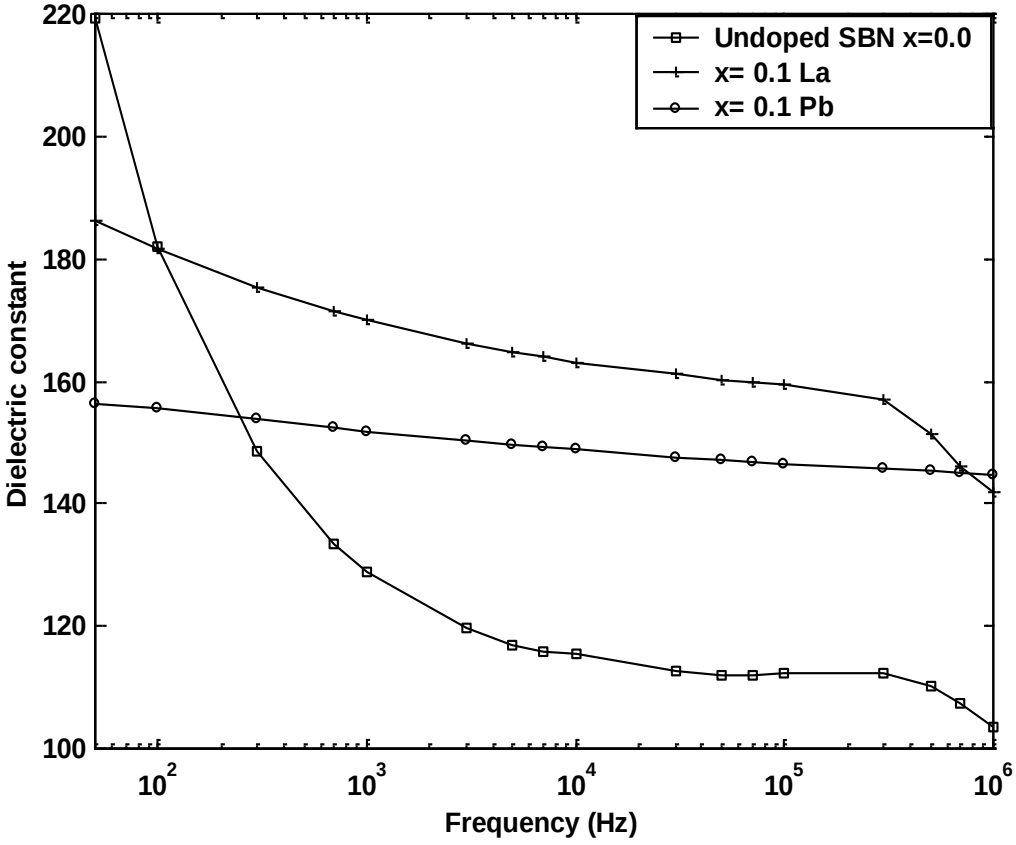


Fig.2 Dielectric constant versus frequency behavior of ceramics

In Fig.3, dielectric constant versus temperature behavior of the studied compositions is shown; these observations have been plotted using Microcal Origin 5.1. These measurements were taken at 100 KHz frequency above which space charge and orientational polarizations cease off and only ionic and electronic contributions persist (ref. 15). These ceramics are known to have single-phase ferroelectric to paraelectric phase transitions corresponding to Curie temperature where dielectric constant is maximum ($\varepsilon_{max.}$) (ref. 18, 19). The Curie temperature decreases on introduction of lanthanum and lead; however, negligible decrease of 5 ^{o}C (undoped SBN- 437 ^{o}C, 0.1 SPBN- 432 ^{o}C) on introducing lead is possibly due to same valence state and nearly equal ionic radius. In addition, the dielectric peak is broadened on introduction of lanthanum, which is known to be observed due to either as defect induced (ref. 20) or relaxor type phase transition (ref. 21) where formation of micro polar regions occurs and each of such regions has its own transition temperature. The observed change in Curie temperature is attributed to change in tetragonal lattice strain (ref. 12). In ferroelectric materials, lanthanum doping is known (ref. 22) to reduce lattice distortions by stabilizing stable non-ferroelectric phases e.g. tetragonal phase. The same is observed in the present work as indicated from tetragonal lattice strain (c/a) values in Table 1.

Dielectric loss versus temperature measurements are shown in Fig. 4. The dielectric loss values of all the three samples are nearly same up to 350 ºC beyond which remarkable changes are observed. Above this temperature, the lead doped sample shows maximum loss and lanthanum doped sample shows low loss values. The undoped sample behaves in an intermediate manner. The reason for observed loss behavior of doped sample could be the high solid solubility of lanthanum (ref. 14) and its non-volatile nature, due to which possibly this results in low dielectric loss values at high temperatures (ref. 23). The lead doped sample shows higher loss values possibly because of the defects produced in the system due to the evaporation of volatile lead at high temperatures.

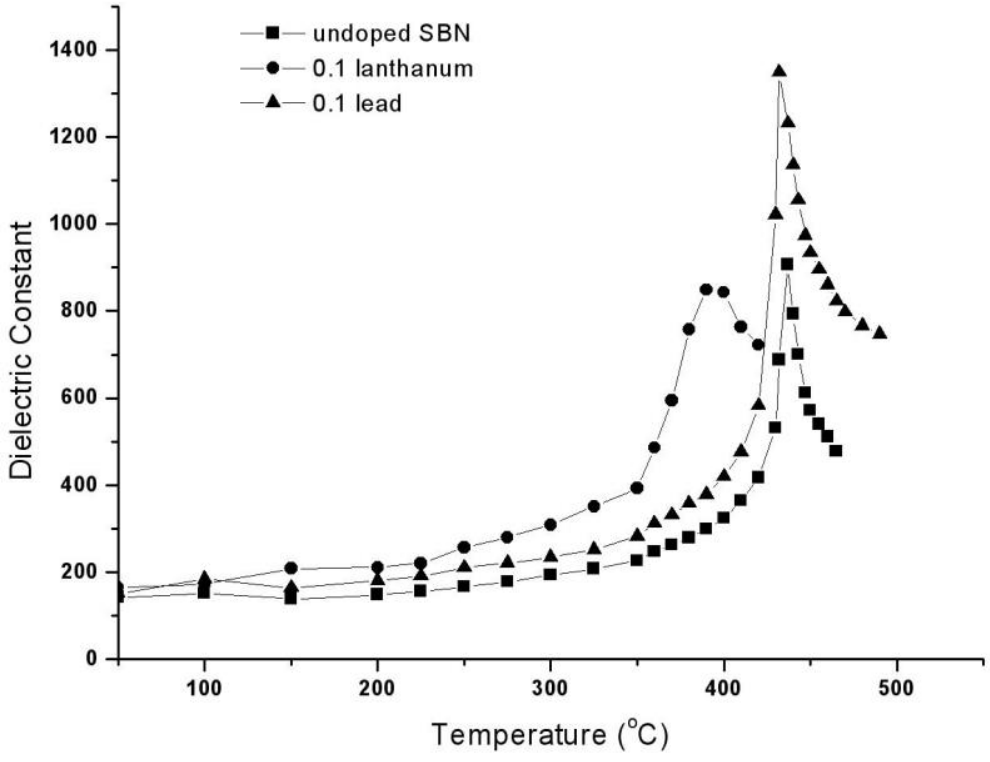


Fig.3 Dielectric constant versus temperature plots

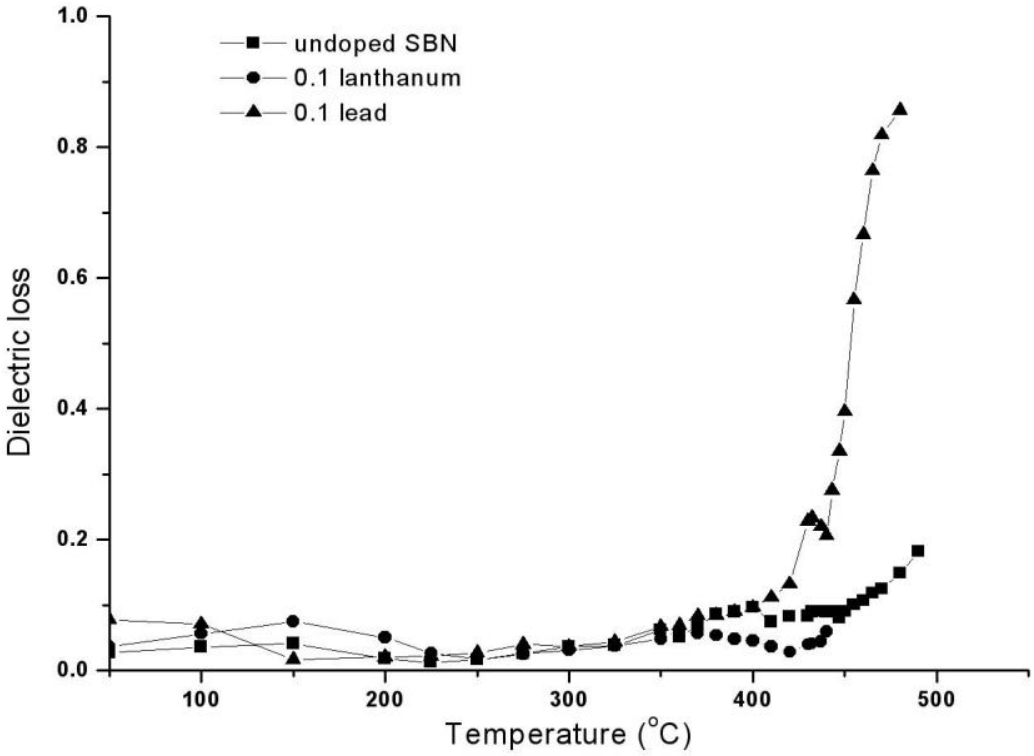


Fig.4 Dielectric loss versus temperature plots

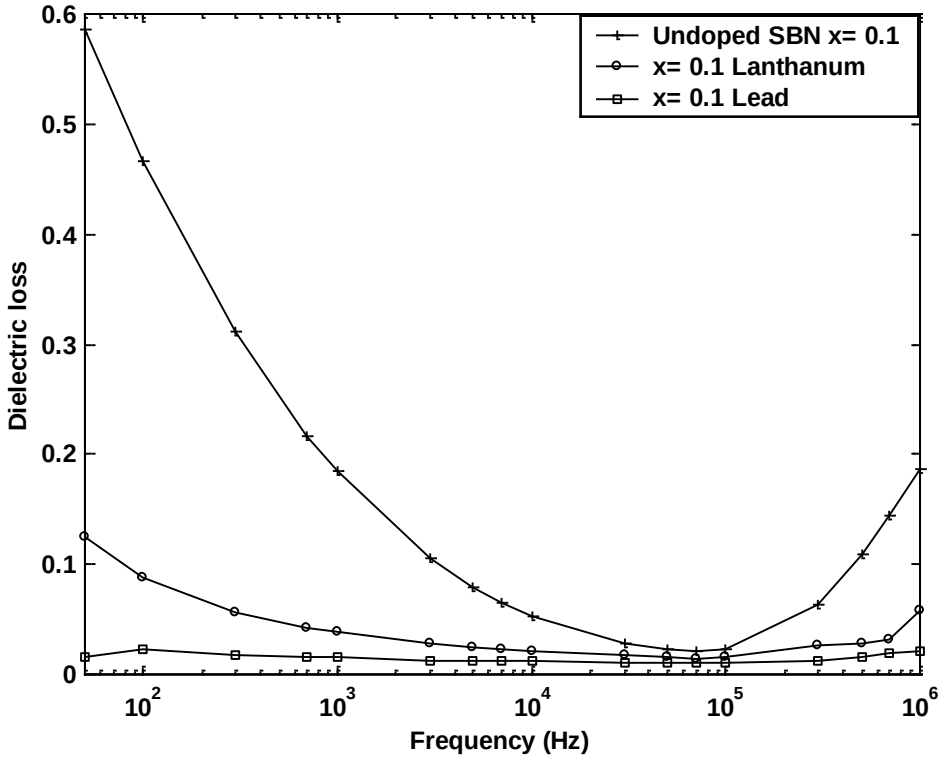


Fig.5 Dielectric loss versus frequency behavior of all samples

Fig.5 shows the dielectric loss as a function of frequency at room temperature. The undoped SBN sample shows large loss which decreases with frequency taking a dip around 80 KHz and thereafter increases again; which is known dielectric dispersion behavior (ref. 15). Among the doped samples, the lead doped sample shows low loss values throughout the frequency range. The lanthanum doped sample shows dielectric loss behavior similar to that of undoped sample though of much lesser magnitude. The increase in loss values for the undoped sample beyond 100 KHz is attributed to resonance losses (ref. 15). The dielectric loss behavior for samples with large space charge is explained by relaxation loss mechanism, which responds at lower frequencies as ion-jump/ electron hopping (<100 KHz) (ref. 24). The resonance losses respond at higher frequencies (>100 KHz) as frequency of applied field matches with the natural frequency of vibration of an ion/electron (ref. 25). The same is observed in Figs.2 and 5 where a rapid fall in dielectric constant and dielectric loss values with increase in frequency up to 1 KHz is observed. The lanthanum and lead doped samples show reduced loss values beyond dip frequency because of depletion of oxygen vacancies by doped element and high frequency of resonance.

CONCLUSIONS

The doped samples show better structural and dielectric response. The doping of lanthanum results in the formation of similar tetragonal crystal system and room temperature dielectric loss is found to be lower for lanthanum doped sample as compared to undoped sample. The broadened dielectric response with temperature offers a range of temperature for the phase change. The lead doped sample is best for high frequency applications as it shows low dielectric loss throughout the frequency range up to 1 MHz. Additionally, this sample offers highest dielectric constant at Curie temperature and does not show frequency relaxation up to 1 MHz.

REFERENCES

1. B.Aurivillius, Arki kemi 1, 463 (1949); 1, 499, (1949); 2, 579 (1950)
2. C.A.P. de Arauzo, J.F.Scott, Science, 246 (1989) 1400
3. A.I.Kingon, Current Opinion in Solid State & Materials Science, 1999, Elsevier, p.39
4. E.C.Subbarao, Jour. of Chem. Phys., 34, 1961
5. K.Shoji, M.Aikawa, Y.Uehara, K.Sakata, Japanese Jour. of Applied Phys., 37, 1998, 5273
6. A.I.Kingon, Nature, Vol.401, October 1999, 658
7. V. Shrivastava, A.K.Jha, R.G.Mendiratta, Solid State Communications, Vol. 133, Issue 2, January 2005, p.125
8. V. Shrivastava, A.K.Jha, R.G.Mendiratta, Physica B, Vol. 371, Issue 2, 31 January 2006, p.337
9. M.J.Forbess, Y.Wu, S.Seraji, C.P.Nguyen, G.Z.Cao, Applied Physics Letters, Vol.76, No.20, May 2000
10. D.Y.Suarez, Jour. of Materials research, Vol.16, No.11, November 2001, 3139
11. M.P.Moret, R.Zallen, R.E.Newnham, P.C.Joshi, S.B.Deshu, PRB, Vol. 57, No.10, March 1998 (A)

12. *Ferroelectric Materials & their applications* by Yuhuan H. Xu, Elsevier Science Publishers B.V., 1991, p.131
13. *Ferroelectric Crystals* by F.Jona, G.Shirane, Pergamon Press, Oxford, 1962, p.276
14. CRC Handbook of Chemistry and Physics edited by David R. Liden, 81st Edition, 2000-2001
15. *Principles of Electronic ceramics* edited by R.C.Buchnan, Marcel Dekkar, New York, 1991,p.38-39
16. G.H.Haertling, Jour. of American Ceramic Society, 82, 4, 797-818, 1999
17. *Phys. of Solid Dielectrics* by Ion Bunget, Mihai Popescu, Elsevier, Amsterdam, Oxford, New York, 1984, p.345
18. E.C.Subbarao, Physical Review, Vol.122, No.3, May1961, 804
19. G.A.Smolenskii, V.A.Isupov, A.I.Agranovskaya, Fiz. Tverdogo Tela 1, 169, 1959
20. Y.Ding, J.S.Liu, J.S.Zhu, Y.N.Wang, Jour. of Appl. Physics, Vol.91, No.4, February 2002, p.2255
21. L.E.Cross, Ferroelectrics, 76, 241, 1967
22. *Ferroelectric Materials & their applications* by Yuhuan H. Xu, Elsevier Science Publishers B.V., 1991, p.165
23. E.C.Subbarao, Colloids and Surfaces A, Physicochemical Engg. Aspects 133, 1998, 3-11
24. *Phys. of Solid Dielectrics* by Ion Bunget, Mihai Popescu, Elsevier, Amsterdam, Oxford, New York, 1984, p.244
25. Ferroelectric Devices by K.Uchino, Marcel Dekkar Inc., NewYork, 2000, p.93

View publication stats

Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-019-01419-7

ORIGINAL RESEARCH

Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets

Akshi Kumar[1] · **Geetanjali Garg**[1]

Received: 21 February 2019 / Accepted: 1 August 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract
Sarcastic expressions tend to flip the polarity of posts when being analyzed for sentiments. Detecting sarcastic tone, which conveys a sharp, bitter, or cutting expression, remark or taunt in natural language is tricky even for humans, making its automated detection more arduous. Computational models for sarcasm detection have often relied on the content of utterances in isolation whereas using contextual information definitely improves it. This work is a preliminary to understand the what, how and why of using context in sentiment analysis. The concept of 'context in use' is described by exemplifying content-based local and global context to predict sarcasm in user-generated social textual data. In this research Twitter data of benchmark SemEval 2015 Task 11 and nearly 20 k posts from Reddit are classified as sarcastic or non-sarcastic using three predictive learning models. The first model is based on the conventional Term Frequency-Inverse Document Frequency (TF-IDF) weighting which is trained over three classifiers, namely the Multinomial Naïve Bayes, Gradient Boosting and Random Forest and Ensemble Voting is utilized to generate the output. In the second model, a combination of semantic (sentiment) and pragmatic (punctuation) features are considered to model the context along with the top-200 TF-IDF features and results are observed using five baseline classifiers (Decision Tree, Support Vector Machine, Random Forest, K-Nearest Neighbour and Multi Layer Perceptron). The final model uses deep learning, that is, the Long–Short-Term-Memory (LSTM) and its variant Bi-directional LSTM applying GloVe (Global Vectors for Word Representation) for building semantic word embeddings and learning context. The empirical study using training set and test set performance metrics (Accuracy, Recall, Precision, F1 Score) is done to compare the three learning models for sarcasm classification within two datasets, and it is observed that Bi-directional LSTM model has the highest accuracy of 86.32% and 82.91% for the Twitter and Reddit datasets respectively.

Keywords Sarcasm · Shallow classifier · Deep classifier · Social media

1 Introduction

The pervasive social web is a rich source of real-time, opinionated user-generated multimodal content available online on resources like review forums and microblogging sites. Automated text analytics using this opinion-rich data has many promising practical applications, such as sentiment analysis (Pang and Lee 2008; Pak and Paroubek 2010; Liu 2015), aggression detection (Chen et al. 2018), and rumour detection (Zubiaga et al. 2017). Sentiment Analysis has been thriving to facilitate knowledge extraction for decision making within this social setting. It is often described as typical text classification task that indispensably relies on the understanding of the human language and emotions expressed via textual or non-textual content. But human expressions are extremely complicated as statements can convey a wide range of emotions, and requires contextualization of words to bridge the gap between what you've experienced and what you are trying to say. Accurately determining polarity from the legit literal text is imperative for real-time sentiment analysis. However, the growing use of emblematic language markers such as punctuations (awesome!!!!!!), emojis (,, ❤),wordplay (greatttttt for great), creative spellings (2 day for today), Internet slangs (OMG for 'Oh My God') (Satapathy et al. 2017), code-switching (Nilep 2006) and code-mixing (Barman et al. 2014; Das and Gambäck 2014) are some commonly seen phenomenon which increase the complexity of computational linguistics to analyze the social media content, thus making sentiment analysis a non-trivial

✉ Akshi Kumar
akshikumar@dce.ac.in

[1] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India

Published online: 12 August 2019



Fig. 1 Misinterpreted sarcastic tweet

challenge with a lower prediction accuracy. Figure 1 represents an instance of a bot misinterpreting a sarcastic tweet and responding inappropriately.

Sentiment analysis in microblogs majorly deals with textual content analysis. The primary challenges to this sentiment analysis based social media intelligence include sarcasm, implicit knowledge, nature of text, thwarting and domain specificity, amongst others (Joshi et al. 2017). Tools to analyze the sentiments get trapped in the intricacies of sarcasm, irony and mixed feelings making it tricky to perceive polarity and lower the prediction accuracy. Tools and techniques to predict sarcasm in text are therefore crucial as figurative language permeates daily life and social media (Hercig and Lenc 2017). "Sarcasm is a type of sentiment where people express their negative feelings using positive or intensified positive words in the text" (Bharti et al. 2016). It is an expression representative of conflict between the apparent and the applied. For example, a tweet/post "*I love being cheated on!*" demonstrates this conflict between the obvious state of '*being cheated*' and the articulation '*love*'. The inconsistency, contrast and shifts within the polarities of sentiments validate sarcasm as a distinctive case of sentiment analysis.

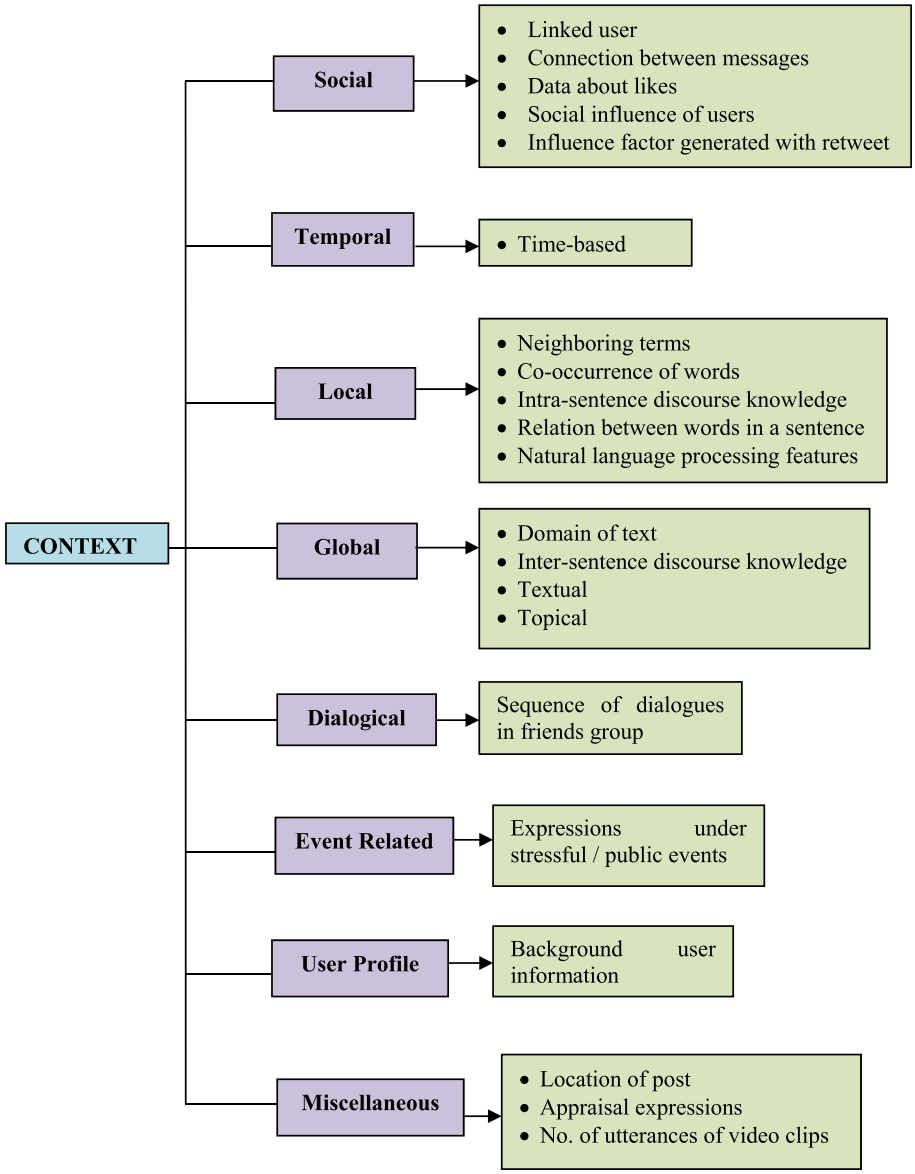
Sarcasm is topic-dependent and highly contextual; therefore, polarity shift and other contextual clues can help detect sarcasm from text by resolving the ambiguity of meaning and improve the generic sentiment classification of voluminous user-generated social textual data. Detecting sarcastic tone is very difficult to accomplish without having a sufficient knowledge of the 'context' of the situation, the specific topic, and the environment (Kumar and Garg 2019). Understanding context is one of the most difficult aspects of content moderation. Context in sentiment analysis is defined as any complementary source of evidence which can either intensify or flip the polarity of content. The accuracy of polarity classification will thus depend on a context vector and the learning model will ensure the overall decision making (classification) is more reliable. The use of various factors like dependency based rules, topic, user profile, social influence, product ontologies, lexical knowledge, discourse knowledge, textual, dialogical, appraisal expressions (e.g. attitude, engagement, graduation, etc.), opinion homophily, temporal changes in the data, number of followers, domain knowledge have been reported in literature to define context for improving the classification accuracy. The following Fig. 2 depicts the various types of 'context' defined in literature to complement sentiment analysis.

In this study, content-based local and global contextual information is used to build feature-variant predictive learning models which classify text into the sarcastic or non-sarcastic categories. Twitter data of the SemEval 2015 Task 11 and 20 k posts from Reddit is used for classification. We examine three sets of features, namely the lexical (TF-IDF based), pragmatic (punctuation-based) and the semantic (sentiment and word embedding based) for sarcasm detection. The features are used to train and evaluate different classifiers, shallow, ensemble voting and deep classifiers. The first model is based on the conventional Term Frequency-Inverse Document Frequency (TF-IDF) statistical weighting for feature extraction which is trained over three classifiers, namely the Multinomial Naïve Bayes, Gradient Boosting and Random Forest and ensemble voting is utilized to generate the output. In the second model, a combination of semantic (sentiment) and pragmatic (punctuation) features is taken with the top-200 TF-IDF features to build the context and results are observed using five shallow classifiers (K-Nearest Neighbor, Support Vector Machine, Random Forest, Decision Tree and Multi Layer Perceptron). The third model uses deep learning-based architectures, that is, the Long–Short-Term-Memory (LSTM) and its variant Bi-directional LSTM applying GloVe embedding for building contextual vectors with representative semantic word as features. The empirical analysis is done using classifier's key performance indicators (KPIs). The primary contributions are as follows:

- Understand sarcasm as a special case of context-based sentiment analysis
- Define and capture content-based local and global context using three features: lexical (TF-IDF based), pragmatic (punctuation-based), semantic (sentiment and word embedding based)
- Build a learning model for predicting sarcasm in real-time social data

Thus, the intent of this study is primarily to comprehend explicit linguistic strategies used to capture context and express sarcasm in text. This study also investigates the

Fig. 2 Types of '*Context*'



behavior of shallow classification and deep classification models to find the best model for finding sarcastic expressions in online content. The organization of the work presented in this paper is as follows: Sect. 2 confers the relevant work within the research area of automatic sarcasm detection. The next section, Sect. 3 illustrates the working of the three models. Section 4 illustrates the results and its discussion followed by Sect. 5 giving the conclusion of the research with future potential directions.

2 Literature review

The contemplation and vocalization of public opinion on online social portals is a goldmine for extracting and analyzing the content available. Twitter has been the quintessential source for mining the sentiment-rich data. 'Sentiment Analysis' was first introduced by Dave et al. (2003) and within the past 15 years the research interest within the domain has increased manifold (Pang and Lee 2008; Kumar and Sebastian 2012). Literary resources pertaining to practical and theoretical machine-learning based sentiment analysis on user-generated web content are expansive. Significant studies demonstrate sarcasm in a sentence by conflict and contrast of sentiment polarity (Camp 2012; Riloff et al. 2013; Joshi et al. 2017). Detecting sarcastic tone in natural language text is a well-recognized problem within the research area of sentiment analysis.

In 2010, researchers (Davidov et al. 2010) proposed a model of sarcasm detection in Twitter data and Amazon reviews. The authors used #sarcasm as a filter to obtain sarcastic tweets and used content words and punctuation related

features in their work and obtained F-measures of 0.55 and 0.83 for the respective datasets. Another work reported in 2011 (González-Ibánez et al. 2011) divided the twitter data into three types—sarcastic, positive and negative and used lexicon based sentiment scores such as WordNet Affect in their study. In 2013, authors (Liebrecht et al. 2013) presented an approach for classifying Dutch tweets as sarcastic using unigrams, bigrams and trigrams as features. A model which used a lexicon based approach using the assumption that sarcasm employs a contrast between positive and negative sentiments was also proposed (Riloff et al. 2013). Another study (Liu et al. 2014) incorporated the intrinsic imbalance between the sarcastic and non-sarcastic texts while modeling their feature set specifically for social media and compared their work with imbalanced classification methods. A hypothesis that Twitter texts are closer to spoken language and hence such features can be incorporated in classification model was also introduced (Barbieri et al. 2014).The researchers (Ptáček et al. 2014) worked on a similar problem but emphasized more on Czech language twitter using n-gram and Part-of-Speech features.

In 2015, a Hadoop based framework to capture real time tweets and process it with a set of algorithms which identified sarcastic sentiment efficiently was proposed (Bharti et al. 2015). In the same year a user-centric approach for detecting sarcasm by designing a behavioral modeling framework fine-tuned for sarcasm was also proposed (Rajadesingan et al. 2015). Further, a contextual approach to detect sarcasm by including information about the author and the audience and their interaction on the social media platform was given in a study (Bamman and Smith 2015). Researchers (Wang et al. 2015) explored intrinsic factors of tweets, including authors' past and use these in a sequential classification approach to detect sarcastic tweets. Another work proposed (Sulis et al. 2016) investigated the hashtags #sarcasm, #irony and #not and the similarities and differences between the different figures of language. A two-layered approach to classify French Tweets, by using a generic classifier followed by checking its pragmatic context using search engines to further refine it was also proposed (Karoui et al. 2015).

A number of machine learning techniques and more recently deep learning techniques have been applied in a variety of research domains (Wang et al. 2018, 2019; Tang et al. 2018; Zhang et al. 2018a, b, c, d). Deep learning techniques attempt to artificially simulate the hierarchical learning approach of the human brain and generalize globally. With the growing popularity of deep learning techniques in natural language processing task (Young et al. 2018), some of these have been exploited within the domain of automatic sarcasm detection too. Authors (Felbo et al. 2017), proposed a hybrid of attention based BiLSTM and CNN to detect emotions on Twitter. A variety of features such as emotions; sentiment and personality based were used by authors (Poria et al. 2016). Authors (Zhang et al. 2016) viewed tweets with reference to their contextual environment and use recurrent neural networks as a part of their model. In another reported work (Joshi et al. 2016) too used similarity between word embeddings as features for sarcasm detection. In the same year, a novel convolution network model was proposed where the model learnt user embeddings and utterance-based embeddings (Amir et al. 2016). A combination of convolution neural network, a recurrent neural network variant and a deep neural network was given (Ghosh and Veale 2016) and showed improved performance using deep learning architecture. The study conducted in this research examines machine-learning based and deep-learning based models with a varied set of features for building a robust and efficient automatic sarcasm detection system.

In this work, three novel models have been proposed for sarcasm detection and comparison between the performances of three models has been done to evaluate the best model. The models define and capture context in local and global manner. The first model is based on lexical feature using TF-IDF and ensemble voting of shallow classifiers for classification of sarcastic and non-sarcastic tweets. The second model not only uses lexical features of tweets but also use punctuation-based features as well as semantic features to increase the efficiency of sarcasm detection model. The second model uses shallow classifiers for classification of sarcastic and non-sarcastic tweets. The third model uses semantic features in the form of word embeddings using GloVe. These models are chosen so as to identify the effect of all the categories of features such as syntactic, semantic and pragmatic and also, to compare the performance of shallow classifiers and deep learning for sarcasm detection.

3 Automatic sarcasm detection

The generic model for predicting a sarcastic text automatically has primarily four components, namely (i) Data acquisition and pre-processing (ii) Feature Extraction (iii) Learning using Shallow/Deep Classifiers and (iv) Evaluation using performance measures. The task begins by acquiring the data, which in this study includes two datasets, namely the shared task of SemEval 2015—task 11 and Reddit Posts. The unstructured input data has to be converted into an array of representative features to facilitate training and testing of classifiers. Thus, pre-processing task is carried out, which structures the input data for feature extraction. The relevant features (statistical; sentiment-based; punctuation-based; word embedding features) are used to automatically label each tweet/post as sarcastic or non-sarcastic using different classification algorithms. Performance results obtained are then assessed on the basis of evaluation metrics.

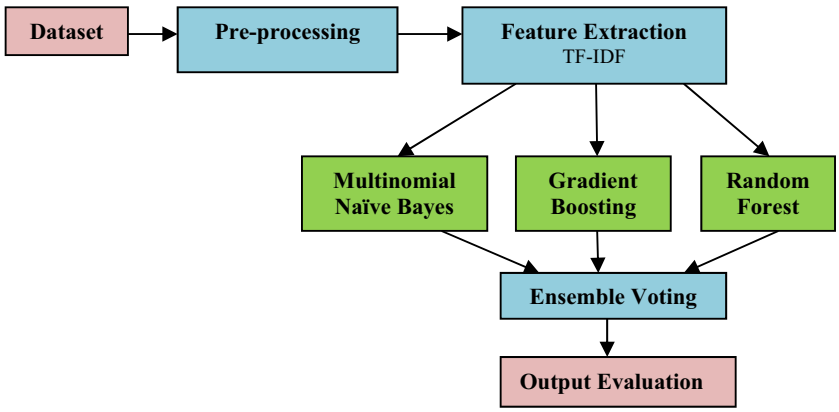


Fig. 3 Systematic flow of the first model

3.1 Models

In this experimental study we investigated three predictive models for automatic sarcasm detection. Each model uses a varied set of features which builds context vector locally and/or globally and is implemented choosing between the set of shallow or deep classifiers. The following sections briefly discuss the models whereas the details of pre-processing, features extraction and classification algorithms used in each model are expounded in the subsequent sub-sections.

3.1.1 First model: lexical feature + ensemble voting of shallow classifiers

The first model is based on the Bag-of-Words (BoW) model which uses TF-IDF (Bhatia and Kumar 2008) as a lexical level feature to extract the most descriptive terms in a text. Thus, to assess the significance of a word in the corpus, the TF-IDF features are generated using TF-IDF Vectorizer which is subsequently used as input parameters for training the classification algorithm.

In this first model, three classifiers are used, namely the Multinomial Naïve Bayes, Gradient Boosting and Random Forest. One of the simplest ways of combining the predictions from multiple machine learning algorithms is by using voting ensembles. A voting classifier is then used to combine model predictions into ensemble predictions. It averages the predictions of the sub-models. This model is the simplest but it does not capture the semantics, position of the terms or their co-occurrences in text. The Fig. 3 depicts the systematic flow of this model.

3.1.2 Second model: lexical-pragmatic-semantic (sentiment) features + baseline shallow classifiers

The second model combines a variety of features (lexical, pragmatic and semantic). Pragmatic features are linguistic markers that rather than relying on the information apropos the literal meaning of the text comprehend the way of utterance. It describes the "meaning in context" where, context means the situation in which an utterance is made; by whom, to whom, and the immediate circumstances. Prominent categories of pragmatic markers within text include ellipses, quotation marks, exclamation points, question marks, emoticons, caps, parentheses, hyphens and vocalization signals, amongst others. Whereas pragmatics in language is about markers from the point of view of usage, semantics markers are concerned with the study of meaning.

Sarcasm is invariably defined as a specialized sentiment across literature studies. We use the sentiment and punctuation features as defined in the reported work (Bouazizi and Ohtsuki 2016) on 'pattern-based approach to sarcasm detection' as semantic and pragmatic features, respectively. The details of these features are given in Sect. 3.4.2. These pragmatic and semantic features are then combined with the top 200 TF-IDF features (intrinsic filtering) to generate a feature matrix which is used as input to the baseline classifiers. The Fig. 4 depicts the systematic flow of this model.

3.1.3 Third model: semantic (word embedding) features + deep learning

Deep learning excels at finding useful representations of the data for a particular task. It has the capability to extrapolate novel features from a limited set of features in a training set, without a human intervention. That is, it searches for and seeks further features that correlate to the previously known ones without the need to label everything (Young et al. 2018). Motivated by the collective intelligence of deep learning, in the third model we demonstrate the use of LSTM and bidirectional LSTM models, the variants of Recurrent Neural Network (RNN) on benchmark datasets for improved sarcasm classification. Word embeddings are imperative to any deep learning based natural language processing task. In this model, the word embedding is done using GloVe (Pennington et al. 2014). The Fig. 5 depicts the systematic flow of this model. The dense layer is fully connected layer, so all the neurons in a layer are connected to those in a next layer. The dropout drops connections of neurons from the dense layer to prevent overfitting.

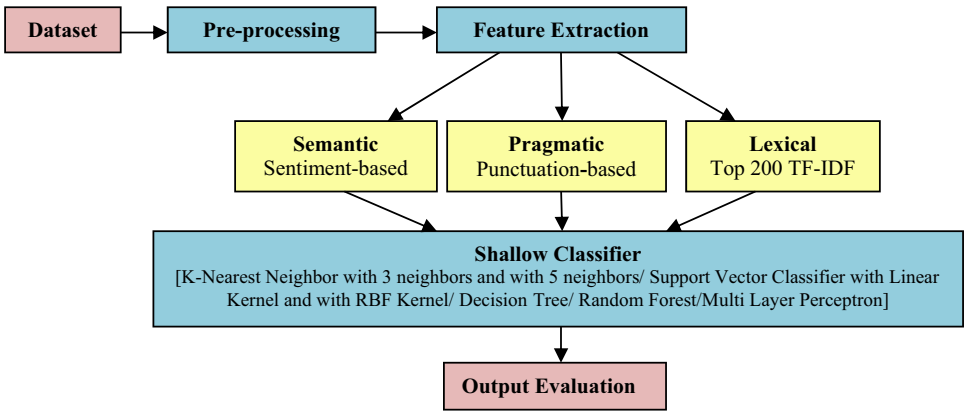


Fig. 4 Systematic flow of the second model

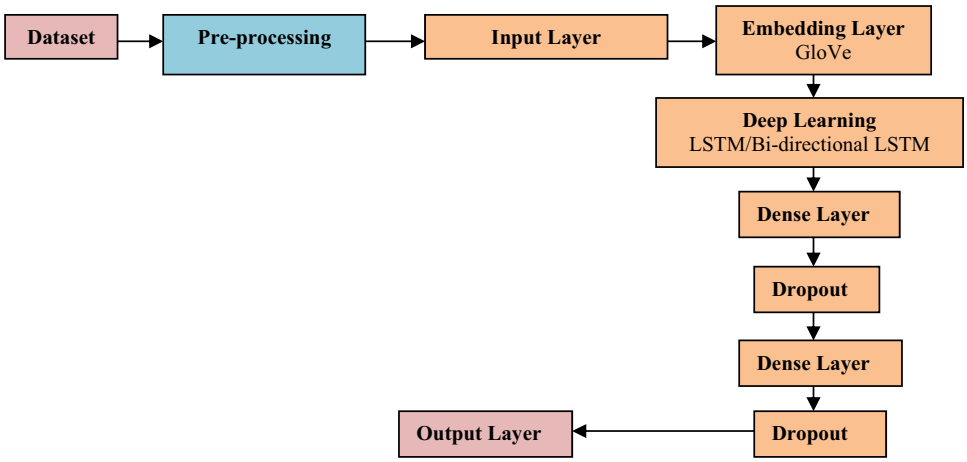


Fig. 5 Systematic flow of the third model

3.2 Datasets

To evaluate the performance of the three feature-variant predictive models, two datasets are used for empirical analysis. These datasets are, namely, the benchmark Twitter corpus: SemEval 2015 Task 11 and the Kaggle's Reddit dataset. The length of the text in the Reddit is greater than the 140 character (now 280) limit set by Twitter, and this was significant to understand how the models worked on longer texts. This helped in validating the results and in robustness check of the learning models. The Table 1 presents the various statistics of the two datasets.

The characteristic features of these datasets are briefly discussed next.

3.2.1 SemEval 2015 Task 11

SemEval 2015 task 11 (Ghosh et al. 2015) challenge was defined for "sentiment analysis of figurative language on Twitter". The sarcasm corpus is constructed using hashtags #irony, #sarcasm and #not (considered similar for the empirical analysis).For the negative class data, SemEval 2013 task 2 dataset was used (Nakov et al. 2013). The dataset is a balanced once as it contains almost equal number of samples from the positive and negative class. A total of 15,961 tweets were retrieved, categorized as 7994 sarcastic and 7324 non-sarcastic.

Table 1 Comparison of statistics pertaining to different datasets used

Dataset	Average length	Average no. of words	Average no. of punctuations	Average length of words
SemEval	104	18	2.4	4.2
Reddit	130	22	3	5

3.2.2 Reddit posts

Reddit is a social web portal where users post their thoughts and views on number of topics in pre-defined categories known as subreddits. The "/s" subreddit contains sarcastic content. Hence, posts taken from this subreddit have been labeled as sarcastic while others are negatively labeled. Reddit has no restrictions on the number of characters in a text post and we thus performed our experiments on it to investigate how well our model works on longer text. This is an important aspect since longer text poses the problem of context where sarcasm can only be inferred from a previous statement made in the text. Shorter texts are therefore less "tricky" to classify. This Reddit dataset for this study

was taken from the Kaggle competition series (Khodak et al. 2017) and it contains 19,737 total posts, categorized as 9999 sarcastic and 9398 non-sarcastic.

3.3 Pre-processing

Post data acquisition, to logically mine the text in tweets and Reddit posts, pre-processing is done to clean and transform the data for extraction of features. This pre-processing includes:

- Removal of duplicates tweets, stop words, number in tweets with placeholders, mentions, etc.
- Replacement of URLs, special characters such as @, #.
- Natural Language tool-Kit (NLTK)[1] for tokenization.
- Porter's stemmer for stemming to the root word.
- Removal of non-ASCII English character.
- Part-of-Speech tagging is also done to extract common structural patterns such as verb, adverb, adjective and noun.

For example, we consider a sample tweet as follows:

"@Raj_232 As a kid, I dreamt of being 30 with a Good CAR, BIG House, and Highly Paid Job. At least one dream came true: I'm 30. https://twitter.com/__yarelii/status/447952784277925888"

The Pre-processing takes place in the sample tweet as follows:

- Firstly, the given tweet is broken down into stream of tokens as shown below. These tokens are then sampled in lower case, to maintain uniformity.

['@raj_232', 'as', ' a', 'kid', ',', 'i', 'dreamt', 'of', 'being', '30', 'with', 'a', 'good', 'car', 'big', 'house', 'and', 'highly', 'paid', 'job', '.', 'at', 'least', 'one', 'dream', 'came', 'true', ':', 'I', 'am', '30', '.', 'https://twitter.com/__yarelii/status/447952784277925888']

- Next, Hashtags, Numbers in the given tweet, URLs, Mentions is substituted with the placeholders as depicted below.

['AT_USER', 'as', ' a', 'kid', ',', 'i', 'dreamt', 'of', 'being', 'NUMBER', 'with', 'a', 'good', 'car', 'big', 'house', 'and', 'highly', 'paid', 'job', '.', 'at', 'least', 'one', 'dream', 'came', 'true', ':', 'I', 'am', 'NUMBER', '.', 'URL']

- Last pre-processing step involves normalization of words by transforming them into singular, active and present tense. This allows easier parsing and accurate feature acquisition.

['AT_USER', 'as', ' a', 'kid', ',', 'i', 'dream', 'of', 'being', 'NUMBER', 'with', 'a', 'good', 'car', 'big', 'house', 'and', 'high', 'pay', 'job', '.', 'at', 'least', 'one', 'dream', 'come', 'true', ':', 'I', 'am', 'NUMBER', '.', 'URL']

- Although all the punctuation marks in a tweet are removed as the part of the cleaning process; however, the count of each punctuation mark is kept as we use them as pragmatic features to train the model. In this work, five punctuation-based features that represent figurative text and provide symbolic clues within the tweet are used. These include exclamation marks (!), question marks (?), periods (.), capital letters and use of "or".

3.4 Feature extraction

As discussed previously, three categories of features, namely, lexical, pragmatic and semantic features are used in this study. The first model makes use of a conventional lexical feature, which is the TF-IDF to train three classifiers that generate a voting based output. The second model uses punctuation-based features as pragmatic markers and sentiment-based features as semantic markers. These pragmatic and semantic features are combined with lexical features to train and test the baseline classifiers. In the final model, semantic relatedness is captured using the pre-trained GloVe word embeddings in deep recurrent neural network models, namely the LSTM and Bi-directional LSTM. A brief discussion on the features used is given next.

3.4.1 Lexical feature: TF-IDF

TF-IDF is a statistical weighting scheme, widely used in text mining tasks. It is a combination of two scores: term frequency and inverse document frequency. Term-frequency is a measure of the frequency of occurrence of a particular term in a single document, normalized by dividing with the total number of terms present in that document, i.e. the

[1] https://www.nltk.org/.

document length to enable comparisons between documents of different lengths. It is denoted as given in the following equation:

$$\mathrm{TF}_{t,d} = \mathrm{f}_{t,d} \Big/ \sum_{t \in d} t \tag{1}$$

where $\mathrm{TF}_{t,d}$ is denotes the term frequency of term t in document d, $\mathrm{f}_{t,d}$ is denotes the frequency of term t in document d and $\sum_{t \in d} t$ is denotes the total number of terms present in document d.

Document frequency, on the other hand, measures the number of documents containing a particular term t. Term-frequency gives equal importance to each term in a document, which can be misleading as certain terms like articles and prepositions occur very frequently in documents but are hardly of any relevance. This shortcoming is overcome by using the inverse document frequency which scales down the weights of frequent terms and gives more importance to not so frequent terms. It is calculated as given in the following equation:

$$\mathrm{IDF}_t = \log \frac{N}{\mathrm{dft}} \tag{2}$$

where IDF_t is refers to the inverse document frequency of term t, N is the total number of documents in the corpus, df_t is the document frequency of a term t across the corpus.

The TF-IDF value is then computed for each term as given in the following equation:

$$\mathrm{TF} - \mathrm{IDF} = \mathrm{TF}_{t,d} \times \mathrm{IDF}_t \tag{3}$$

Thus, the TF-IDF score is lowest for terms that occur frequently in almost all documents, highest for terms that occur frequently in a small number of documents and in between for terms that occur a few times in some documents or occur in many documents.

3.4.2 Pragmatic features: punctuation-related

Punctuations, wordplay or uppercase alphabets characterize symbolic clues which help comprehend the context within the text. Frequency encoding has thus been done for these pragmatic features as shown in the following Table 2 (Bouazizi and Ohtsuki 2016):

3.4.3 Semantic features

While pragmatic features are concerned with the language use, semantic features typically convey the conceptual relationship between words. In this study, two types of semantic features capture the relatedness within words namely, the sentiment and the word embeddings.

Table 2 Pragmatic features used

rec	Frequency of recurring alphabets, that is, if alphabet recurrence > 2, then set the feature to true else false
exclm	Frequency of exclamation marks
quest	Frequency of question marks
dot	Frequency of dots
upper	Frequency of capital letters
quotes	Frequency of ‘’ or “”

- Sentiment-based features

The sarcastic sentences are skewed in the sense that the sentiment polarity of the literal meaning may differ radically from the intended figurative meaning. The shift in sentiment (positive to negative or negative to positive) within these sentences is a strong indicator of sarcasm. For example, in the sentence “*Oh, how I love this feeling of being hurt and alone*” is a clear pointer to contradicting a fact. In this case, the fact (i.e., “*being hurt and alone*”) and the contradictory statement to that fact (i.e., “*I love*”) express sentiment polarity shifts. Therefore, any type of inconsistency or contradiction between sentiments within the tweet/post can be considered hinting sarcasm. The sentiment-based features defined in a previous reported work (Bouazizi and Ohtsuki 2016) have been used in this study. The authors characterize two lists of words classified as “positive words” and “negative words” which contain words with the positive emotional content, i.e. emotional positive terms (e.g., “love”, “enjoy”, “happy”, etc.) and that with negative emotional content i.e. emotional negative terms (e.g., “hate”, “worry”, “sad”, etc.). These lists of words are created using the pos_tag library under NLTK (Loper and Bird 2002). Next, using these lists, for each tweet or post the number of positive words (pw) and negative words (nw) are counted. Also, the adjectives, adverbs and verbs have higher emotional content as compared to nouns, therefore, all the words, either positive or negative that have the associated POS tag, are counted another time to create two additional features that are represent the number of highly emotional positive terms (PW) and highly emotional negative terms(NW) respectively. Sometimes, emotional content is conveyed through hashtags too. For example, in the tweet, “*Thanks a lot for always helping me #ihateyou*”, the hashtag “*#ihateyou*” tells that the user is not actually thanking the addressed user, but was rather extremely disliking him for not helping him. In addition to the already mentioned features, some features that are related to the contrast between the sentimental components are also extracted. Contrast means the coexistence of both a negative and a positive component in the same tweet. The ratio of emotional words is calculated denoted as ρ(t) expressed in the following equation:

Table 3 Sentiment-based features used

pw	Count of words with positive sentiment
nw	Count of words with negative sentiment
PW	Count of words with highly emotional positive content
NW	Count of words with highly emotional negative content
ρ score	Finds contrast between the above sentimental components, here delta is equal to 3

$$\rho(t) = \frac{(\delta \cdot PW + pw) - (\delta \cdot NW - nw)}{(\delta \cdot PW + pw) + (\delta \cdot NW - nw)} \quad (4)$$

where pw is the count of words with positive sentiment, nw is the count of words with negative sentiment, PW is the count of words with highly positive emotional content, NW is the count of words with highly negative emotional content, ρ is the score determined to find contrast between the above sentimental components, t is the tweet, δ is a weight bigger than 1 given to the highly emotional words and is set to 3.

Thus, the sentiment related features considered for this study are summarized in Table 3.

- Word embedding-GloVe

Word embeddings facilitate learned word representations. The benefits of extracting features based on word embedding to detect sarcasm have been recently reported (Joshi et al. 2016). In our study, for building word embeddings, GloVe (Pennington et al. 2014) is used. It is a count-based model to capture the semantic relatedness of words and builds the learned representation of a real-valued vector of words as input. It generates a 'word vector table' representing words by feature vectors. This log-bilinear model studies the relationship of words by counting the number of times they co-occur. Thus, this model aids in mapping all the tokenized words in each tweet/post to its respective 'word vector table'. Proper padding is done for unifying the feature vector matrix.

Thus, the feature set generated in this step is used to train and test the classifier for two pre-defined categories namely, sarcastic and non-sarcastic. A brief discussion on the classifiers used in this study is given next.

3.5 Shallow classifiers

Shallow classifiers refer to the type of task-specific algorithms where the feature extraction is problem specific and the model is trained using these handcrafted features. Shallow structure models generate generalized prediction using supervised learning through classification. Shallow classifiers employ statistical machine learning models like naive bayes, SVM, decision tree, random forest, etc., which perform satisfactorily on the numeric features, but fail to understand the context of the text. Deep learning architectures [such as recurrent neural networks (RNN) and convolutional neural network (CNN)] techniques of word embeddings can more efficiently and effectively apprehend the contextual semantics of the information and enhance the generic classification task. The Table 4 concisely describes the shallow classifiers used in this study.

3.6 Deep learning model

"Deep Learning" also known as "Deep structured Learning" or "Hierarchical Learning" involves learning of deep representation of data be it structured or unstructured and allow to build a solution optimized from algorithm to solve Machine Learning problems. Deep Learning has become growing trend to abstract better results when data is large and complex. It consists of a combination of Machine Learning and Artificial Neural Network (ANN) which refers to the depth of the network as it has more than one hidden layer(s).

Deep learning models in literature are defined as neural networks with a large number of parameters and layers in one of four fundamental network architectures:

- Unsupervised pre-trained networks
- Convolutional neural networks
- Recurrent neural networks
- Recursive neural networks

A *Convolutional Neural Network* is basically a standard neural network that has been extended across space using shared weights. A *Recurrent Neural Network* is basically a standard neural network that has been extended across time by having edges which feed into the next time step instead of into the next layer in the same time step. It has cycles inside that imply the presence of short memory in the net. A *Recursive Neural Network* is more like a hierarchical network where there is really no time aspect to the input sequence but the input has to be processed hierarchically in a tree fashion. The following Fig. 6 represents the hierarchy of deep learning and its variants.

As described previously, in this research, LSTM and Bi-LSTM have been used to build the deep learning model for sarcasm detection.

3.6.1 Long–short-term model (LSTM)

The long–short-term memory (LSTM) was introduced as a modified version of the classic RNN with the ability to adjust states to determine what to store and what to forget. The RNNs are enhanced with the intent to improve its memory capabilities forming the LSTM (Salehinejad et al. 2017). The primary

Table 4 Shallow classifiers used

Technique	Description
Naïve Bayes (NB)	It is based on the Bayes Theorem. Three versions of Naive Bayes models exists, namely, the Multinomial, Binarized and Bernoulli. In this study, the Multinomial Naïve Bayes explicitly models the word counts and works for discrete data
Support vector classifier (SVC)	The SVM Classifier (SVC) is described by a decision plane (hyper-plane) which builds decision boundaries for separating group of instances as different class members. The function of kernel is to take data as input and simulate the projection of the initial data in a feature space with higher dimension
SVC (with Linear Kernel)	A linear kernel uses linear functions $k(x_i, x_j) = (x_i \cdot x_j + 1)^1$
SVC (with RBF Kernel)	It is a general-purpose kernel and is used when there is no prior knowledge about the data. It uses the following equation: $k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$
K-Nearest neighbor (KNN)	K- nearest neighbors is a non-parametric classification algorithm that finds the cluster closest in distance to a particular data point
Multi layer perceptron (MLP)	MLP is a type of neural networks which consists of input, hidden and output nodes. It is a self-adaptive and data driven technique
Decision tree (DT)	It is a machine learning algorithm that forms a tree based classification where the internal nodes represent features and the leaf nodes represent final classes
Random forest (RF)	The Random Forest is an ensemble classifier based on Decision Trees. The random forest makes a prediction about the class, not simply based on one decision trees, but by an (almost) unanimous prediction, made by 'K' decision trees
Ensemble vote classifier	The Ensemble Vote Classifier is used to combine the results of different machine learning classifiers for classification. It is a meta-classifier and predicts the final class label via majority voting which is the class label that has been predicted most frequently by the classification models

purpose was to deal with the vanishing gradient and exploding gradient problems in the standard RNN. During training, errors back-propagate to adjust weights while maintaining the training properties. Figure 7 illustrate the structure of LSTM.

The calculation for the forget, the input and the output gate along with the input cell state is done using the following Eqs. 5–10:

$$i_t = \sigma\left(w_{i_x} x_t + w_{i_h} h_{t-1} + b_i\right) \tag{5}$$

$$f_t = \sigma\left(w_{f_x} x_t + w_{f_h} h_{t-1} + b_f\right) \tag{6}$$

$$o_t = \sigma\left(w_{o_x} x_t + w_{o_h} h_{t-1} + b_o\right) \tag{7}$$

$$s_t = \tan h\left(w_{s_x} x_t + w_{s_h} h_{t-1} + b_s\right) \tag{8}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot s_t \tag{9}$$

$$h_t = \tan h\left(c_t\right) \odot o_t \tag{10}$$

where $\odot$ represents element wise product. $w_i w_f w_o w_s$ are weighing factors used for mapping the hidden layer input to the three gates and the input cell state. $b_i b_f b_o b_s$ are bias vectors. The σ is the gate activation function(sigmoid function). tanh is the hyperbolic tangent function. $Y_t = \left[h_{t-n}, \ldots, h_{t-1}\right]$ represents the final output of a LSTM layer that is the vector of all outputs.

3.6.2 Bi-directional LSTM (Bi-LSTM)

This consists of two networks where one reads information in forward direction and another reads it in the reverse direction. The output thus generated includes both past and future context. Each element of output vector generated by Bi-LSTM layer is calculated by using the following Eq. 11 (Graves et al. 2013):

$$y_t = \sigma(h^{\rightarrow} t, h^{\leftrightarrow} t) \tag{11}$$

where σ. function is used to combine the two output sequences. It can be a concatenating function, a summation function, an average function or a multiplication function. The final output of a Bi-LSTM layer can be represented by a vector as given in the following equation:

$$Y_t = \left[y_{t-n}, \ldots, y_{t-1}\right] \tag{12}$$

Wrapping the LSTM hidden layer with a Bi-directional layer creates a Bi-LSTM. By default, the output values from these LSTMs will be concatenated.

4 Results

This section presents and discusses the performance results achieved by the classifier using the three models on SemEval 2015 and Reddit datasets based on key performance indicators. Accuracy, Precision, Recall and F1 score have been

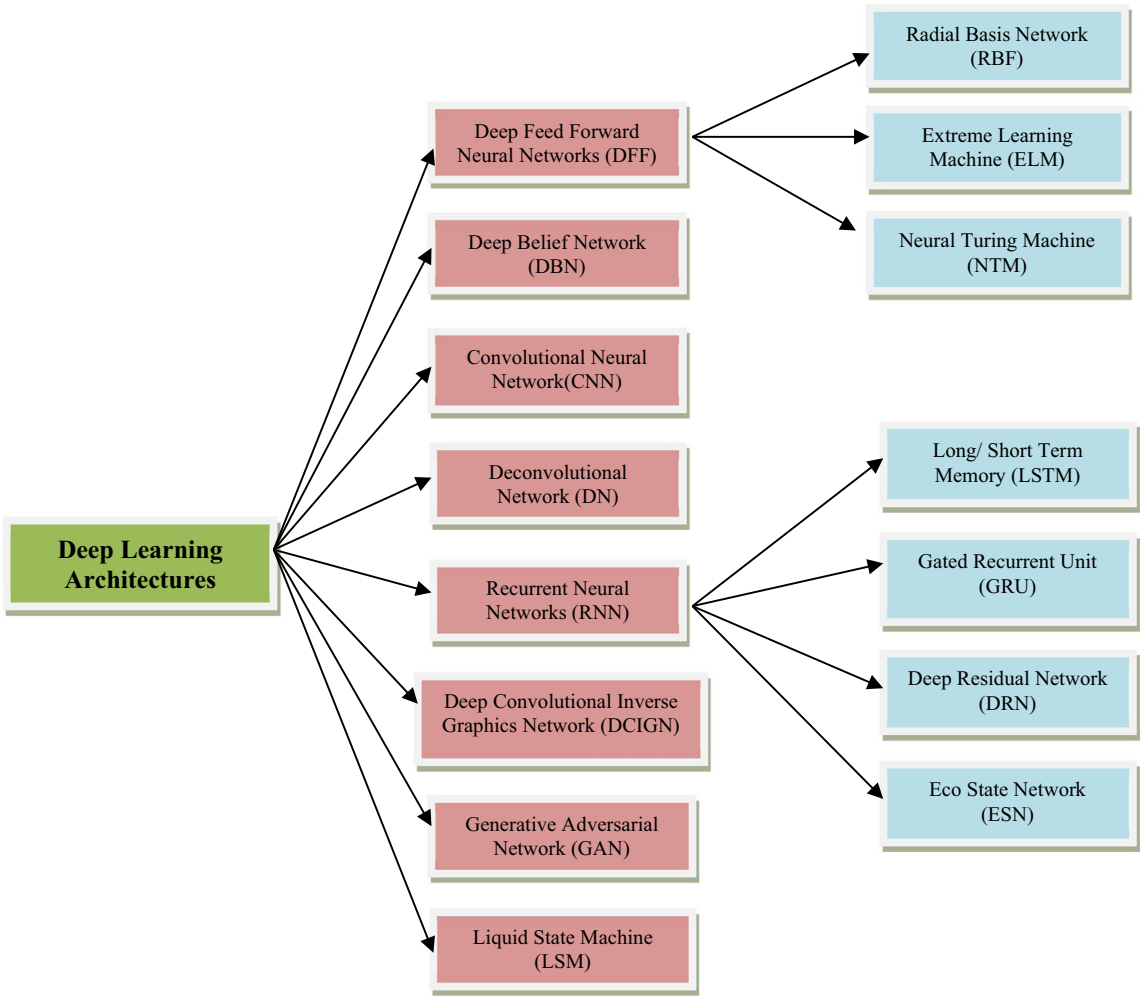


Fig. 6 Deep learning architectures

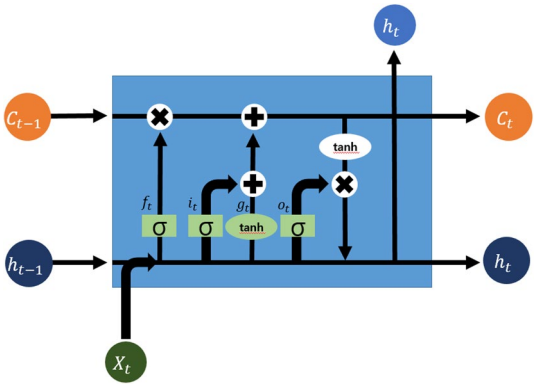


Fig. 7 Structure of LSTM

used as the performance measure for each model. "Accuracy is defined as the proximity of a measurement to its true value, i.e., it is measured as a proportion of true positives and true negatives among total inspected cases". "Precision defines the exactness of any classifier, i.e., it is measured as ratio of true positives to all the predicted positives". "Recall is defined as the ratio of true positives to all the actual positives". "F1 score is the harmonic mean of Precision and Recall" (Powers 2011). All the values are expressed in percentages (%).

In order to attain the best performance results optimal selection of parameters is crucial. The Table 5 gives the values of parameters used in this research.

4.1 Results using the first model

The Table 6 illustrates the results of the implementing the first model. The results show an accuracy of 80.23% when the model is applied on the SemEval Dataset while an accuracy of 75.31% is achieved when applied on the Reddit dataset. The Fig. 8 depicts the results graphically.

Table 5 Hyper-parameter values

Hyper-parameter	Value
Dimension of GloVe vectors	200
Hidden units of LSTMs (Forward, Backward)	500 each
Mini Batch Size	10
Regularization	Dropout Operation
Drop-out Rate	0.5: word embedding; 0.2: Bi-LSTM
Learning rate	0.2

Table 6 Performance Results using First Model

	SemEval dataset	Reddit dataset
Accuracy	80.23	75.31
Recall	80.74	74.22
Precision	76.39	71.30
F1 Score	78.51	72.73

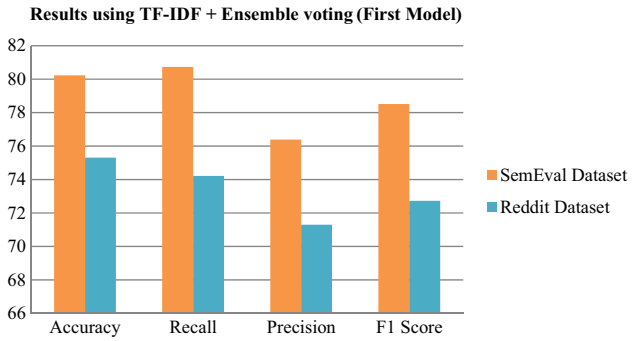


Fig. 8 Results using the first model

4.2 Results using the second model

Our second model extracts 11 features (sentiment +punctuation) which are concatenated with the top TF-IDF features generated. This feature matrix is used to train the five baseline classifier (SVM, KNN, DT, RF and MLP). We considered KNN with 3 and 5 neighbours and the support vector classifier with both Linear and RBF Kernel. The Tables 7, 8, 9, 10, 11, 12, 13 depict the results.

The Fig. 9 depicts the results graphically.

It can be observed from the above results that the MLP gives the highest accuracy for the datasets, that is, 83.56% for the SemEval dataset and 77.95% for the Reddit dataset. Next to MLP is the RF classifier. The least accuracy of 73.27% is observed using DT for the SemEval dataset and 73.25% for Reditt dataset using SVC with RBF kernel.

Table 7 Performance Results using Second Model (KNN with 3 neighbors)

	SemEval Dataset	Reddit Dataset
Accuracy	75.61	74.12
Recall	75.83	72.57
Precision	67.11	63.24
F1 Score	71.20	67.58

Table 8 Performance Results using Second Model (KNN with 5 neighbors)

	SemEval dataset	Reddit dataset
Accuracy	76.88	77.13
Recall	78.73	79.27
Precision	67.35	69.39
F1 Score	72.60	74.01

Table 9 Performance Results using Second Model (SVC with Linear Kernel)

	SemEval dataset	Reddit dataset
Accuracy	80.72	73.25
Recall	89.56	90.13
Precision	67.43	71.26
F1 Score	76.94	79.59

Table 10 Performance Results using Second Model (SVC with RBF Kernel)

	SemEval dataset	Reddit dataset
Accuracy	81.03	75.34
Recall	85.81	77.89
Precision	68.73	64.32
F1 Score	76.32	70.45

Table 11 Performance Results using Second Model (Decision Tree)

	SemEval dataset	Reddit dataset
Accuracy	73.27	76.37
Recall	67.18	71.38
Precision	67.74	64.59
F1 Score	67.46	67.81

4.3 Results using the third model

The accuracy results of this model are represented in Tables 14 and 15.

Table 12 Performance results using second model (random forest)

	SemEval dataset	Reddit dataset
Accuracy	82.69	77.35
Recall	89.13	87.44
Precision	69.08	72.20
F1 Score	77.84	79.09

Table 13 Performance Results using Second Model (Multi Layer Perceptron)

	SemEval dataset	Reddit dataset
Accuracy	83.56	77.95
Recall	80.26	75.39
Precision	72.98	71.83
F1 Score	76.45	73.56

Table 14 Performance Results using Third Model (LSTM)

	SemEval dataset	Random Tweets
Accuracy	84.89	80.28
Recall	83.51	78.01
Precision	86.78	74.98
F1 Score	80.39	76.91

Table 15 Performance Results using Third Model (Bi-LSTM)

	SemEval dataset	**Random Tweets**
Accuracy	86.32	82.91
Recall	84.39	80.27
Precision	81.61	78.76
F1 Score	85.25	79.45

It can be observed that the Bidirectional LSTM outperforms the LSTM when implemented on both the SemEval and the Reddit datasets. In case of SemEval dataset it gives accuracy of 86.32% while in the case of Reddit dataset it gives the accuracy of 82.91%. The graphical representation of results is given in the Fig. 10.

To better understand the deep learning architectures for sarcasm detection, the results of LSTM and Bi-LSTM were compared with CNN too. The classification accuracy was observed as: CNN < LSTM < Bi-LSTM as shown in the following Fig. 11.

4.4 Comparison of the three models

The Bi-directional LSTM achieves the highest classification accuracy of 86.32% for the Twitter dataset and 82.91% for the Reddit dataset to detect sarcasm. The second-best results are obtained using LSTM with the accuracy of 84.89% for Twitter Dataset and 80.28% for Reddit dataset. The Fig. 12 depicts the comparison graphically.

5 Conclusion

Automatic sarcasm detection as a typical text classification task has the primary goal to distinguish between sarcastic or non-sarcastic textual content. Sentiment Analysis is certainly one of the areas that can benefit most from sarcasm detection. This study probed the problem of detecting sarcasm in social media from the computational linguistic perspective by relying on lexical, pragmatic and semantic cues. Context was defined using various features such as n-grams with TF-IDF encoding, frequency encoding for

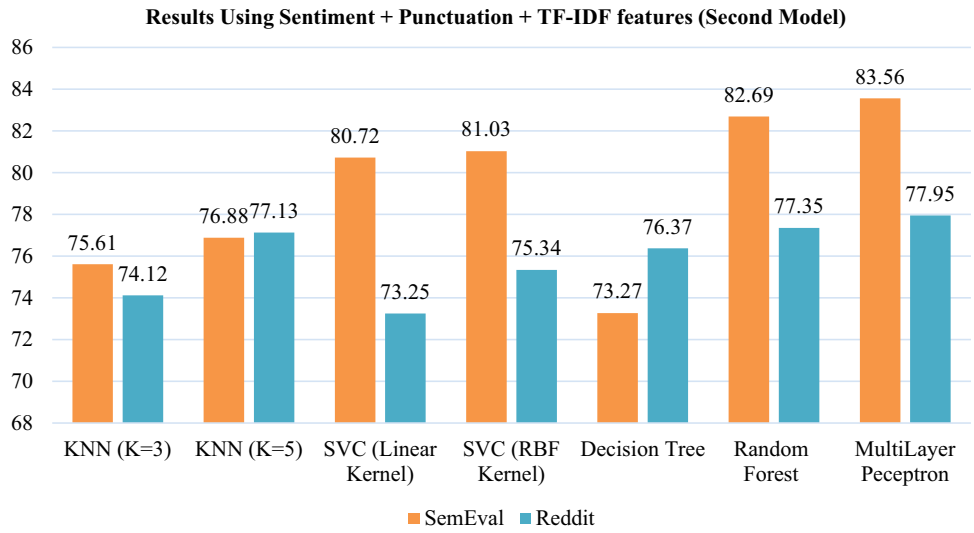


Fig. 9 Results using the second model

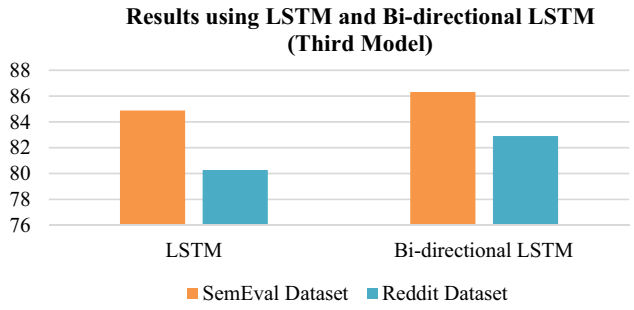


Fig. 10 Results using the third model

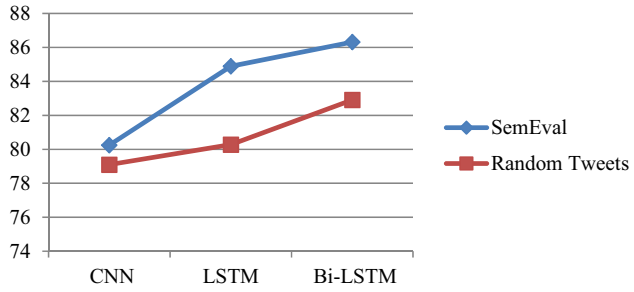


Fig. 11 Accuracy of deep models

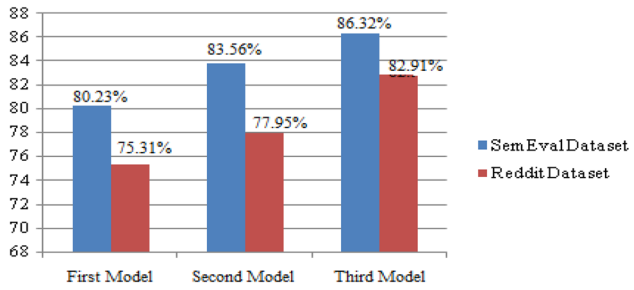


Fig. 12 Comparison of the three models

punctuations, hand-crafted features and pre-trained word embeddings. Three models were proposed using these linguistic context markers to train and test shallow as well as deep classifiers which predict sarcasm effectively and efficiently in two experimental datasets. It was observed that the Bi-directional LSTM using pre-trained vectors from the GloVe word embedding is the best sarcasm prediction model with an average accuracy of approximately 85% (~84.6). The preliminary results are clearly motivating. Deep learning architectures have been used for various natural language processing tasks and can be applied to random, live twitter streams for enhanced classification performance. More specifically, the recent attention and augmented recurrent neural networks architectures can be built to model a capacity to dynamically shift/select attributes such that, given a context (input and task pair), the overall decision making (classification) is more reliable. Also, as the accuracy of sarcasm classification mainly depends on the context vector, the multi-faceted concept of 'context' needs proper scrutiny. Further, the growing use of typo-graphic images, that is text represented as an image further characterizes the power of expressiveness in online social data and detecting sarcastic tone in it is a promising direction of research. Also, the use of code-mix and code-switch languages, intentional ambiguity, novelty in vocabulary and the "self-tagging" or "crowd-sourced" datasets create new challenges in sarcastic tone detection in short text, making it a dynamic area of research with lots of open problems.

Compliance with ethical standards

Conflict of interest The authors certify that there is no conflict of interest in the subject matter discussed in this manuscript.

References

Amir S, Wallace BC, Lyu H, Silva PC (2016) Modelling context with user embeddings for sarcasm detection in social media. arXiv preprint arXiv:1607.00976

Bamman D, Smith NA (2015) Contextualized sarcasm detection on Twitter. ICWSM. 2:15

Barbieri F, Saggion H, Ronzano F (2014) Modelling sarcasm in twitter, a novel approach. In: Proceedings of the 5th workshop on computational approaches to subjectivity, sentiment and social media analysis, pp 50–58

Barman U, Das A, Wagner J, Foster J (2014) Code mixing: a challenge for language identification in the language of social media. EMNLP 2014:13

Bharti SK, Babu KS, Jena SK. (2015). Parsing-based sarcasm sentiment recognition in twitter data. In: Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining, ACM, pp 1373–1380

Bharti SK, Vachha B, Pradhan RK, Babu KS, Jena SK (2016) Sarcastic sentiment detection in tweets streamed in real time: a big data approach. Digit Commun Netw 2(3):108–121

Bhatia MPS, Kumar A (2008) Information retrieval and machine learning: supporting technologies for web mining research and practice. Webology 5(2):5

Bouazizi M, Ohtsuki TO (2016) A pattern-based approach for sarcasm detection on Twitter. IEEE Access 4:5477–5488

Camp E (2012) Sarcasm, pretense, and the semantics/pragmatics distinction. Noûs 46(4):587–634

Chen J, Yan S, Wong KC (2018) Verbal aggression detection on Twitter comments: convolutional neural network for short-text sentiment analysis. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3442-0

Das A, Gambäck B (2014). Identifying languages at the word level in code-mixed Indian social media text. In: Proceedings of the 11th international conference on natural language processing, Goa, India, 2014, pp 169–178

Dave K, Lawrence S, Pennock DM (2003) Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th international conference on World Wide Web, ACM, pp 519–528

Davidov D, Tsur O, Rappoport A (2010).Semi-supervised recognition of sarcastic sentences in twitter and Amazon. In: Proceedings of the fourteenth conference on computational natural language learning, association for computational linguistics, pp 107–116

Felbo B, Mislove A, Søgaard A, Rahwan I, Lehmann S (2017) Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. arXiv preprint arXiv:1708.00524

Ghosh A, Veale T (2016) Fracking sarcasm using neural network. In: Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis, pp 161–169

Ghosh A, Li G, Veale T, Rosso P, Shutova E, Barnden J, Reyes A (2015) Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. In: Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015), pp 470–478

González-Ibánez R, Muresan S, Wacholder N. (2011).Identifying sarcasm in Twitter: a closer look. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies: short papers association for computational linguistics vol 2, pp 581–586

Graves A, Jaitly N, Mohamed AR. (2013). Hybrid speech recognition with deep bidirectional LSTM. In: 2013 IEEE workshop 2013 on automatic speech recognition and understanding (ASRU), pp 273–278

Satapathy R., Guerreiro, C., Chaturvedi, I., Cambria, E (2017) Phonetic-based microtext normalization for twitter sentiment analysis. In: ICDM, pp 407–413

Hercig T, Lenc L (2017) The impact of figurative language on sentiment analysis. In: Proceedings of the international conference recent advances in natural language processing, pp 301–308

Joshi A, Tripathi V, Patel K, Bhattacharyya P, Carman M (2016) Are word embedding-based features useful for sarcasm detection? arXiv preprint arXiv:1610.00883

Joshi A, Bhattacharyya P, Carman M (2017) Automatic sarcasm detection: a survey. ACM Comput Surv 50(5):73

Karoui J, Benamara F, Moriceau V, Aussenac-Gilles N, Hadrich Belguith L (2015) Towards a contextual pragmatic model to detect irony in tweets. Association for Computational Linguistics (ACL)

Khodak M, Saunshi N, Vodrahalli K (2017) A large self-annotated corpus for sarcasm. arXivpreprint arXiv:1704.05579

Kumar A, Garg G (2019) Systematic literature review on context-based sentiment analysis in social multimedia. Multimed Tools Appl. https://doi.org/10.1007/s11042-019-7346-5 **(in press)**

Kumar A, Sebastian TM (2012) Sentiment Analysis: a perspective on its past, present and future. Int J Intell Syst Appl 4(10):1–14

Liebrecht CC, Kunneman FA, van den Bosch APJ (2013) The perfect solution for detecting sarcasm in tweets #not. In: Proceedings of the workshop on computational approaches to subjectivity, sentiment, and social media by the North American Chapter of the Association for Computational Linguistics (WASSA-NAACL'13)

Liu B (2015) Sentiment analysis mining opinions, sentiments, and emotions. Cambridge University Press, Cambridge

Liu P, Chen W, Ou G, Wang T, Yang D, Lei K. (2014) Sarcasm detection in social media based on imbalanced classification. In: International conference on web-age information management. Springer, Cham 2014, pp 459–471

Loper E, Bird S. (2002). NLTK: The natural language toolkit. In: Proceedings of the ACL-02 workshop on effective tools and methodologies for teaching natural language processing and computational linguistics, Association for Computational Linguistics, vol 1, pp 63–70

Nakov P, Rosenthal S, Kozareva Z, Stoyanov V, Ritter A, Wilson T (2013) Semeval-2013 task 2: Sentiment analysis in twitter. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), Vol 2. Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, Georgia, USA, June. Association for Computational Linguistics, pp 312–320

Nilep C (2006) Code switching in socio-cultural linguistics. Colorado Res Linguist 19(1):1–22

Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. Int Lang Resour Eval 2010:1320–1326

Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr 2(1–2):1–135

Pennington J, Socher R, Manning C (2014).Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp 1532–1543

Poria S, Cambria E, Hazarika D, Vij P (2016). A deeper look into sarcastic tweets using deep convolutional neural networks. In: Proceedings of international conference on computational linguistics, pp 1601–1612

Powers DMW (2011) Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. J Mach Learn Technol 2(1):37–63

Ptáček T, Habernal I, Hong J (2014) Sarcasm detection on Czech and English twitter. In: Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers, pp 213–223

Rajadesingan A, Zafarani R, Liu H (2015) Sarcasm detection on twitter: a behavioral modeling approach. In: Proceedings of the eighth ACM international conference on web search and data, ACM, pp 97–106

Riloff E, Qadir A, Surve P, Silva LD, Gilbert N, Huang R (2013) Sarcasm as contrast between a positive sentiment and negative situation. EMNLP 13:704–714

Salehinejad H, BaarbeJ, Sankar S, Barfett J, Colak E, Valaee S (2017) Recent advances in recurrent neural networks. arXiv preprint arXiv:1801.01078

Sulis E, Farías DI, Rosso P, Patti V, Ruffo G (2016) Figurative messages and affect in twitter: differences between# irony,# sarcasm and# not. Knowl-Based Syst 15(108):132–143

Tang C (2018) Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. Front Neurosci 12:818

Wang Z, Wu Z, Wang R, Ren Y. (2015). Twitter sarcasm detection exploiting a context-based model. In: International conference on web information systems engineering. Springer, Cham 2015, pp 77–91

Wang S-H, Junding S, Preetha P, Guihu Z, Yudong Z (2018) Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units. J Real-Time Image Proc 15(3):631–642

Wang S-H, Muhammad K, Hong J, Sangaiah KA, Zhang Y-D (2019) Alcoholism identification via convolutional neural network based on parametric ReLU, dropout, and batch normalization. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3924-0

Young T, Hazarika D, Poria S, Cambria E (2018) Recent trends in deep learning based natural language processing. IEEE Comput Intell Mag 13(3):55–75

Zhang M, Zhang Y, Fu G (2016) Tweet sarcasm detection using deep neural network. In: Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers, pp 2449–2460

Zhang YD, Chichun P, Junding S, Chaosheng T (2018a) Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU. J Comput Sci 28:1–10

Zhang YD, Yin Z, Xiao-Xia H, Hong C, Shuihua W (2018b) Seven-layer deep neural network based on sparse autoencoder for voxelwise detection of cerebral microbleed. Multimed Tools Appl 77(9):10521–10538

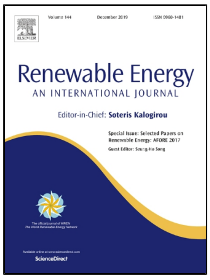
Zhang YD, Chichun P, Xianqing C, Fubin W (2018c) Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling. J Comput Sci 27:57–68

Zhang YD, Muhammad K, Chaosheng T (2018d) Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform. Multimed Tools Appl 77(17):22821–22839

Zubiaga A, Aker A, Bontcheva K, Liakata M, Procter R (2017) Detection and resolution of rumours in social media: a survey. arXiv preprint arXiv:1704.00656

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

Journal Pre-proof



EXERGO-ENVIRONMENTAL ANALYSIS OF AN INDIRECT FORCED CONVECTION SOLAR DRYER FOR DRYING BITTER GOURD SLICES

S. Vijayan, T.V. Arjunan, Anil Kumar

PII:	S0960-1481(19)31253-4
DOI:	https://doi.org/10.1016/j.renene.2019.08.066
Reference:	RENE 12145
To appear in:	*Renewable Energy*
Received Date:	29 April 2019
Accepted Date:	12 August 2019

Please cite this article as: S. Vijayan, T.V. Arjunan, Anil Kumar, EXERGO-ENVIRONMENTAL ANALYSIS OF AN INDIRECT FORCED CONVECTION SOLAR DRYER FOR DRYING BITTER GOURD SLICES, *Renewable Energy* (2019), https://doi.org/10.1016/j.renene.2019.08.066

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier.

Journal Pre-proof

EXERGO-ENVIRONMENTAL ANALYSIS OF AN INDIRECT FORCED CONVECTION SOLAR DRYER FOR DRYING BITTER GOURD SLICES

S.Vijayan[a]*, T.V.Arjunan[a], Anil Kumar[b*]

[a]Department of Mechanical Engineering, Coimbatore Institute of Engineering and Technology, Coimbatore 641 109, India
[b]Department of Mechanical Engineering, Delhi Technological University, Delhi-110042, India

*Corresponding Author: vijayandakshinya@gmail.com (S.Vijayan), anilkumar76@dtu.ac.in (Anil Kumar)
Contact number: +91-9865266082, +91 422-2970703
FAX number: +91 422-2970704

Abstract

In this study, a low cost indirect forced convection solar dryer integrated with porous bed sensible heat storage medium has been developed and studied under the climatic conditions of Coimbatore for drying bitter gourd slices. The experimental setup consists of a solar collector with $2m^2$ area, drying chamber and a centrifugal blower to force the air. The main objective of the study is to investigate the effect of mass flow rate of air on the exegetic and pickup efficiencies of the solar drying system for drying bitter gourd slices. The results indicate that the mass of the bitter gourd slices was reduced from 4000 to 723 grams in 7 hours at the mass flow rate of air 0.0636 kg/s in solar drying system. The average exergy efficiency values vary from 28.74 % to 40.67 % for the mass flow rates of air from 0.0141 to 0.0872 kg/s respectively. The average pickup efficiency of the drying air varied from 54.29% to 17.18% for various mass flow rates of air and it decreases with increase in mass flow rate of air. The effective moisture diffusivity of bitter gourd slices was observed from 8.6293×10^{-10} to 12.9585×10^{-10} m^2/s, whereas for open sun drying it gives the lowest value of 0.9568×10^{-10} m^2/s. Environmental impact analysis indicated that the energy payback time for the indirect solar dryer is found to be 2.21 years only. The CO_2 mitigation and earned carbon credit values for the developed system are 33.52 tons and INR 10894 to 43576 for the expected lifetime of 35 years. Impact analysis shows that the solar dryer is a suitable solution for preserving the agricultural products with environmental sustainability.

Keywords: Indirect solar dryer; Exergy efficiency; Bitter gourd; Effective moisture diffusivity; Embodied energy

Journal Pre-proof

1. INTRODUCTION

Preservation of agricultural products is considered as an essential process to ensure the availability of the products for later use and also to reduce the post harvest losses. One of the best known preservation techniques to mankind is of drying and it is practiced since ancient days. Solar drying method seems as advantageous one as it helps to overcome the major issues faced by the developing countries such as rapidly increasing price of fossil fuels and threatening environmental pollution. Solar dryers are utilizing the sun's energy to remove the excessive moisture from the wet products and they offer a wide range of advantages such as (i) superior quality of the dried products, (ii) speedy process, (iii) reduced losses (iv) more control over the process and (v) low floor space requirements, when compared with the open sun drying method (Prakash & Kumar, 2013; Vijayan et al., 2016)

Solar dryers can be classified based on the mode of receiving energy from sun as direct, indirect and mixed mode. In direct solar dryer, the products are directly exposed to sun's radiation which results in increase in temperature of the product and thereby the moisture is removed. In case of indirect solar drying, the products are kept in closed container and only hot air produced from collector is supplied to the products to remove the moisture. At last, the mixed mode solar dryer is combination of both direct and indirect solar drying concepts. Among these dryers, the indirect mode solar dryers are widely preferred for their advantageous features like superior quality of the dried products and having better control over the drying process when compared with the open sun and direct mode dryers (Vijayan et al., 2017; Singh et al., 2018).

Several researchers have developed a wide variety of solar dryers including direct, indirect and mixed mode dryers for drying the agricultural and food products with performance enhancement methods such as energy storage (sensible/latent heat), recirculation, backup heating, etc (Prakash et al., 2016). Usage of energy storage materials in the solar dryers improves the reliability and also the performance of the system as it addresses the intermittent nature of solar energy. Energy storage process can reduce the rate mismatch between energy supply and energy demand. A solar dryer without heat storage provides air with large variations in temperature to the dryer, and drying of food is not possible during partial clouds and late evening hours. In such cases a solar dryer with thermal storage is required. Reliability of a solar dryer can be increased by storing excess energy during peak time and using it in off-sun hours or when the intensity of solar energy is not inadequate (Vijayan et al. 2017b).

Thermodynamic analysis of the drying process is preferred for assessing the performance of the solar dryers in terms of energy and exergy parameters. Energy analysis is a quantitative way of investigating a system and exergy is qualitative analysis (Aviara et al, 2014). Energy analysis is the first law of thermodynamics, which involved with the basic energy balance to and from the system. However, energy analysis has some inabilities in presenting the real characteristics of the system e.g energy analysis does not object if heat is considered to be transferred spontaneously in the direction of the increasing temperature. It also does not distinguish the quality of the energy, e.g., 1W of heat equals 1W of work or electricity (Akpinar & Koçyiğit, 2010). Exergy is the second law of thermodynamics, which is defined as the maximum amount of work, which can be produced by a stream of matter, heat or work as it comes to equilibrium with a reference environment and is an important tool to analyse, optimise and thus, to improve the energy efficiency of a process (Dincer & Sahin, 2004). Exergy method considers both the system and also its environment properties during the analysis, because the performance of a system depends on both the characteristics. Exergy method accounts both the conservation of mass and energy concepts along with the second law of thermodynamics for the analysis, design and improvement of energy and other systems. Dincer (2000, 2002) stated that the exergy is a suitable technique for designing a system with more efficient by reducing the energy losses and a primary tool in best addressing the impact of energy resource utilization on environment. Sami et al. (2011) pointed that the exergy is a significant property to estimate the maximum useful work that can be converted from the system, when it is brought to thermodynamic equilibrium from its reference state.

Exergetic performance of various types of solar dryers have been investigated and reported for wide range of agricultural products. Midilli and Kucuk (2003) carried out a detailed study on energy and exergy performance of solar drying process of shelled and unshelled pistachios using an indirect solar cabinet dryer. Exergetic inflow increases almost linearly with increase in inlet temperature of air and found that the exergetic efficiency of the cabinet was in the range of 15.65 to 100%. In the similar way, the energy and exergy performance of forced convection solar dryer was investigated for drying of parsley leaves by Akpinar (2011). Exergy efficiency of the drying process was observed between 35.76 to 86.79 % and found that Energy Utilization Ratio (EUR) decreased with increase in drying time while exergy efficiency increased. Akpinar (2010) developed an indirect forced convection solar dryer and investigated its energy and exergy performance for drying mint leaves. Exergetic efficiency of the drying chamber was varied from 34.76% to 87.72% and

increases with increase in drying time. Exergetic performance of a solar tunnel dryer was investigated for drying jackfruit leather. The average dryer exergy efficiency was 41.42% and highlighted that it decreases with increase in drying time (Chowdhury et al., 2011). Fudholi et al. (2014) have developed an indirect solar dryer with a double pass collector, auxiliary heater and a drying chamber. Drying characteristics of red seaweed as well as the performance of the system were investigated. The collector, overall dryer pickup and exergy efficiencies were 35, 27, 95 and 30 %, respectively. Rabha et al. (2017) developed a forced convection solar tunnel dryer integrated with latent heat storage. The energy and exergy performance of the system for drying ghost chilli pepper and ginger was studied. Exergetic efficiency was 63% and 47% for ghost chilli and sliced ginger respectively. It was also concluded that the exergy efficiency may vary from product to product in the same drying system. A low cost and user friendly mixed mode solar tunnel dryer was developed for drying agricultural products and exergy analysis was investigated for drying turmeric along with the thin layer drying characteristics. Exergetic efficiency of the system was varied from 23.25 to 73.31% and with the average of 49.12% (Karthikeyan and Murugavelh, 2018). A solar dryer integrated with packed bed energy storage medium was developed and performance of the system was studied using orange slices (Atalay, 2019). The exergy efficiency of the system without and with energy storage was found in the range of 50.18 to 66.58% and 54.71 to 68.37% respectively. They concluded that the energy storage in the drying system results in improved energy and exergy efficiencies for drying operation.

The detailed literature review reveals that the exergy analysis is an effective tool to investigate the quality of the drying process and limited literature discusses the effect of mass flow rate on the exergetic performance as well as the pickup efficiency of the solar drying process. Most of the studies involved with the energy and exergy performance analysis are not explored the sustainability of the drying process. Further, there is no information found on exergy analysis on drying of bitter gourd slices and as well the effective moisture diffusivity values of bitter gourd slices under solar drying process. In this study, a low cost indirect solar dryer integrated with porous bed sensible heat storage medium (pebbles) was developed with the locally available materials suitable for rural area. Based on the literature, the main objectives of this present work are coined as (i) to investigate the effect of mass flow rate of air on the exergic performance of the system for drying bitter gourd slices (ii) to estimate the effective moisture diffusivity value of bitter gourd slices for solar drying and open sun drying methods to analyse the thin layer drying behaviour of the system (iii) to perform environmental impact analysis to study the effectiveness of the system.

Journal Pre-proof

2. MATERIALS AND METHODS

2.1. Material

Momordica charantia is the scientific name of bitter gourd and also known as bitter melon, balsam pear, pare, or karela. It is a widely grown and consumed vegetable in Asia, East Africa, India, and South America. Bitter gourd is rich in carbohydrates, minerals, proteins, nutritional values such as vitamin C, vitamin A, calcium, potassium, iron and fiber (Lucas et al. 2010). It is well known for its medicinal properties such as antidiabetic, anticancer, antivirus, cholesterol lowering effects, and anti-inflammation. Traditionally, it is also used for treating ailments like digestive disorders, ulcer, hyper tension, microbial infections, hyperlipidemia and menstrual problems (Jospe and Jini, 2013). Bitter gourd is also perishable vegetable as it contains a very high moisture content of 92 % wet basis (Vijayan, 2016).

2.2. Experimental setup

The indirect forced convection solar dryer was constructed and installed at Solar Energy Research Centre, Coimbatore Institute of Engineering and Technology, Coimbatore (India). Experimental setup is comprised of a centrifugal blower, flat plate collector with packed bed sensible heat storage medium and a drying chamber. The flat plate collector dimensions are 2m length, 1m width and 150mm height. It is provided with a black coated Galvanized Iron sheet corrugated absorber plate to receive maximum amount of solar radiation. In order to reduce the top heat loss coefficient, a window glass cover of 4 mm thick was mounted 30 mm above the absorber plate along with the necessary rubber beadings to completely prevent the possibility of leakage of air. The schematic diagram of the experimental setup is shown in Fig. 1.

Journal Pre-proof

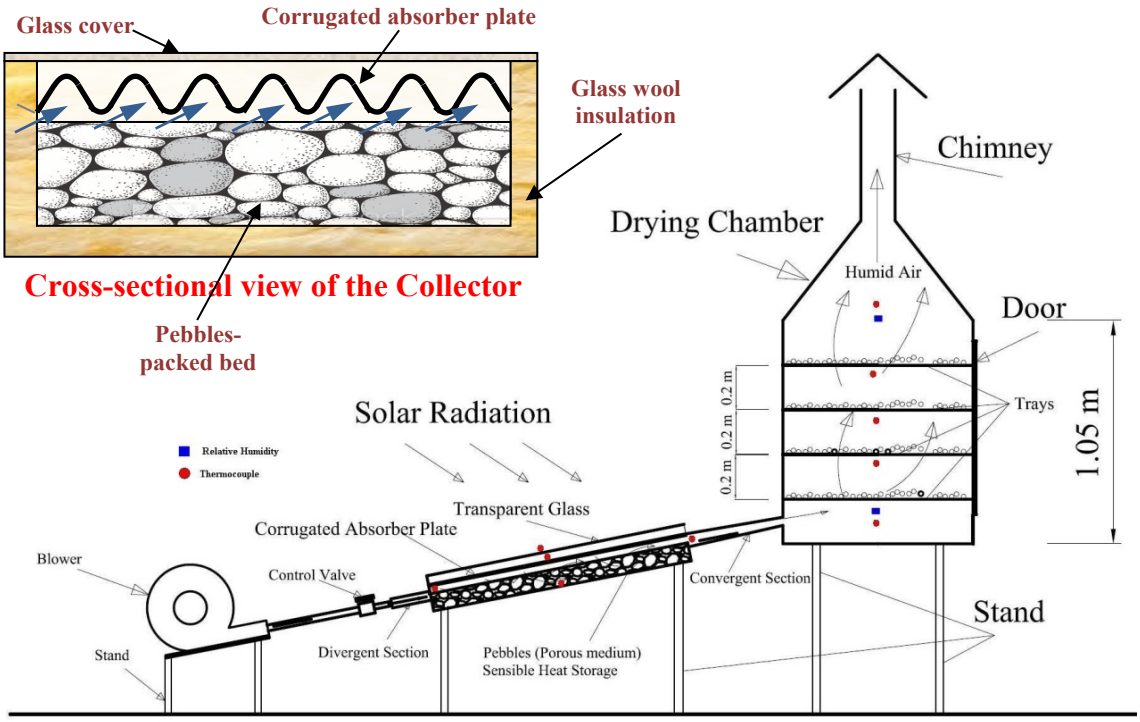


Fig.1 Schematic Diagram of the Experimental Setup

The gap between the absorber plate and the back plate was filled with pebbles of approximately 30mm size to height of 60mm from bottom to store the excessive heat during peak sunshine hours and release the same during off sunshine hours. The bottom and sides of the collector were insulated with glass wool of 50mm and 25mm thickness respectively to reduce the conductive and convective heat loss to environment. The arrangements of the insulation material, packed bed, the corrugated absorber plate and the air flow direction are clearly shown in cross sectional view of the collector (Fig. 1). The air flow takes place between the absorber and back plate of the collector in the perpendicular direction to the cross-section of the collector as shown in Fig. 1. To ensure long life of the system outer surfaces of glass wool was completely covered with 1.2 mm thick galvanized iron sheets and the exposed parts of the system were completely painted to protect it from environmental effects. The collector was positioned at an angle equal to the latitude of the location (i.e 11° with the horizontal) to receive the maximum solar radiation (Elminir et. al., 2006),. A divergent section was provided at the entry of the solar collector to ensure the fully developed flow of air. A centrifugal blower was used to force the hot air from the collector to drying chamber and a manually operated ball valve was mounted between the blower and collector to adjust the flow rate of air to the system. The drying chamber was made with 1.2mm thick galvanized iron sheet for the dimensions of 0.75m × 0.75m × 1.05m. It has four perforated

aluminium trays to spread the wet products. All the sides of the drying chamber were insulated with the glass wool of 25mm thick and an insulated door was one side of the drying chamber. An insulated convergent section was used to connect the collector with the drying chamber. The photograph of the developed system and the arrangements of the bitter gourd slices in drying experiments are shown in Fig. 2.

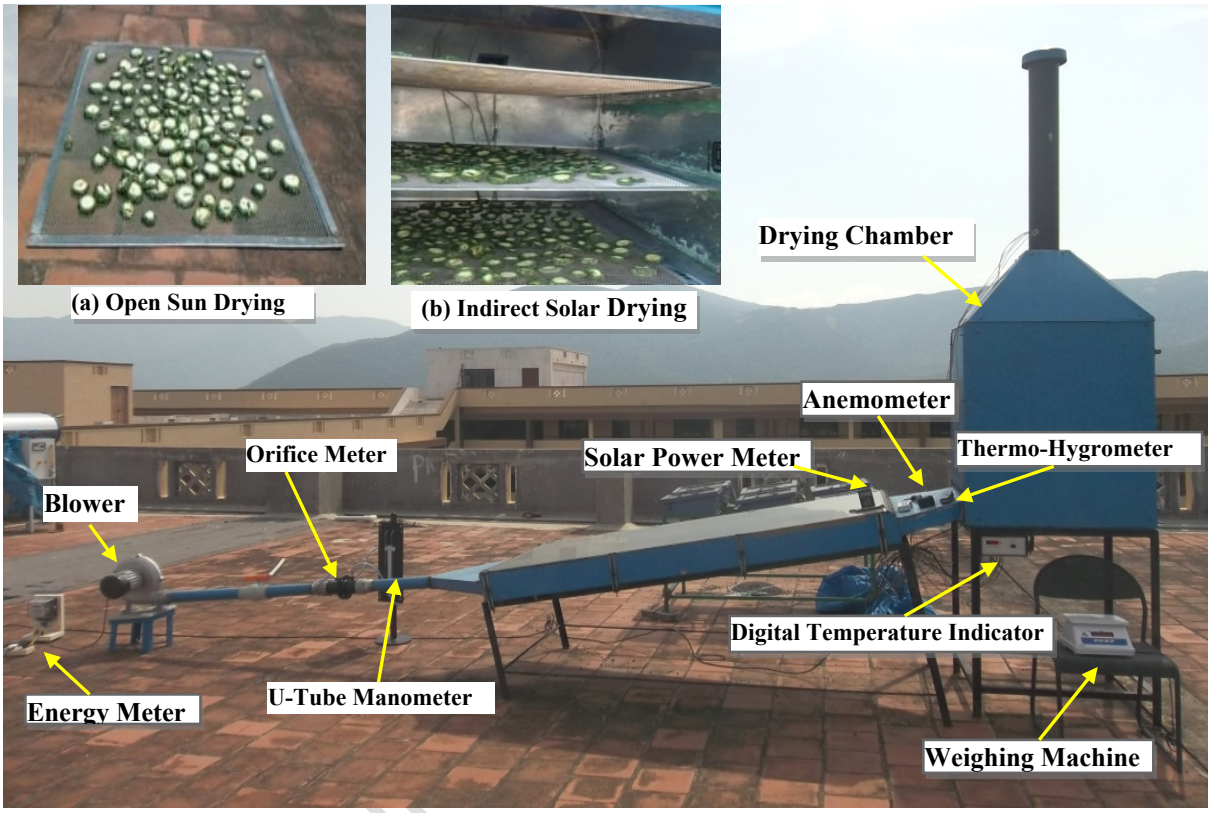


Fig.2 Experimental setup with various instruments and arrangements of bitter gourd slices (a) open sun drying (b) indirect solar drying

The solar insolation was measured using TES (Taiwan) make solar power meter having an accuracy of ±10 W/m^2. Temperature values at various locations were measured using calibrated k-type thermocouples connected with a digital temperature indicator having resolution of 0.1°C. The relative humidity of ambient air, drying chamber inlet and outlet hot air were observed using a thermo-hygrometer having an accuracy of ± 2% and resolution of 0.01%. The mass flow rate of air through the system was measured using orifice meter with a U tube manometer. Weight of the samples was checked using a digital weighing machine having a resolution of 1g.

The drying experiments were conducted with the bitter gourd slices using an indirect forced convection solar dryer with packed bed heat storage with the mass flow rate of air from 0.0141 kg/s to 0.0872 kg/s. In all the drying experiments, around 4 kg of sliced bitter gourd in indirect solar drying (1 kg in each tray) and 1 kg in open sun drying are loaded. In all the cases of drying experiments products are loaded at 9.00 am and continued till the product reaches the stagnation point. The experimental setup was allowed to run for 30 minutes (to attain steady state) before loading the products into the drying chamber.

2.3. Experimental uncertainty

Uncertainties in the experimental studies are inevitable due to the reasons of the improper instrument selection, environmental conditions, instrument accuracy, readability and human errors. In drying experiments, it is very important to estimate the maximum possible uncertainties in the independent variables (temperature, velocity, solar radiation and weight) and calculated parameters (mass flow rate, efficiency, heat gain, drying rate). The uncertainties of the variables are temperature measurement (w_T) ± 0.1 °C, relative humidity (w_{RH}) ± 3 %, air velocity (w_v) ± 0.1m/s, solar radiation (w_I) ± 10 w/m^2, and weight (w_m) ± 0.001 kg. The result R is a given function in terms of the independent variables. Let w_R be the uncertainty in the result and w_1, w_2,……,w_n be the uncertainties in the independent variables. The result R is a given function of the independent variables x_1, x_2, x_3,….., x_n. The uncertainty in the result can be calculated as (Akpinar, 2010):

$$w_R = \sqrt{\left(\frac{\partial R}{\partial x_1} w_1\right)^2 + \left(\frac{\partial R}{\partial x_2} w_2\right)^2 + \ldots + \left(\frac{\partial R}{\partial x_n} w_n\right)^2} \quad (1)$$

3. PERFORMANCE ANALYSIS

The performance of the indirect forced convection solar dryer integrated with porous bed thermal energy storage medium for drying of bitter gourd slices was investigated using the parameters moisture ratio, drying rate, pick-up efficiency, dryer thermal efficiency, exergy efficiency and exergy sustainability indicators. The instantaneous moisture content of the products can be determined using the Equation (2) in wet basis (Fudholi et al., 2014):

$$M_t = \frac{W_t - W_d}{W} \quad (2)$$

where, W_t is the weight of the product at instant t in kg, W_d is the dry weight in kg and W is the total weight of the product in kg.

The moisture ratio (MR) is the ratio between the moisture content at time 't' and the initial moisture content of the product and it is determined as (Akpinar, 2010; Vijayan, 2016):

$$MR = \frac{(M_t - M_e)}{(M_i - M_e)} \tag{3}$$

The drying rate of the products can be found using the following Equation (4) (Shrivastava & Kumar, 2016):

$$DR = \frac{dM}{dt} = \frac{M_{t+\Delta t} - M_t}{\Delta t} \tag{4}$$

where, M_t is the moisture content of the product and Δt is the time period.

The pick-up efficiency of the solar dryer is another important parameter for evaluating the performance of the system. It is a parameter which determines the evaporative capacity of the drying chamber. The pick-up efficiency can be calculated using the Equation (5):

$$\eta_{pick\ up} = \frac{h_o - h_i}{h_{as} - h_i} = \frac{W_o - W_t}{m_a t(h_{as} - h_i)} \tag{5}$$

The thermal efficiency of the dryer is the ratio of energy utilized effectively to evaporate the moisture content from the products to the total energy supplied. It is calculated using the Equation (6):

$$\eta_d = \frac{Moisture\ evaporated\ (kg) \times latent\ heat\ of\ evaporation\ (J/kg)}{\text{Total Energy Supplied (J)}} \tag{6}$$

The exergy values are calculated by using the characteristics of the working medium from a first-law energy balance. For this purpose, the general form of exergy equation applicable for a steady flow system may be expressed as in Equations (7), (8) & (9) (Fudholi *et. al.* 2014):

$$Ex = \dot{m}_a C_p \left[(T - T_a) - T_a ln\left(\frac{T}{T_a}\right) \right] \tag{7}$$

For exergy inflow of drying chamber

$$Ex_{dci} = \dot{m}_a C_p \left[(T_{dci} - T_a) - T_a ln\left(\frac{T_{dci}}{T_a}\right) \right] \tag{8}$$

For exergy outflow of drying chamber

$$Ex_{dco} = \dot{m}_a C_p \left[(T_{dco} - T_a) - T_a ln\left(\frac{T_{dco}}{T_a}\right) \right] \tag{9}$$

However, during the solar drying process, the exergy losses are determined using the Equation (10):

$$Ex_{loss} = Ex_{dci} - Ex_{dco} \quad (10)$$

The exergy efficiency can be defined as the ratio of exergy use (investment) in the drying of the product to exergy of the drying air supplied to the system. However, it is explained as the ratio of exergy outflow to exergy inflow for drying chamber. Considering this definition, the exergy efficiencies of drying chamber can be determined. Thus, the general form of exergy efficiency is written as Equation (10):

$$\eta_{Ex} = \frac{Ex_{dco}}{Ex_{dci}} = 1 - \frac{Ex_{loss}}{Ex_{dci}} \quad (11)$$

Exergy sustainability indicators are the parameters used to identify, how the exergy efficiency and loss values affect the sustainable development of the drying process. These parameters are useful in achieving more efficient, environmentally, sustainable and economic energy use in drying systems. The sustainability indicators used in this study are improvement potential (IP), waste energy ratio (WER) and sustainability index (SI). Caliskan at al. (2017) highlighted that the sustainability of the system varies with the temperature and mass flow rate of air. The behavior of the sustainability indicators is almost similar to the exergy efficiency and they can be expressed as follows (Atalay, 2019):

$$IP = (1 - \eta_{Ex}) Ex_{loss} \quad (12)$$

$$WER = \frac{Ex_{loss}}{Ex_{dci}} \quad (13)$$

$$SI = \frac{1}{1 - \eta_{Ex}} \quad (14)$$

3.1. Determination of effective moisture diffusivity

The rate of moisture movement in the product during drying is described using a term known as Effective moisture diffusivity (D_{eff}). Effective moisture diffusivity, which is affected by composition, moisture content, temperature and porosity of the material, is used due to the limited information on the mechanism of moisture movement during drying and complexity of the process. The drying rate of the process can be stated using the moisture ratio for a long drying process (Doymaz, 2007) as given in Equation (15) & (16):

$$\frac{dM}{dt} = D\frac{d^2M}{dr^2} \tag{15}$$

$$MR = \frac{M_t - M_e}{M_0 - M_e} = \frac{8}{\pi^2}\exp\left(-\frac{\pi^2.D_{eff}}{4\,r^2}.t\right) \tag{16}$$

where, D_{eff} is the effective diffusivity coefficient (m^2/s); r is the thickness of the slab (m), t is the drying time (sec) and n is the positive integer. Only the first term of Equation (16) can be used for long drying times (Lopez 2000): Several researchers demonstrated that Equation (16) could be further simplified to a straight-line equation as Equation (17) (Sacilik 2006; Wang *et. al.* 2007):

$$\ln(MR) = \ln\left(\frac{8}{\pi^2}\right) - \left(\frac{\pi^2.D_{eff}}{4\,r^2}.t\right) \tag{17}$$

The effective moisture diffusivity (D_{eff}) is typically determined by plotting experimental drying data in terms of ln (MR) versus time and the slope (k_o):

$$k_0 = \frac{\pi^2\,D_{eff}}{4\,r^2} \tag{18}$$

where, r bitter gourd half thickness of bitter gourd slices in mm

3.2. Environmental Impact Analysis

Any built facility consumes a major quantity of energy during its life cycle stages construction, operational and disposal. The total energy consumption can be categorized into two components such as embodied energy (EE) and operational energy (OE). Embodied energy is defined as the total amount of energy spent during the extraction of raw materials, the transport, manufacturing, construction, use (maintenance and renovation) and disposal phases (Alwan & Jones 2014). The operational energy is the energy required to supply for its operation or application periods. The operational energy of a system comparatively occupies a large portion of total energy consumption. However, the energy efficient structures design, the embodied energy is an increasingly high proportion of the total life-cycle energy consumption.

In recent years, reduction of the total energy (EE and OE) consumption in built facilities is drawing the attention of the investigators in order to reduce the utilization of conventional energy resources. Solar dryers are the systems operating with the solar energy, which is freely available, eco-friendly and clean renewable energy, therefore, the solar dryers are considered as one of the energy efficient structures. The energy effectiveness of the

system can be evaluated by the total energy consumption and the environmental impact of the system is assessed with the indicators like embodied energy, energy payback period, and emission of greenhouse gas during its life cycle.

3.2.1. Energy payback time (EPBT)

The energy payback time (EPBT) is the term used to express the time period required to pay back the energy spent or utilized during the production of the raw materials used for the construction of the system. It can be found using the following Equation (19) (Kesavan et al., 2019):

$$Energy\ payback\ time = \frac{Embodied\ energy\ (kW\mathrm{h})}{Annaul\ Energy\ output\ (kW\mathrm{h}/yr)} \quad (19)$$

The annual thermal energy output of the dryer can be evaluated using (20):

$$Annual\ thermal\ Energy\ output\ (E_{aout}) = Daily\ thermal\ output\ \times N_d \quad (20)$$

where N_d is the total number sunshine days in a year, in India approximately 290 days are with good sunshine but here it is assumed as 250 days. The daily thermal output of the dryer can be determined as follows (21):

$$Daily\ thermal\ output\ of\ dryer\ (kW\mathrm{h}) = \frac{Moisture\ evaporated\ (kg) \times latent\ heat\ of\ evaporation\ (J/kg)}{3.6\ \times 10^6} \quad (21)$$

3.2.2. Carbon Dioxide (CO_2) emission

The emission of average carbon dioxide (CO_2) is approximately equivalent to 0.98 kg of CO_2/kWh in the case of electricity generated by coal. The CO_2 emission per year can be calculated using the Equation (22) (Chauhan et al. 2018):

$$CO_2\ emissions\ per\ year = \frac{Embodied\ Energy\ \times 0.98}{Life\ time} \quad (22)$$

If the transmission and domestic losses accounted for the estimation, then the Equation (22) becomes

$$CO_2\ emissions\ per\ year = \frac{E_{in}}{L} \times \frac{1}{1 - L_a} \times \frac{1}{1 - L_t} \times 0.98\ kg \quad (23)$$

where E_{in} is embodied energy, L is life time of dryer, L_t is transmission losses and L_a is domestic appliances losses. If the losses are assumed as L_t =40% and L_a =20% due to old appliances (Tiwari & Tiwari, 2016):

Then the Equation (23) becomes

$$CO_2\ emissions\ per\ year = \frac{E_{in}}{L} \times 2.042\ kg \qquad (24)$$

3.2.3. CO_2 mitigation

CO_2 mitigation of indirect solar dryer can be calculated using the Equation (25):

$$The\ net\ CO_2 mitigation\ over\ the\ lifetime = Total\ CO_2\ mitigation - Total\ CO_2 emission$$

$$= (E_{aout} \times L - E_{in}) \times \frac{1}{1-L_a} \times \frac{1}{1-L_t} \times 0.98\ kg \qquad (25)$$

$$Annual\ CO_2 mitigation = (E_{aout} \times L - E_{in}) \times 2.042\ kg \qquad (26)$$

where, L is the life time of the indirect solar dryer (5,15, 25, 35 years), E_{in}- embodied energy, E_{aout}- annual energy output

3.2.4. Carbon Credit Earned

One carbon credit is equal to the mitigation of one ton (1000 kg) of emission and the earned credit of solar dryer is calculated using the Equation (27) (Tiwari & Tiwari, 2016):

$$Earned\ Carbon\ credit = net\ mitigation\ of\ CO_2\ in\ life\ time \times Price\ per\ ton\ of\ CO_2\ mitigati \qquad (27)$$

The input parameter values are indicated in Table 1.

Table 1 Specifications and input parameters of indirect solar dryer

Aperture Area	2 m^2
Dimensions of the collector	2m × 1m
Number of glazing	1
Absorber plate material	Black painted corrugated GI sheet
Packed Bed Material	Pebbles
Size of pebbles	Average size 30-50 mm dia (approx.)
Quantity of pebbles	110 kg
Porosity	0.85
Specific heat of pebble	835 J/kg K
Mass flow rate of air	0.0141 to 0.0872 kg/s
Initial moisture content of bitter gourd slices	92 % (wet basis)
Specific heat of air	1008 J/kg K
Latent heat of vaporization of water	2430 kJ/kg

Thickness of the bitter gourd slice	5 - 7 mm
Bitter gourd diameter	43-51 mm
Drying time	6 to 10 hours
No. of sunny days in a year	250 days
Expected life time of the system	15 years

4. RESULTS AND DISCUSSIONS

The exergy and environmental impact analysis of the indirect forced convection solar dryer integrated with sensible heat storage material for drying bitter gourd slices were analyzed using the experimental observations. The experimental observations were made during the summer periods and the days with similar solar insolation profiles were only accounted for further data analysis and the average insolation values of the days were around 720 W/m^2. The uncertainties in the calculated parameters of mass flow rate, useful collector gain, moisture content, drying rate, pickup and exergy efficiency are ±0.141, ±0.1997, ±0.141, ±0.244, ±0.223, and ±0.173 respectively and all these values are within the permissible limit.

Fig. 3 shows the hourly variations of solar insolation, absorber plate temperature, collector inlet and outlet temperatures, and ambient temperature for the mass flow rate of 0.0872kg/s. The solar insolation value was reached the maximum value of 996W/m^2 at 13:00 hrs. The temperature of collector components was varied with respect to solar insolation i.e increases with increase in solar insolation and vice versa. The maximum collector outlet temperature of air was attained as 45.3°C at the time of peak solar insolation period and the corresponding absorber plate temperature was noted as 68.2°C. Interestingly, the collector inlet temperature was 1 to 8°C higher than the ambient temperature; it is due to the reason of blower work.

Journal Pre-proof

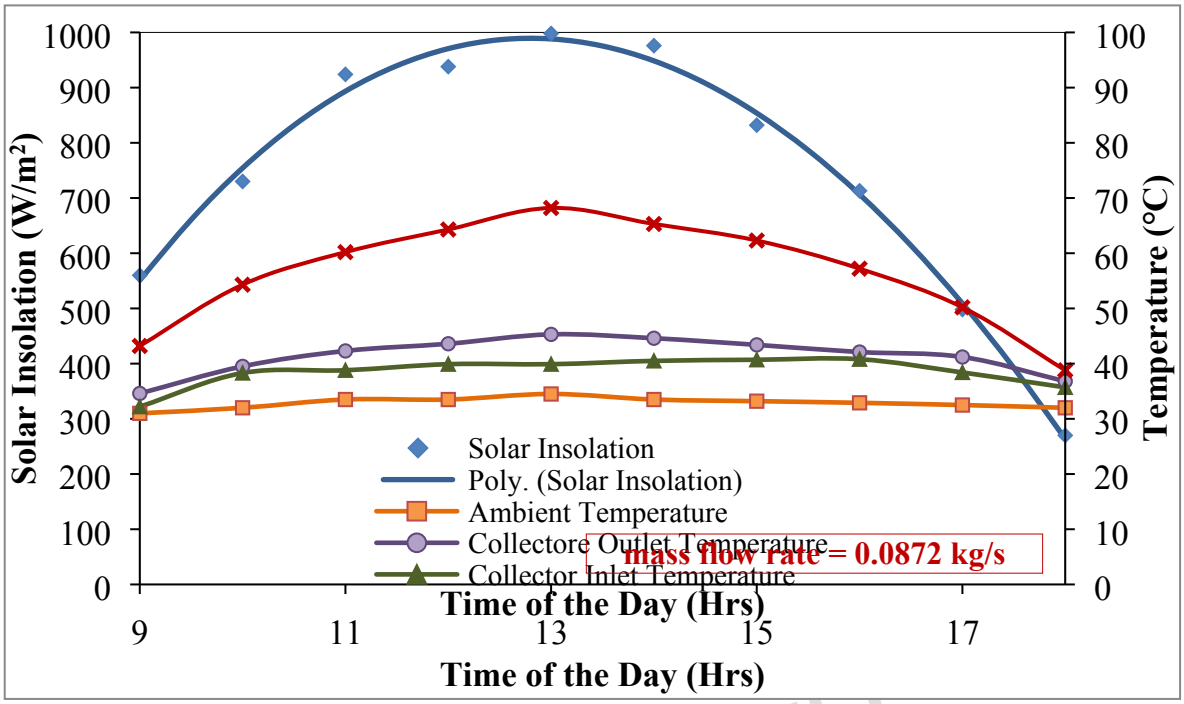


Fig. 3 Solar insolation and temperatures of various elements of collector

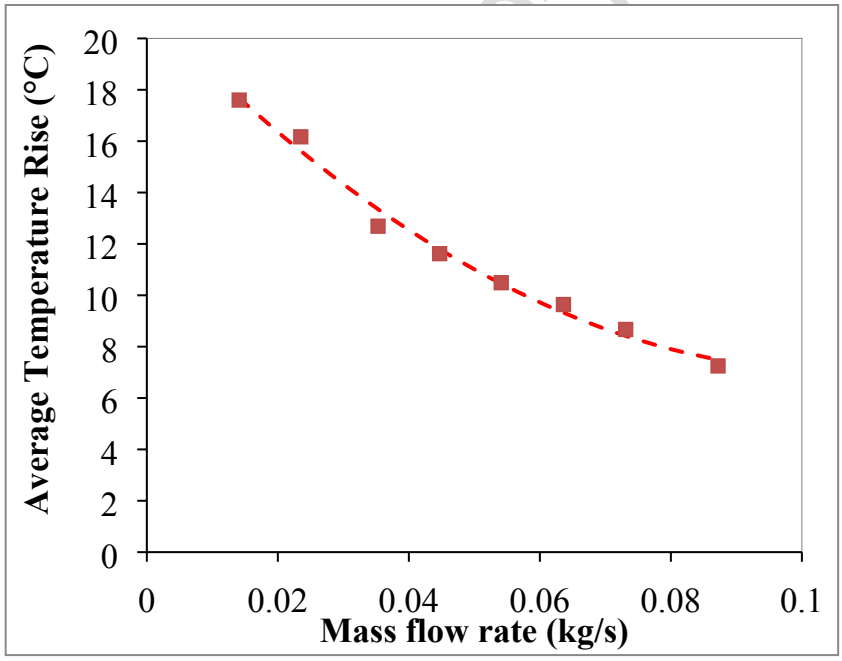


Fig. 4 Average Temperature Rise for different mass flow rates of air

The average temperature rise for various mass flow rates from 0.0141 to 0.0872 kg/s are shown in Fig. 4. The temperature rise decreases when the mass flow rate of air increases, it is due to the fact that the residing or period of contact with the absorber (hot) surface is less when the velocity of air increases. However, the useful energy gained by the air increases

with increase in velocity. The maximum and minimum average temperature rise obtained were 17.6 °C and 8.2 °C for the mass flow rates 0.0141 and 0.0872 kg/s respectively. The measured average velocity of the hot air from outlet of the solar collector i.e inlet of the drying chamber was measured using the anemometer is indicated in Table 2. The velocity of the air increases with increase in mass flow rate of air through the system. The velocity is reduced immediately after entering the drying chamber due to enlarged volume of the chamber. The temperature and velocity of the hot air are the important parameters for the drying process (Mohan and Talukdar, 2013).

Table 2 Average velocity of the air at the drying chamber inlet

Mass flow rate of air (kg/s)	0.0141	0.0235	0.0353	0.0447	0.0541	0.0636	0.0731	0.0872
Average velocity at the outlet of collector (i.e entrance of the drying chamber) (m/s)	0.50	0.83	1.25	1.58	1.92	2.26	2.59	3.09

4.1. Drying Kinetics of Bitter Gourd Slices

The hourly variations of moisture content of sliced bitter gourd for different mass flow rates of air in solar drying and open sun drying are shown in Fig. 5. It clearly indicates that the moisture content of the bitter gourd exponentially decreases with increase in drying time. Only for the mass flow rate of 0.0141 kg/s the drying process was continued for the next day and for the remaining flow rates the drying process was completed within a day. In solar drying, the drying time of bitter gourd slices was varied from 7 hours to 10 hours for different mass flow rates of air. For the lower mass flow rates, the drying process took longer period due to the reason of insufficient velocity to cause the movement of moisture from the interior to the surface of the product as well as from surface to surrounding air. The initial moisture content in 4 kg of bitter gourd slices was reduced from 3.65 kg of water (92% wet basis) to a safe moisture content of 0.35kg of water (9% w.b) in 7 hours using indirect solar drying at the mass flow rate of 0.063kg/s. But in open sun drying the initial moisture content of around 0.85 kg of water is brought to 0.090kg of water in 10 hours.

Journal Pre-proof

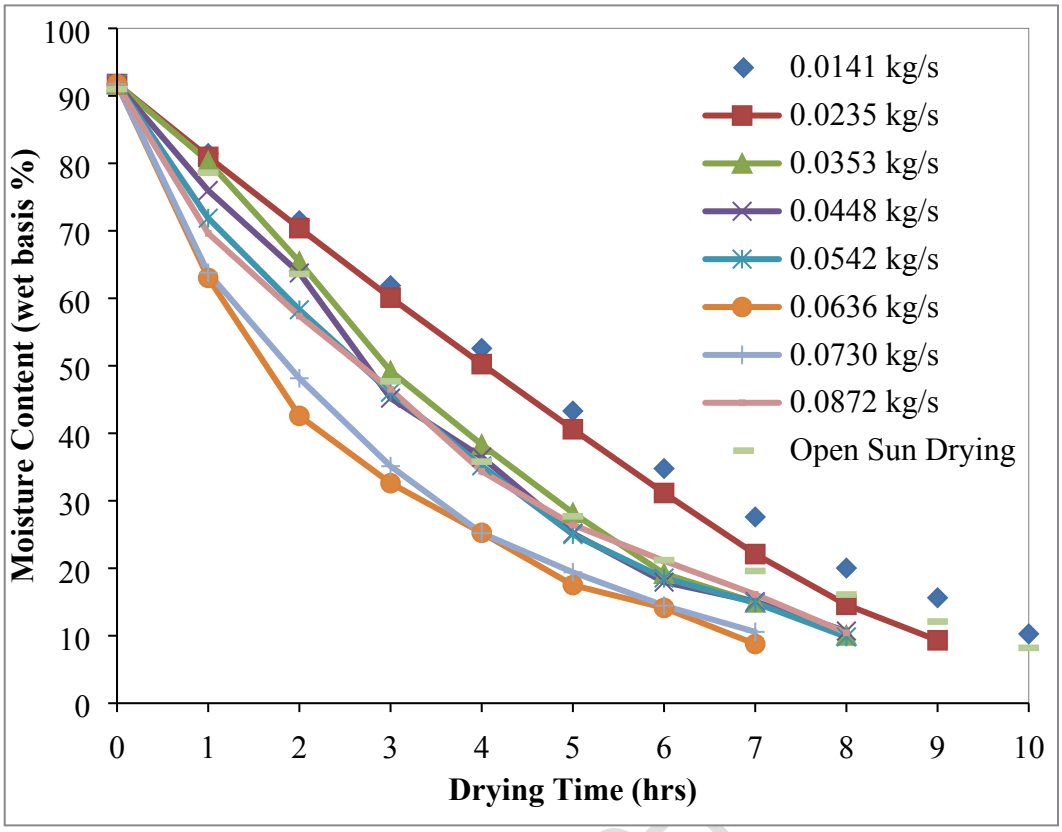


Fig. 5 Moisture content for open sun and solar drying of bitter gourd slices

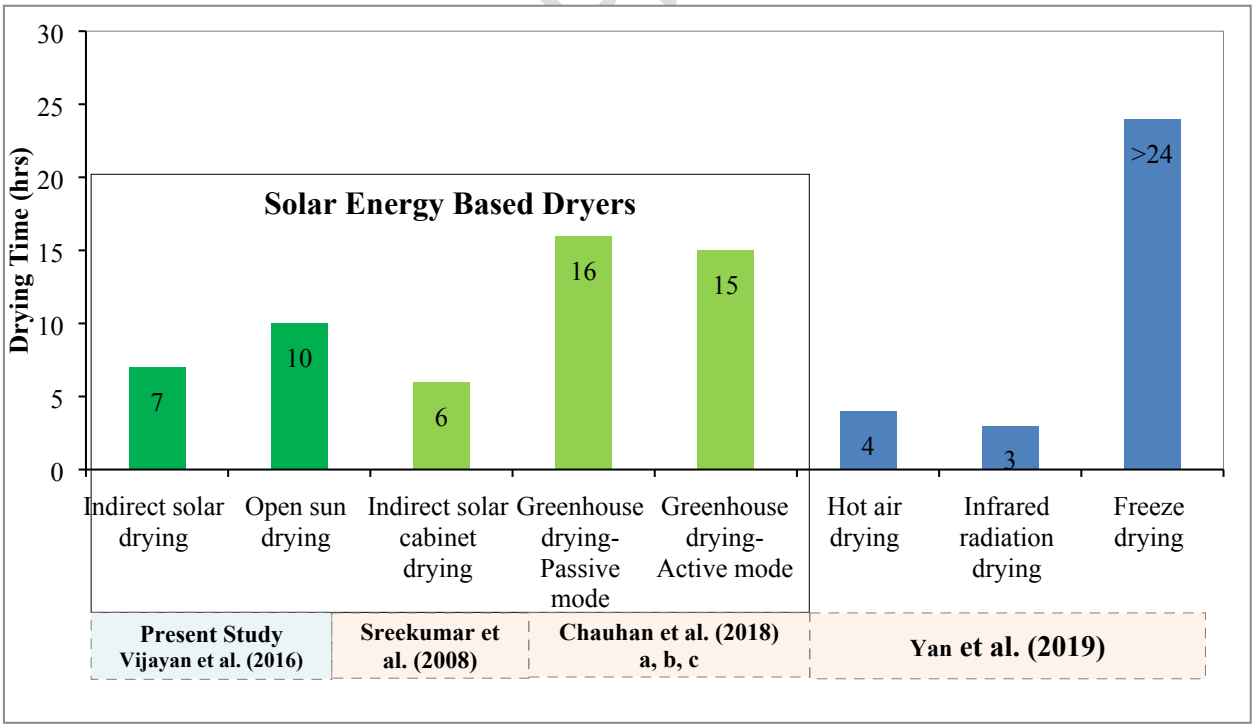


Fig. 6 Drying time of bitter gourd slices in various dryers

Drying results of the bitter gourd slices in indirect solar dryer is compared in Fig. 6 with the results of other type of dryers reported in the literature such as solar cabinet dryer, greenhouse dryer, hot air dryer, infrared radiation drying and freeze drying. The drying time of indirect solar dryer is lower than the freeze drying (more than 24 hours), greenhouse drying as well as open sun drying methods, on the other hand the hot air and infrared radiation dryers are having less than the half of the drying time of indirect solar drying. However, these methods are high energy consumption, high cost, low efficiency and inferior quality of the dried products (Yan et al., 2019). The indirect solar dryers are capable of producing superior quality dried products at economically as well eco-friendly way (Vijayan et al., 2016).

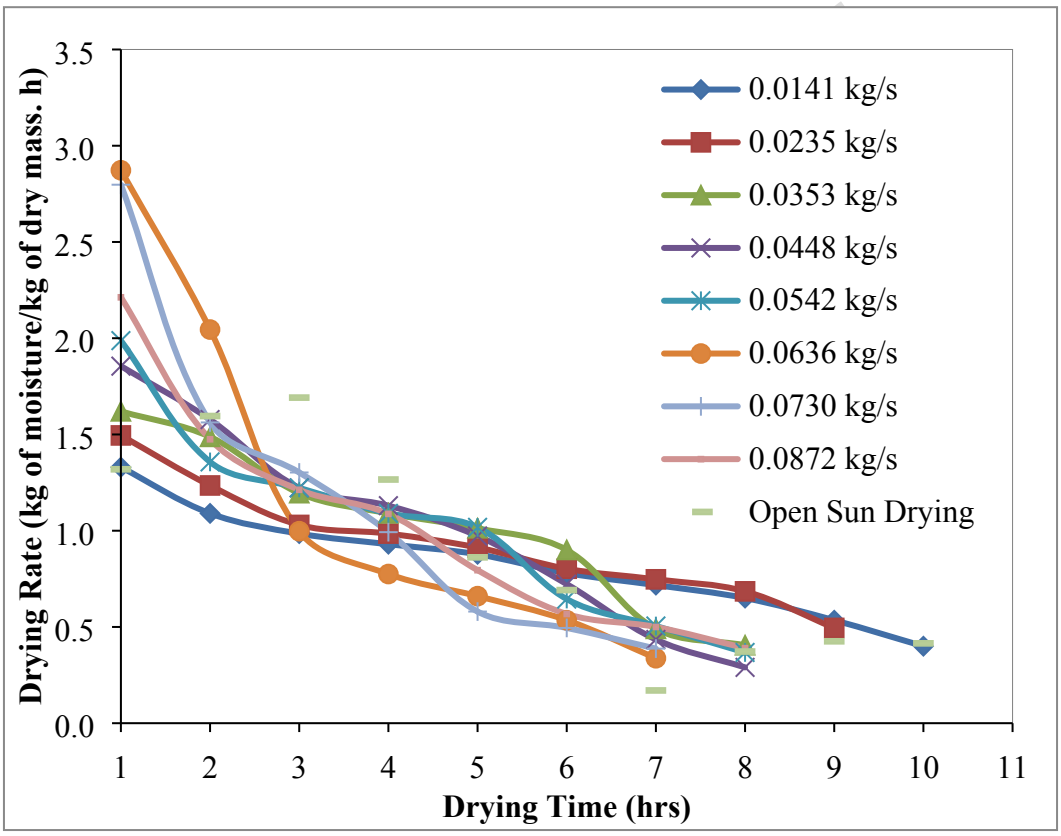


Fig.7 Drying rate with drying time in solar drying of bitter gourd slices for various mass flow rates

Fig.7 shows the variations of drying rate (kg of moisture/kg of dry mass. h) with drying time in solar drying of bitter gourd slices for the mass flow rates of 0.0141 kg/s to 0.0872 kg/s. The drying rate mainly depends on physical characteristics of the product and the external parameters of the surrounding air (temperature, velocity, flow rate and relative

humidity). The drying rate increases with increase in mass flow rate till it reaches 0.0636 kg/s and further increase in mass flow rate decreases the drying rate. The combined effect of velocity and temperature of air causes increase in drying rate and reduces the drying time and the velocity of air may be higher for the air beyond 0.0636 kg/s mass flow rate but the temperature of the air is insufficient to induce the migration of moisture from the interior of the product to its surface. Variation of drying rate with the moisture content (wet basis) of bitter gourd slices for various mass flow rates of air is shown in Fig. 8. The drying rate curve of open sun drying is exhibiting some fluctuation due to high dependency of solar insolation and the surround air temperature (Vijayan et al, 2017). Drying rate curve for all the mass flow rates containing only the falling rate period, it means the constant drying rate period is absent. The drying rate curve for the mass flow rates 0.0636, 0.0730 and 0.0872 kg/s are steeper than the other drying rate curves. For the lower mass flow rates, the drying rate is less due to the insufficient velocity at the drying chamber.

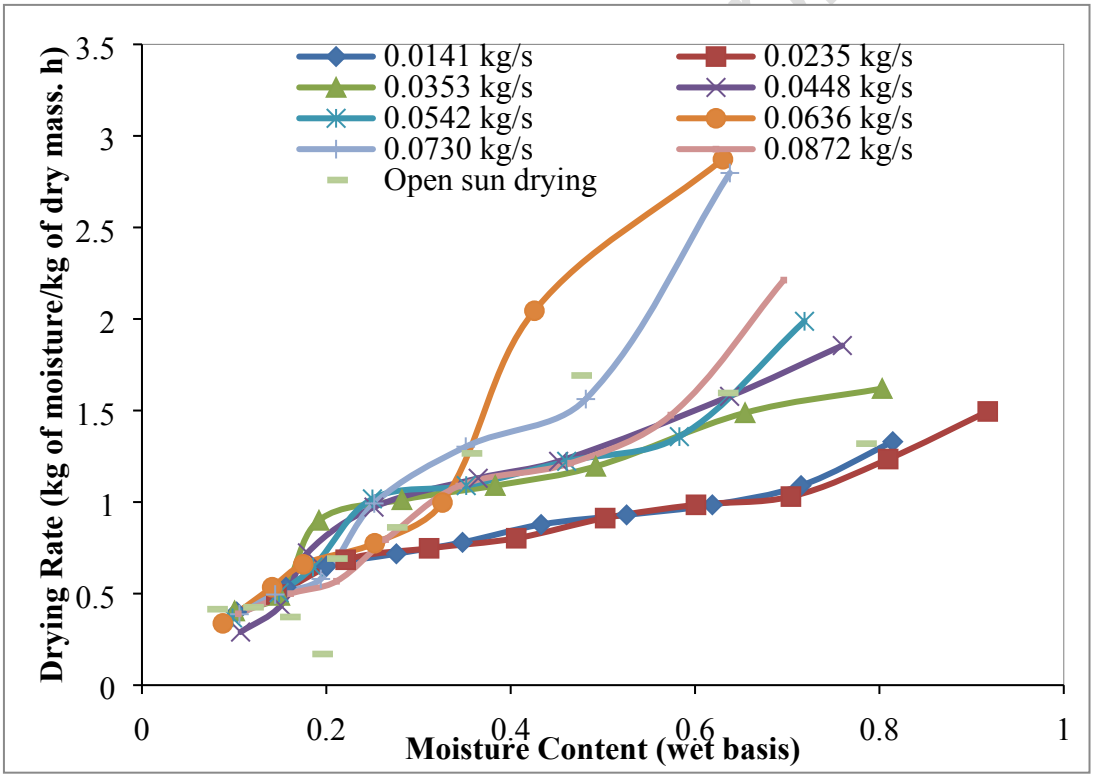


Fig. 8 Moisture content with drying rate at different flow rate of air

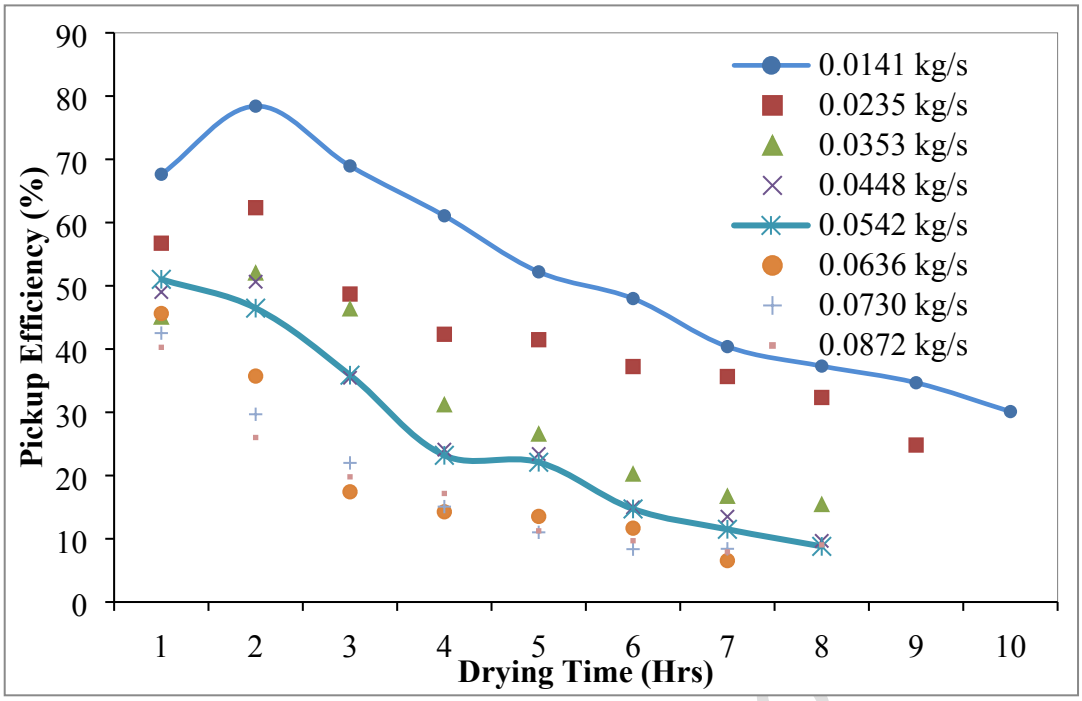
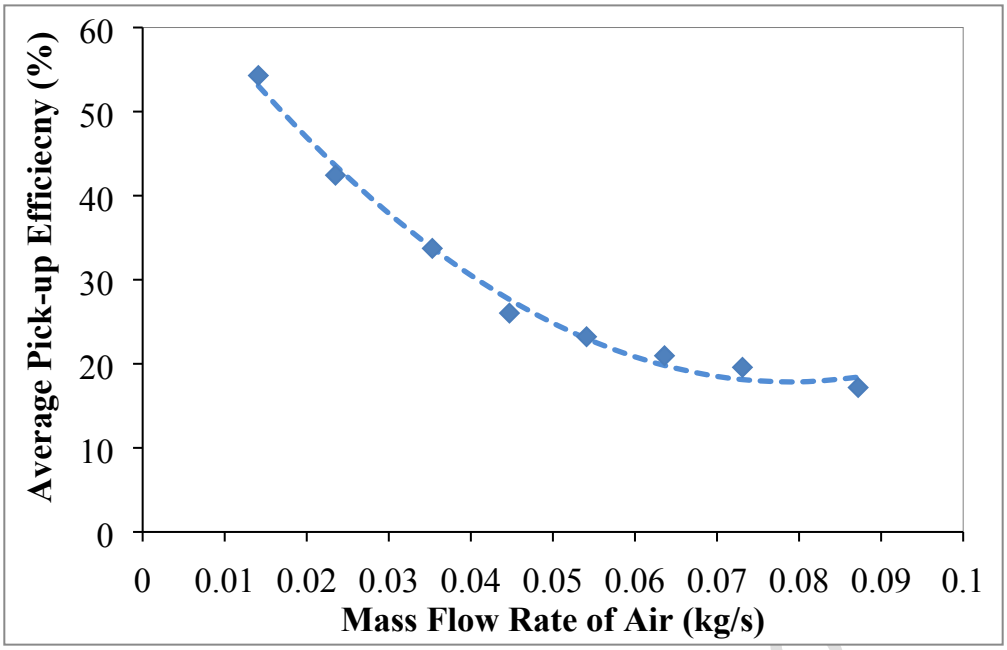


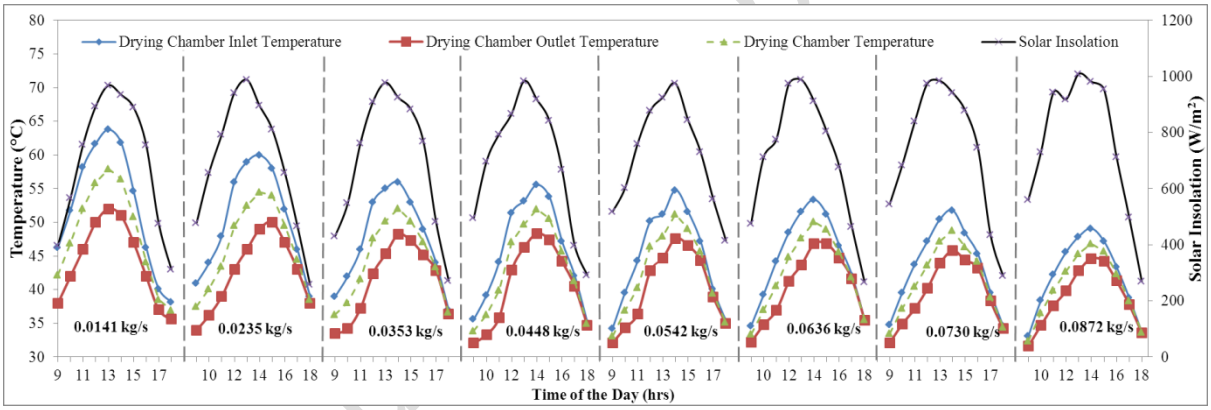
Fig.9 Pickup efficiency with drying time for various mass flow rates of air

The hourly variations pickup efficiency of the convective air with drying time for various mass flow rates of air from 0.0141 to 0.0872kg/s are shown in Fig. 9. During the initial period of drying the pickup efficiency is high and decreases when the drying progresses. It is due to evaporation of free moisture content available at the surface of the product. Over the drying period, the free moisture content available at the surface of the product decreases and the rate at which the migration of moisture from the interior to the surface is also decreased. The maximum and minimum pickup efficiency observed for the typical mass flow rate of 0.0141 kg/s is 78.4% and 30.11% respectively. The variations of average pickup efficiencies for various mass flow rate of air are shown in Fig. 10. The pickup efficiency exponentially decreases with increase in mass flow rate of air. Pickup efficiency is the ability of the surrounding air to evaporate the moisture from the product and the temperature of air plays a significant role in inducing the drying process. But the increase in mass flow rate of air decreases the dying chamber inlet temperature which results in reduction in moisture pickup potential of air. The variations of drying chamber inlet, outlet and chamber temperature of air for various mass flow rates are shown in Fig. 11. It clearly indicates the deceasing trend of the drying chamber temperature and more over the temperature profile follows the solar insolation of the day.



458
459
460 **Fig.10** Pickup efficiency with mass flow rate of air

461



462
463 **Fig. 11** Drying chamber inlet, outlet and average temperature of air with time

464

465 Variation of overall dryer efficiency with mass flow rates of air for drying bitter gourd
466 slices is shown in Fig. 12. Dryer thermal efficiency value varies from 12.11% to 19.21% for
467 various mass flow rates of air. The dryer efficiency increases with increase mass flow rate till
468 it reaches the mass flow rate of 0.0636 kg/s and further increase in mass flow rate decreases
469 the dryer efficiency. The reduction in efficiency is due to the increase in drying time.

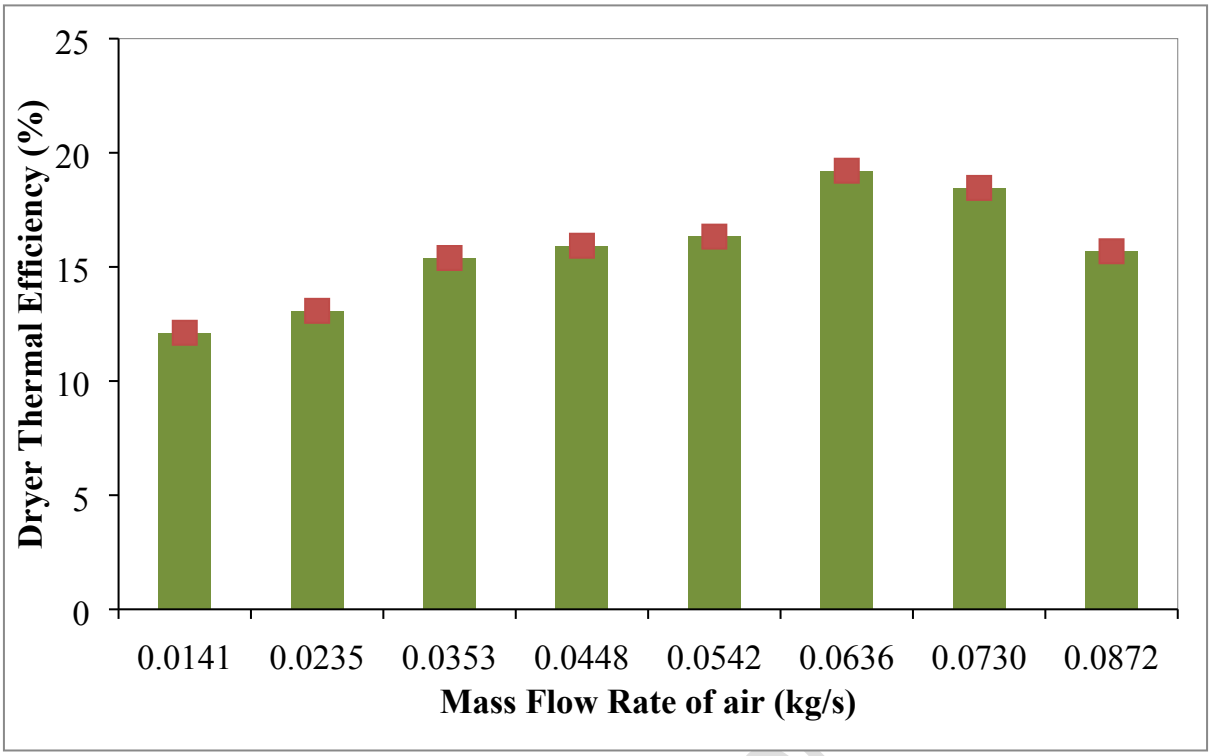


Fig.12 Dryer Thermal Efficiency

The effective moisture diffusivity for bitter gourd slices under various drying conditions was calculated from the Equation (14). Effective moisture diffusivities of bitter gourd slices at a different mass flow rate and open sun drying were obtained from the graph plotted between ln (MR) and drying time (t). Fig. 13 shows the variation of ln (MR) versus drying time in sec for all the mass flow rates of drying and the open sun drying of bitter gourd slices. The values of effective moisture diffusivity for various mass flow rates vary from 8.6293×10^{-10} to 12.9585×10^{-10} m^2/s, whereas for open sun drying it gives the lowest value of 0.9568×10^{-10} m^2/s. It yields the maximum value at the mass flow rate of 0.0636 kg/s due to the fact that the drying time is lower than other flow rates. Table 3 shows the values of line equations, R^2 values and effective moisture diffusivities for mass flow rates.

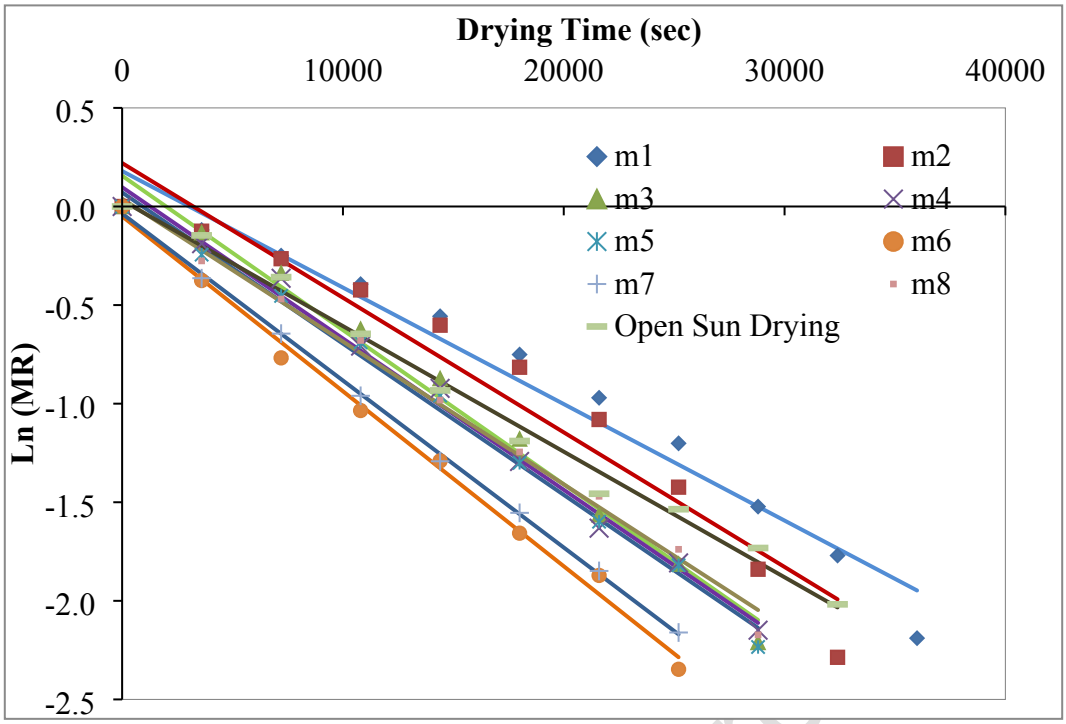


Fig. 13 Drying time vs ln (MR)

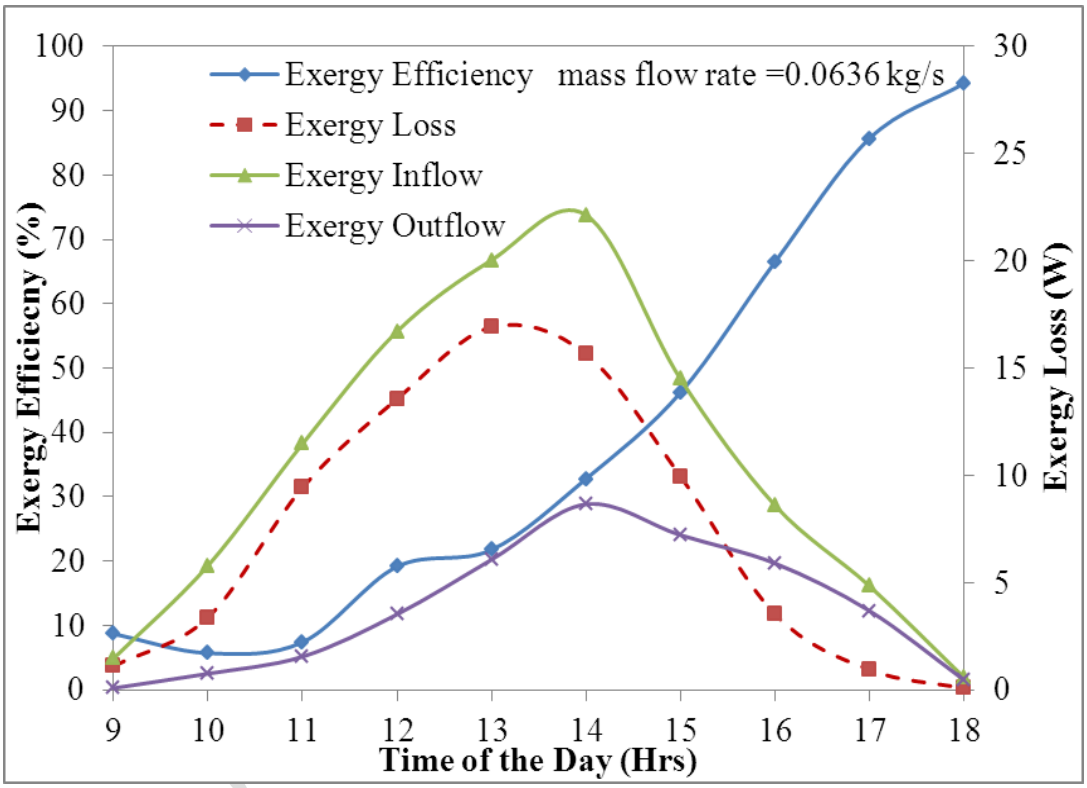
Table 3 Effective Moisture Diffusivity values for solar drying and open sun drying

Mass flow rate (Kg/s)	Equation	R^2	Effective Moisture Diffusivity (m^2/s)
0.0141	$y = -5.9084 \times 10^{-5} + 0.1796$	0.9685	8.6293×10^{-10}
0.0235	$y = -6.8278 \times 10^{-5} + 0.2201$	0.9499	9.9720×10^{-10}
0.0353	$y = -7.8233 \times 10^{-5} + 0.1558$	0.9871	11.4259×10^{-10}
0.0448	$y = -7.6724 \times 10^{-5} + 0.0977$	0.9928	11.2060×10^{-10}
0.0542	$y = -7.6685 \times 10^{-5} + 0.0712$	0.9937	11.1999×10^{-10}
0.0636	$\boldsymbol{y = -8.8726 \times 10^{-5} + 0.0499}$	**0.9945**	$\mathbf{12.9585 \times 10^{-10}}$
0.0730	$y = -8.4679 \times 10^{-5} + 0.03600$	0.9991	12.3672×10^{-10}
0.0872	$y = -7.2426 \times 10^{-5} + 0.0393$	0.9925	10.5778×10^{-10}
Open sun drying	$y = -6.5511 \times 10^{-5} + 0.0499$	0.9919	0.9568×10^{-10}

4.2. Exergetic performance of the dryer

Exergy analysis is very helpful in identifying the potential improvements in the operating conditions. The variations of exergy efficiency, exergy inflow, outflow and loss for

494 the typical mass flow rate 0.0636 kg/s are shown in Fig. 14. Dryer exergy inflow, outflow and
495 loss in the drying chamber increases during the first five hours of drying then it gradually
496 decreases till the end of the day because the exergy depends on the solar radiation. At the
497 peak radiation time, the exergy loss also attains the maximum value and the exergy efficiency
498 gradually increases with increase in time of the day. It shows that the exergy efficiency
499 increases when the drying process progresses because the exergy outflow becomes closer the
500 exergy inflow to the drying chamber. Similar kind of patterns for exergy inflow, outflow and
501 loss were reported by Chowdhury *et al.* (2011) and Akpinar (2011). Exergy takes into
502 account of the availability for expressing the performance of solar drying systems.



503
504

505 **Fig. 14** Exergy Efficiency and Exergy loss values for the mass flow rate of 0.0636 kg/s

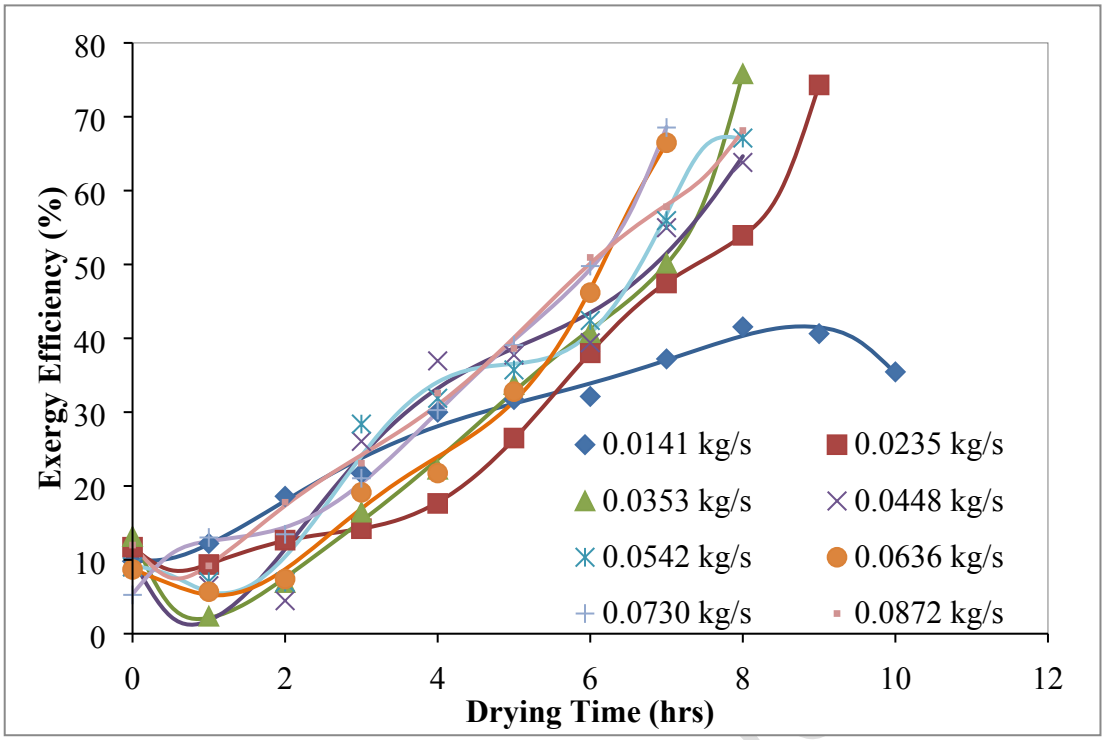


Fig. 15 Influence of mass flow rate of air on exergy efficiency

The effect of mass flow rate of air on the exergetic performance of the system for drying bitter gourd slices is shown in Fig.15 It indicates that, the exergy efficiency for all the mass flow rates increase with increase drying time, due to the reason of reduction in exergy destruction during the drying process. It is due to the reason that the difference in air temperature of drying chamber inlet and outlet is reduced, which is clearly seen in Fig. 11. At the end of the drying process the drying chamber inlet temperature is slightly lesser than the outlet temperature. The variations of average exergy loss and exergy efficiency values with the mass flow rates are shown in Fig. 16. The exergy efficiency increases and the exergy loss decreases with increase in mass flow rate of air as reported by Akbulut & Durmus (2010). The average exergy loss values are 9.185, 8.315, 8.094, 7.891, 7.548, 7.475, 6.825 and 6.333 W. The average exergy efficiency values vary from 28.74% to 40.679% for the mass flow rates 0.0141 kg/s to 0.0872 kg/s respectively. Exergy analysis should be used for evaluating the solar dryer and can also be used for designing an efficient with highest possible thermodynamic efficiencies.

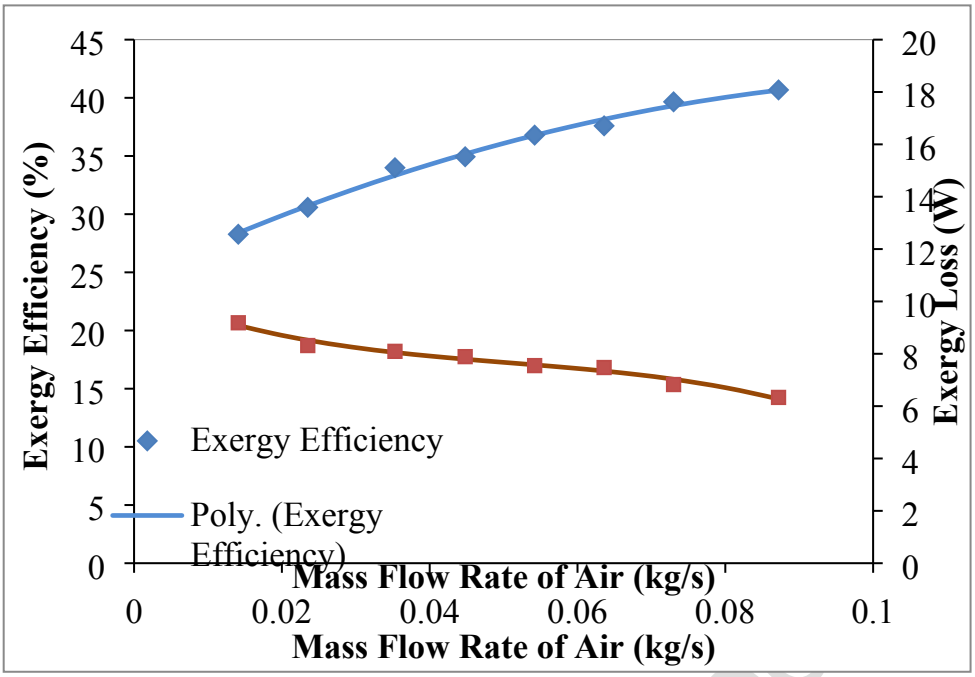


Fig. 16 Exergy Efficiency and Exergy loss with mass flow rate of air

Exergy Sustainability Indicators were calculated using the Equations (12), (13) and (14) for the drying of bitter gourd slices. The estimated values are indicated in Table 4 for various mass flow rates of air. Improvement Potential (IP) values for drying of bitter gourd slices for various mass flow rates of air were found in the range of 6.59 to 3.76W. It decreases with increase in mass flow rates of air, it means the possibility of the improving the drying process is reduced. Average sustainability index (SI) was expressed in terms of exergy efficiency; it was observed in the range of 1.39 to 1.69. It increases with increase in mass flow rate of air because the exergy efficiency of bitter gourd drying increases.

Table 4 Exergy Sustainability Indicators for drying of bitter gourd slices

Mass flow rate of air (kg/s)	Improvement Potential (IP)	Waste Exergy Ratio (WER)	Sustainability Index (SI)
0.0141	6.587805	0.813871	1.39
0.0235	5.771627	0.759024	1.44
0.0353	5.342315	0.733494	1.51
0.0448	5.134203	0.664411	1.54
0.0542	4.77106	0.652638	1.58
0.0636	4.574728	0.657867	1.63
0.0730	4.118659	0.643222	1.66

0.0872	3.756999	0.631227	1.69

536

537 The results of the present study can be compared with the indirect solar dryer
538 developed by Atalay (2019) for drying orange slices. The sustainability index was varied
539 from 2.15 to 3.17 and the exergy efficiency was also varied from 59.4 to 63.34 %.

540 **4.3. Environmental Analysis**

541 **Embodied Energy**

542 The embodied energy of the different materials used for the fabrication of the indirect
543 forced convection solar dryer is listed in Table 5.

544 **Table 5** Embodied energy for the fabrication of indirect solar dryer

S. No.	Component in Solar dryer	Material	Embodied energy (kWh/Kg) (Shrivastava & Kumar, 2017)	Quantity (Kg)	Total Embodied Energy (kWh)
1.	Glass Cover	Glass	7.28	4.8	34.944
2.	Absorber plate	Galvanized Iron sheet (0.45 mm thick)	9.636	3.02	29.101
3.	Outer covers	Galvanized Iron sheet (1.2 mm thick)	9.636	14.6	140.686
4.	Insulation	Glass wool	4.044	6.45	26.084
5.	Piping	Steel	8.89	4.32	38.405
6.	Frames	Mild Steel	8.89	41.5	368.935
7.	Drying trays	Aluminium mesh	55.28	4.54	250.971
8.	Fittings (nuts, bolts, screw, and rivets)	Steel	8.89	1	8.890
9.	Chimney	PVC pipe	19.39	1.5	29.085
10.	Coatings	Paint	25.11	2	50.220
11.	Packed bed	Pebbles	0.0278	110	3.056
12	Blower				
	i) Copper wire	Copper	19.61	0.5	9.805
	ii) Casings, fan, shaft etc	Steel	8.89	13.4	119.126
Total Embodied Energy (kWh)					**1109.307**

545

Journal Pre-proof

The indirect forced convection solar dryer is made up of different materials and the entire mass of the system is 207.63 kg. The mass percentages of the materials are shown in Fig.17. The packed bed energy material (gravels) occupies the major portion of 53% (110 kg) next to that piping and supporting structure takes 22% (45.82 kg) and the remaining mass percentages are shared by the other materials such as glass cover, insulation material, absorber plate etc.

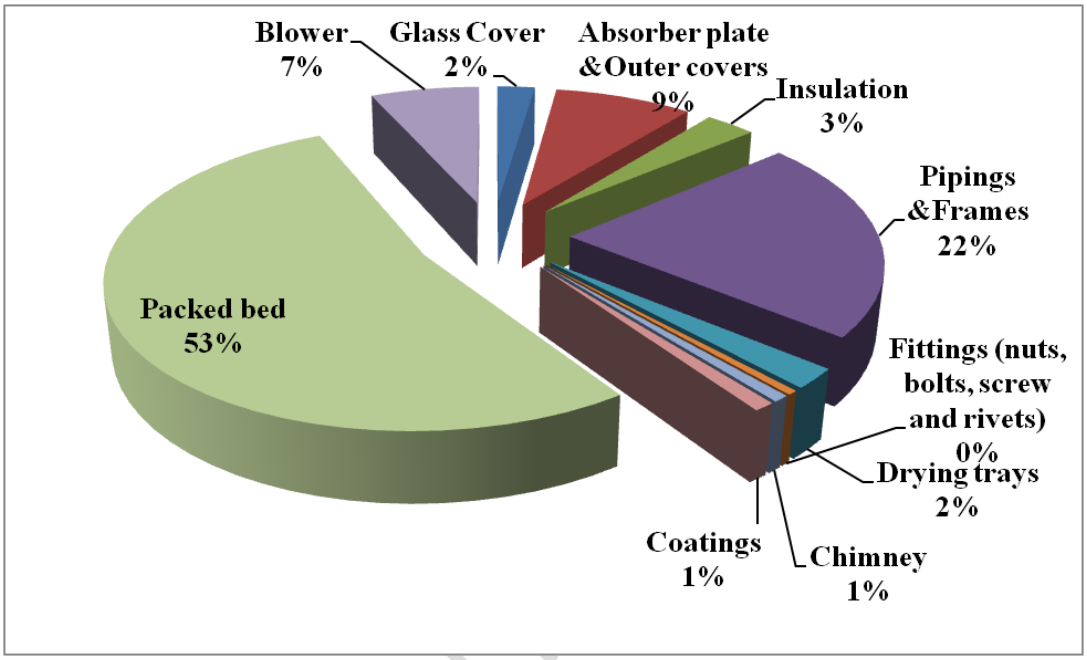


Fig. 17 Break-up of mass of different materials used for indirect solar dryer

The detailed breakup of the embodied energy shares of different materials used for the fabrication of indirect forced convection solar dryer is shown in Fig. 18. The total embodied energy for the developed system is 1109.307 kWh. The piping and supporting structures are made up of mild steel material; it shares about 37% of total embodied energy. The second major embodied energy is about 23% contributed by the drying trays because they were made up of aluminium material. Moreover, aluminimum material consumes relatively higher embodied energy than the other construction materials used. The embodied energy for packed bed is approximately nil because it does not require energy to produce. The pebble materials are naturally produced and easily available at low cost.

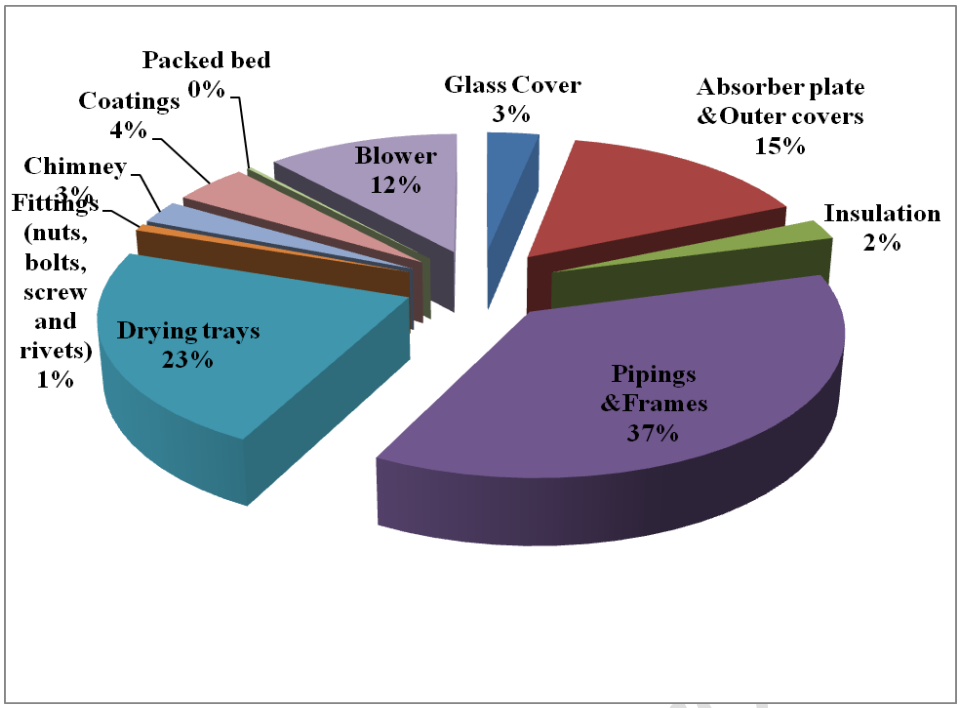


Fig. 18 Break-up of embodied energy of different materials used for indirect solar dryer

The indirect forced convection solar dryer evaporates 3.32 kg of moisture in 7 hours at the mass flow rate of 0.0636 kg/s air and the dryer is available for operation from morning 09:00 to 18:00 hr i.e 9 hours per day. However, this time of operation may vary depending upon the season, on an average the evaporative capacity of the dryer is taken as 3.5 kg per day. The latent heat of vaporization of water is taken as 2060 kJ/kg. The energy payback time for the indirect solar dryer is found to be 2.21 years. This period is much lower when compared to the lifetime of the developed solar dryer (35 years). The quantity of CO_2 emission per year was estimated using the Equation (22) for the life span 5 to 35 years and the results are shown in Table 6.

Table 6 CO_2 emission per year for different life span of indirect solar dryer

Life of solar Dryer (years)	5	15	25	35
CO_2 emission (Kg/year)	217.43652	72.47884	43.487304	31.06236

The calculated CO_2 mitigation of indirect solar dryer for different life span is given in Table 7.

584 **Table 7** CO_2 mitigation (Tons) for different life span of indirect solar dryer

Life of solar Dryer (years)	5	15	25	35
CO_2 mitigation (Tons)	2.847	13.072	23.296	33.520

585
586

587 If CO_2 emission is being traded from 5 to 20 US$ per tones (Tiwari & Tiwari, 2016)
588 of CO_2 mitigation, then the carbon credit earned by the system for the life span 5, 15, 25 and
589 35 are given in Table 8.

590 **Table 8** Carbon credit earned for different life span of indirect solar dryer

Life of solar Dryer (years)	5	15	25	35
Carbon credit earned (INR)	925.43 to 3701.71	4248.26 to 16993.02	7571.09 to 30284.34	10893.91 to 43575.66

591
592 **Table 9** Environmental impact results comparison for bitter gourd slices drying
593 (Assumed life span 35 years)
594

S.No.	Type of solar dryer	Embodied Energy kWh	Energy Payback Time (EPBT) in years	CO_2 emission kg/year	Co_2 Mitigation In tons	Earned carbon credit In INR	Reference
1	Indirect solar dryer	1109.307	2.21	31.06	33.52	10894 to 43576	Present study
2	Greenhouse dryer –passive mode	554.65	1.68	15.53	33.04	11068 to 44273	Chauhan et al. 2018 (b)
3	Greenhouse dryer - active mode	750.41	2.35	21.01	36.34	12173 to 48675	

595

596 Environmental impact analysis for solar drying process have been carried out for various
597 types of solar dryers such as indirect solar dryer, greenhouse dryer, modified greenhouse
598 dryer etc. Comparison analysis of energy payback and the CO_2 emission are reported by
599 Chauhan at al. (2018b) and highlighted that the modified greenhouse solar dryer is having
600 very low impact on environment. However, the energy payback period estimation depends on
601 the energy output but which varies product to product. Therefore, the results of
602 environmental impact analysis of this present study are compared with the greenhouse dryers
603 for drying bitter gourd slices in Table 9. Comparison analysis indicates that the embodied
604 energy is higher than the greenhouse solar dryers due to the reason of long-lasting structures
605 of the system whereas the greenhouse solar dryer polycarbonate sheets have to be replaced

periodically. The energy payback period is less than the greenhouse solar dryer (active mode).

5. CONCLUSION

Exergy and environmental impact analysis of the indirect forced convection solar dryer integrated with sensible heat storage medium for drying of bitter gourd slices are focused on this study. The mass of the bitter gourd slices is reduced from 4000 to 723 grams in 7 hours at the optimum mass flow rate of air 0.0636kg/s. i.e the initial moisture content is reduced from 92% (wet basis) to 9% (wet basis) by removing the moisture content of 3.277kg from the product. The drying rate is observed maximum at the initial stage of drying due to the fact of quick removal of surface moisture. The maximum drying rate noted is of 2.872kg of water/kg of dry mass for the mass flow rate of air 0.0636 kg/s. The effective moisture diffusivity of bitter gourd slices in solar drying is varied from 8.6293×10^{-10} to 12.9585×10^{-10} m^2/s and the lowest value of 0.9568×10^{-10} m^2/s is observed for open sun drying. The effect of mass flow rate of air on the pickup and exergy efficiencies of the drying process is studied and results indicated that the increase in mass flow rate of air leads to increase in pickup as well as exergy efficiencies of the drying process. The values of average pickup and exergy efficiencies are varied from 17.12% to 54.29% and from 28.27% to 40.68% for the mass flow rates from 0.0141 kg/s to 0.0872 kg/s, respectively. Exergy sustainability indicators such as improvement potential, waste energy ratio and sustainability index values for drying of bitter gourd slices in indirect solar dryer are in the range of 6.58 W to 3.76 W, 0.814 to 0.631 and 1.39 to 1.69 respectively for various mass flow rates of air. Environmental impact analysis indicated that the energy payback time for the developed system is 2.21 years only, which is very less while comparing with the expected life time of 35 years. The carbon dioxide is mitigation is estimated as 33.52 tons over life span. Through the reduction of CO_2 emission, the earned carbon credit values for the developed system is I the range of INR 10894 to 43576. The results of environmental impact assessment indicate that the solar drying systems are the best alternative solutions to protect the environment through the utilization of freely available solar energy source as compared with the other types of drying methods like hot air drying, infrared drying and freeze drying, as they operate with high grade energy (electrical). Moreover, the drying inside the indirect dryer is more consistent and produced a better quality dried product when compared with open sun-dried. The results indicated that the solar dryer is economically viable for small-scale farmers and can be used as women empowerment option in rural areas.

640

641 **Nomenclature**

642

C_p	-	specific heat of air (J/kg °C)
DR	-	drying rate (gm of water/gm. hr)
D_{eff}	-	Effective moisture diffusivity (m^2/s)
E_{aout}	-	*Annual thermal Energy output*
Ex		Exergy (J/kg)
E_{in}	-	Total Embodied Energy
I	-	solar insolation (W/m^2)
L	-	Life time of the system (yrs)
h_{as}	-	adiabatic saturation humidity of air entering the chamber (kg water/kg dry air)
h_i	-	absolute humidity of air entering the chamber (kg water/kg dry air)
h_o	-	absolute humidity of air leaving the chamber (kg water/kg dry air)
m_a	-	mass flow rate of air (kg/s)
M_e	-	equilibrium moisture content on wet basis (%)
M_i	-	initial moisture content on wet basis (%)
M_t	-	moisture content at any time of drying on wet basis (%)
MR	-	moisture ratio
N_d	-	number of sunshine days in a year
R	-	Half thickness of bitter slices (m)
T_a	-	Ambient temperature (°C)
T_{dci}	-	Drying chamber inlet temperature (°C)
T_{dco}	-	Drying chamber outlet temperature (°C)
W_d	-	weight of dry matter (kg)
W_t	-	weight of sample at any time t (kg)
W	-	uncertainty value in a variable
Greek		
$\eta_{pick\ up}$	-	Pickup efficiency (%)
η_{Ex}	-	Exergy Efficiency (%)
Subscripts		
Dci	-	Drying chamber inlet
Dco	-	Dying chamber outlet
Loss	-	losses

643

644 **Acknowledgement**

Journal Pre-proof

Authors are highly thankful to Solar Energy Research Centre, Coimbatore Institute of Engineering and Technology, Coimbatore (India) and Department of Mechanical Engineering, Delhi Technological University, Delhi (India) for providing basic facility to successfully complete this work.

References

1. Akbulut, A., & Durmuş, A. (2010). Energy and exergy analyses of thin layer drying of mulberry in a forced solar dryer. *Energy*, *35*(4), 1754-1763.
2. Akpinar, E. K. (2006). Mathematical modelling of thin layer drying process under open sun of some aromatic plants. *Journal of Food Engineering*, 77(4), 864–70.
3. Akpinar, E. K. (2010), Drying of mint leaves in a solar dryer and under open sun: Modelling, performance analyses., Energy Conversion and Management 51 (2010) 2407–2418
4. Akpinar, E. K. (2011). Drying of parsley leaves in a solar dryer and under open sun: modeling, energy and exergy aspects. *Journal of food process engineering*, *34*(1), 27-48.
5. Akpinar, E. K., & Koçyiğit, F. (2010). Energy and exergy analysis of a new flat-plate solar air heater having different obstacles on absorber plates. *Applied energy*, *87*(11), 3438-3450.
6. Alwan, Z & Jones, P 2014, 'The importance of embodied energy in carbon footprint assessment', Structural Survey, Vol. 32, pp. 49-60.
7. Atalay, H. (2019). Performance analysis of a solar dryer integrated with the packed bed thermal energy storage (TES) system. *Energy*, *172*, 1037-1052.
8. Aviara, N. A., Onuoha, L. N., Falola, O. E., & Igbeka, J. C. (2014). Energy and exergy analyses of native cassava starch drying in a tray dryer. *Energy*, *73*, 809-817.
9. Caliskan, H. (2017). Energy, exergy, environmental, enviroeconomic, exergoenvironmental (EXEN) and exergoenviroeconomic (EXENEC) analyses of solar collectors. *Renewable and Sustainable Energy Reviews*, *69*, 488-492.
10. Chowdhury, M. M. I., Bala, B. K., & Haque, M. A. (2011). Energy and exergy analysis of the solar drying of jackfruit leather. *Biosystems Engineering*, *110*(2), 222-229.
11. Chauhan, P. S., Kumar, A., Nuntadusit, C., & Mishra, S. S. (2018a). Drying Kinetics, Quality Assessment, and Economic Analysis of Bitter Gourd Flakes Drying Inside Forced Convection Greenhouse Dryer. *Journal of Solar Energy Engineering*, *140*(5), 051001.
12. Chauhan, P. S., Kumar, A., & Nuntadusit, C. (2018b). Thermo-environomical and drying kinetics of bitter gourd flakes drying under north wall insulated greenhouse dryer. *Solar Energy*, *162*, 205-216.

13. Chauhan, P. S., Kumar, A., Nuntadusit, C., & Banout, J. (2018c). Thermal modeling and drying kinetics of bitter gourd flakes drying in modified greenhouse dryer. *Renewable energy*, *118*, 799-813.

14. Dincer, I. (2000). Thermodynamics, exergy and environmental impact. *Energy sources*, *22*(8), 723-732.

15. Dincer, I. (2002). The role of exergy in energy policy making. *Energy policy*, *30*(2), 137-149.

16. Dincer, I., & Sahin, A. Z. (2004). A new model for thermodynamic analysis of a drying process. *International Journal of Heat and Mass Transfer*, *47*(4), 645-652.

17. Doymaz, I. (2007). Air-drying characteristics of tomatoes. *Journal of food engineering*, *78*(4), 1291-1297.

18. Fudholi, A., Sopian, K., Othman, M. Y., & Ruslan, M. H. (2014). Energy and exergy analyses of solar drying system of red seaweed. *Energy and Buildings*, *68*, 121-129.

19. Joseph, B., & Jini, D. (2013). Antidiabetic effects of Momordica charantia (bitter melon) and its medicinal potency.*Asian Pacific Journal of Tropical Disease*, *3*(2), 93-102.

20. Karthikeyan, A. K., & Murugavelh, S. (2018). Thin layer drying kinetics and exergy analysis of turmeric (Curcuma longa) in a mixed mode forced convection solar tunnel dryer. *Renewable energy*, *128*, 305-312.

21. Kesavan, S., Arjunan, T. V., & Vijayan, S. (2019). Thermodynamic analysis of a triple-pass solar dryer for drying potato slices. *Journal of Thermal Analysis and Calorimetry*, 136(1), 159-171.

22. Lopez, A, Iguaz, A, Esnoz, A & Virseda, P 2000, 'Thin-layer drying behaviour of vegetable wastes from wholesale market', Drying Technology, Vol.18 no. 4-5, pp. 995-1006.

23. Lucas, E. A., Dumancas, G. G., Smith, B. J., Clarke, S. L., & Arjmandi, B. H. (2010). Health benefits of bitter melon (Momordica charantia). In *Bioactive Foods in Promoting Health* (pp. 525-549). Academic Press.

24. Mohan, V. C., & Talukdar, P. (2013). Design of an experimental set up for convective drying: experimental studies at different drying temperature. *Heat and Mass Transfer*, *49*(1), 31-40.

25. Prakash, O., & Kumar, A. (2013). Historical review and recent trends in solar drying systems. *International Journal of Green Energy*, *10*(7), 690-738.

26. Prakash, O., Kumar, A., & Sharaf-Eldeen, Y. I. (2016). Review on Indian solar drying status. *Current Sustainable/Renewable Energy Reports*, *3*(3-4), 113-120.

27. Rabha, D. K., Muthukumar, P., & Somayaji, C. (2017). Energy and exergy analyses of the solar drying processes of ghost chilli pepper and ginger. *Renewable Energy*, *105*, 764-773.

28. Sacilik, K., Keskin, R., & Elicin, A. K. (2006). Mathematical modelling of solar tunnel drying of thin layer organic tomato. *Journal of food Engineering*, *73*(3), 231-238.

29. Sami, S., Etesami, N., & Rahimi, A. (2011). Energy and exergy analysis of an indirect solar cabinet dryer based on mathematical modeling results. *Energy*, *36*(5), 2847-2855.

30. Shrivastava, V., & Kumar, A. (2016). Experimental investigation on the comparison of fenugreek drying in an indirect solar dryer and under open sun. *Heat and Mass Transfer*, *52*(9), 1963-1972.

31. Shrivastava, V., & Kumar, A. (2017). Embodied energy analysis of the indirect solar drying unit. *International Journal of Ambient Energy*, *38*(3), 280-285.

32. Singh, P., Shrivastava, V., & Kumar, A. (2018). Recent developments in greenhouse solar drying: A review. *Renewable and Sustainable Energy Reviews*, *82*, 3250-3262.

33. Sreekumar, A., Manikantan, P. E., & Vijayakumar, K. P. (2008). Performance of indirect solar cabinet dryer. *Energy Conversion and Management*, *49*(6), 1388-1395.

34. Tiwari, S & Tiwari, G.N. (2016). 'Exergoeconomic analysis of photovoltaic-thermal (PVT) mixed mode greenhouse solar dryer', Energy, Vol. 114, no. 1, pp. 155-164.

35. Vijayan, S., Arjunan, T. V., & Kumar, A. (2016). Mathematical modeling and performance analysis of thin layer drying of bitter gourd in sensible storage based indirect solar dryer. *Innovative food science & emerging technologies*, *36*, 59-67.

36. Vijayan, S., Arjunan, T. V., & Kumar, A. (2017). Fundamental concepts of drying. In *Solar Drying Technology* (pp. 3-38). Springer, Singapore.

37. Vijayan, S., Vellingiri, A. T., & Kumar, A. (2017). Thin layer drying characteristics of curry leaves (Murraya koenigii) in an indirect solar dryer. *Therm Sci*, *21*(2), S359-67.

38. Wang, J, Xiong, YS & Yu, Y 2004, 'Microwave drying characteristics of potato and the effect of different microwave powers on the dried quality of potato', European Food Research and Technology, Vol. 219, no. 5, pp. 500-506.

39. Yan, J. K., Wu, L. X., Qiao, Z. R., Cai, W. D., & Ma, H. (2019). Effect of different drying methods on the product quality and bioactive polysaccharides of bitter gourd (Momordica charantia L.) slices. *Food chemistry*, *271*, 588-596.

Journal Pre-proof

Highlights

- Effect of mass flow rate of air on exergy and pickup efficiencies were studied
- The average pickup efficiency of the dryer varies from 17.12 to 54.29%
- Exergy efficiency varies from 28.27 to 40.68% for 0.0141 to 0.0872 kg/s
- D_{eff} of bitter gourd slices vary from 8.6293×10^{-10} to 12.9585×10^{-10} m^2/s
- CO_2 mitigation & earned carbon credit values are 33.52 tons and INR 10894 to 43576

Journal Pre-proof

Neural Computing and Applications
https://doi.org/10.1007/s00521-019-04433-0

ORIGINAL ARTICLE

Formulation and application of quantum-inspired tidal firefly technique for multiple-objective mixed cost-effective emission dispatch

Kapil Deo Bodha[1] · Vinod Kumar Yadav[2] · Vivekananda Mukherjee[1]

Received: 30 April 2018 / Accepted: 7 August 2019
© Springer-Verlag London Ltd., part of Springer Nature 2019

Abstract
In this manuscript, a new quantum computing-based optimization algorithm is proposed to solve multiple-objective mixed cost-effective emission dispatch (MEED) problem of electrical power system. The MEED problem aims at maintaining proper balance between emission of pollutants and generation of power. The problem has been formulated here using cubic equation to reduce the nonlinearities of the system. It is transformed to single-objective problem by considering max to max penalty factor. The proposed optimization technique is inspired by the concept of quantum mechanics, gravitational force and firefly algorithm (FA) and is termed as quantum-inspired tidal FA (QITFA). The proposed QITFA is tested on IEEE 14-bus and IEEE 30-bus test system for four different load conditions. The obtained results are compared with the results yielded by some other state-of-the-art methods like Lagrangian relaxation method, particle swarm optimization (PSO), simulated annealing, quantum-behaved bat algorithm and quantum PSO. This paper proves the superiority of the proposed QITFA over all these methods. Further, the obtained results also suggest its effective and efficient implementation in MEED problem.

Keywords Economic power dispatch · Emission penalty cost · Firefly algorithm (FA) · Max to max penalty factor · Quantum-inspired tidal firefly algorithm (QITFA) · Tidal firefly algorithm (TFA)

Abbreviations

A^2MO	Artificial mountain ape optimization
ABC	Artificial bee colony
COOA	Competitive optimization algorithm
DBC	Dance bee colony
FA	Firefly algorithm
FMOPSO	Fuzzified many-objective PSO
FSA	Fractal search algorithm
GA	Genetic algorithm
GSA	Gravitational search algorithm
IEEE	Institute of Electrical and Electronics Engineers
LF-VPSO	Levy flight-based voltage PSO
LR	Lagrangian relaxation
MEED	Multiple-objective mixed cost-effective emission dispatch
MHSA	Modified harmony search algorithm
PF	Penalty factor
PSO	Particle swarm optimization
QBAT	Quantum-behaved bat algorithm
QITFA	Quantum-inspired tidal FA
QPSO	Quantum PSO
SA	Simulated annealing
SOA	Spiral optimization algorithm
TDSE	Time-dependent Schrodinger equation
TFA	Tidal FA
TLBA	Teacher-learning-based algorithm

List of symbols

CO_2	Carbon dioxide
d_{ij}	Gap between firefly i and j
D	Number of dimensions
f_i	Tidal component of the ith firefly
G	Exploration factor
G	Gravitational constant

✉ Vinod Kumar Yadav
vinodkumar@dtu.ac.in

[1] Department of Electrical Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, India

[2] Department of Electrical Engineering, Delhi Technological University, New Delhi, India

Published online: 19 August 2019

$h_{i(\max/\max)}$	Max/max penalty factor of the ith generating unit
$J_i(P_i)$	Emission function of the ith $g$enerator
$J_{i,\mathrm{SO_x}}(P_i)$	Emission function for SO_x
$J_{i,\mathrm{NO_x}}(P_i)$	Emission function for NO_x
$J_{i,\mathrm{CO_2}}(P_i)$	Emission function for CO_2
m_1	Mass of one particle
m_2	Mass of the other particle
n_g	Total numbers of generators till last generating unit
n_f	Number of fireflies
NO_x	Nitrogen oxide
P	Generated power
P_D	Power demand
P_L	Total losses
$P_{i,mu}$	Least power limit of the ith generator
$P_{i,mx}$	Extreme power limit of the ith generator
SO_x	Sulfur oxide
$x_{i,m}$	The position of the ith firefly in mth dimension
$x_{j,m}$	The position of the jth firefly in mth dimension
α	Damping factor
α_0'	Random constants in [0, 1]
β'	Random constants in [0, 1]
β	Attraction factor
β_0, γ	Random constants lying in between 0 and 1
$\alpha_{0,i}, \alpha_{1,i}, \alpha_{2,i}, \alpha_{3,i}$	Emission coefficient of the ith generator
$\omega_{0,i}, \omega_{1,i}, \omega_{2,i}, \omega_{3,i}$	Fuel cost coefficient of the ith generator
$\alpha_{0(\mathrm{SO_x}),i}, \alpha_{1(\mathrm{SO_x}),i}, \alpha_{2(\mathrm{SO_x}),i}, \alpha_{3(\mathrm{SO_x}),i}$	SO_x emission coefficient of the ith generator
$\alpha_{0(\mathrm{CO_2}),i}, \alpha_{1(\mathrm{CO_2}),i}, \alpha_{2(\mathrm{CO_2}),i}, \alpha_{3(\mathrm{CO_2}),i}$	CO_2 emission coefficient of the ith generator
$\alpha_{0(\mathrm{NO_x}),i}, \alpha_{1(\mathrm{NO_x}),i}, \alpha_{2(\mathrm{NO_x}),i}, \alpha_{3(\mathrm{NO_x}),i}$	NO_x emission coefficient of the ith generator
$\Psi(d,t)$	Position of the wave
$\hbar$	Planck's constant
u	Random variable ranging from 0 to 1
∇^2	Laplace operator
$t_{\max}$	Maximum number of iterations

1 Introduction

Temperature of the earth is rising continuously since the past three decades due to greenhouse effect [1]. The major greenhouse gases are carbon dioxide, methane, nitrous oxide and ozone. These gases are enormously produced by power plants [2]. A large percentage of electricity is generated by burning coal, oil and natural gases, contributing directly to the production of greenhouse gases. Increasing industrialization is one of the characteristics of evolving societies, which requires uninterrupted power supply and efficient transport system for its growth. Due to this reason, there is an upsurge in emission level of these hazardous pollutants. Therefore, in order to reduce this emission level, entire world is going through a renewable revolution [3]. However, it is not possible to generate all the required power (which is a very large quantity) from renewable sources alone; therefore, generations from burning coal and different sources are required. For the economical operation of power plant, dispatching is done which gives optimal operating point of the generators in the power plant. At this point, fuel consumption to generate the required power is the minimum. This is known as economic dispatch. Owing to the environmental concerns, the power plant also needs to limit the gaseous emission. The main polluting gases are SO_x, CO_2 and NO_x. The emission of SO_x can be reduced using flue gas desulfurization, and NO_x emission can be reduced by modifying the burner gas design while electrostatic precipitator can be used to reduce the particulate matter. Installation of these requires a huge capital investment; and also power plants need to be shut down for these installations which will cause a large amount of outage [3]. The multi-objective mixed economic emission dispatch (MEED) problem is a multi-objective dispatching problem that aims to minimize the fuel cost and emission from power plant at the same time to generate the required power without violating any specified constraints.

In the literature, various techniques were adopted to optimize the MEED problem. Song et al. [4] presented an improved genetic algorithm (GA) based on two fuzzy controllers to solve the MEED problem. Linear as well as nonlinear goal encoding methods were adopted by Nanda et al. [5]. Venkatesh et al. [6] considered line flow constraints to solve MEED problem by assessment and solicitation of evolutionary programming methods, while Abido [7] used only evolutionary algorithms for the same. A fuzzified many-objective particle swarm technique (PSO) (FMOPSO) was proposed by Wang and Singh in [8] to resolve the emission and economic dispatch issue. Raglend et al. [9] juxtaposed all the artificial intelligent algorithms involved in MEED. In [10], the MEED problem is analyzed and optimized by using multi-objective ant swarm optimization method. Likewise, many such algorithms have been proposed in the literature such as gravitational search algorithm [11], improved ant colony optimization [12], teacher-learning-based algorithm [13], artificial bee colony algorithm [14], dance bee colony method [15], spiral optimization algorithm [16–18], PSO [19–21], competitive optimization algorithm (COOA) [22], quasi-

oppositional teaching-learning-based optimization [23, 24], artificial mountain ape optimization (A^2MO) [25], levy flight-based voltage PSO (LF-VPSO) [26], quantum-inspired PSO (QI-PSO) [27], modified harmony search algorithm (MHSA) [28], fractal search algorithm (FSA) [29] and so on. One shortcoming of all the above reported techniques is that it may trap in local minima/maxima which inspired the present authors to develop quantum-inspired tidal firefly algorithm (FA) (TFFA) (QITFA) to resolve this problem.

The MEED problem is formulated by using cubic-based equation/function as cubic function provides better solution capability than its quadratic counterpart [30]. Quadratic polynomial is a very popular method for formulation of MEED problem. However, the quadratic function does not accurately describe the nonlinearities of a power system. As the order of polynomial is increased, the solution becomes more optimized. Therefore, the cubic function is a step closer to the actual response of the power system. The increasing nonlinearities of a power system can be checked by representing MEED problem through cubic equations. On increasing the order of the polynomial, complexity of the system increases. To get a balance between the optimized solution and system complexities, cubic function is utilized in this manuscript. The use of cubic polynomial is now gaining popularity among the researchers and the industry houses. The multi-objective function is converted to a single-objective function using max to max penalty factor [31]. In this manuscript, QITFA is reported for the first time. The proposed algorithm is tested to solve the MEED problem of the power system. The tidal firefly is an enhanced version of FA that eliminates the need of absorption coefficient [32] and, hence, improves the convergence capability of the algorithm. The TFA is applied though the principle of quantum mechanics. The proposed algorithm depends on very few factors. Hence, it is computationally very dominant and converges much faster than the other existing algorithm.

The problem formulation is developed in the next section which includes formation of emission and fuel cost functions and its conversion into a single-objective function using penalty factors (PFs). FA and the TFA methods are discussed in Sect. 3. The proposed QITFA technique is discussed in detail in Sect. 4. In Sect. 5, the proposed method is applied on IEEE 14-bus and 30-bus system with different loads and is compared with GA, Lagrangian relaxation (LR), PSO, simulated annealing (SA), quantum-behaved bat algorithm (QBA) and quantum PSO (QPSO) methods to prove its effectiveness. Finally, the paper is concluded in Sect. 6 indicating scope of future work.

2 Problem formulation

2.1 Objective function

The objectives of the present problem are to minimize the (a) fuel cost, (b) emissions and (c) total cost that includes the cost of fuel and emission penalty cost [33]. These objective functions are governed by certain constraints. The objective is represented by (1)

$$\text{Minimize } F(X) = [G(P), J(P)] \tag{1}$$

where the total cost of burning fuel to generate the power P is denoted by $G(P)$ and total emission of unwanted and harmful pollutants after burning fuel is denoted by $J(P)$.

2.2 Control variable

In (1), the control variable is X. It is a set of unknown variables, and in this case, it is a set of powers of various generating units as shown in (2):

$$X = \left[P_1, P_2, P_3, \ldots, P_{n_g}\right] \tag{2}$$

where power generation from first, second and third units is denoted by P_1, P_2, P_3, respectively. P_{n_g} is the power generation from the last generating unit (n_g).

2.3 Cost-effective dispatch

The main aspect of economic or cost-effective dispatch of generated power is to minimize $G(X)$ which is the total cost incurred due to burning of fuel to generate required demanded power. It is shown in (3) in the form of cubic function:

$$\begin{aligned}\text{Minimize } G(X) &= \sum_{i=1}^{n_g} G_i(P_i) \\ &= \sum_{i=1}^{n_g} \left(\omega_{0,i} + \omega_{1,i}P_i + \omega_{2,i}P_i^2 + \omega_{3,i}P_i^3\right)\end{aligned} \tag{3}$$

where the fuel cost function of the ith generator is denoted by $G_i(P_i)$ and is calculated in \$/h. The fuel cost coefficients of the ith generator are $\omega_{0,i}, \omega_{1,i}, \omega_{2,i}$ and $\omega_{3,i}$, respectively.

2.4 Emission dispatch

The amount of pollutant is determined by the burning of fuel which is required to generate power. The total emission of pollutants is denoted by $J(X)$. It is shown in (4) in the form of cubic function:

$$\text{Minimize } J(X) = \sum_{i=1}^{n_g} J_i(P_i) = \sum_{i=1}^{n_g} \left(\alpha_{0,i} + \alpha_{1,i}P_i + \alpha_{2,i}P_i^2 + \alpha_{3,i}P_i^3\right) \quad (4)$$

where the emission function of the ith generator is denoted by $J_i(P_i)$ and is calculated in kg/h. The emission coefficients of the ith generator are, in order, $\alpha_{0,i}, \alpha_{1,i}, \alpha_{2,i}$ and $\alpha_{3,i}$.

2.5 Power balance constraints

The summation of total demanded power and the total losses (that take place in order to supply the generated power) must be supplied by the total available generators as shown in (5):

$$\sum_{i=1}^{n_g} P_i = P_D + P_L \quad (5)$$

where the total demanded power is denoted by P_D and the total losses are denoted by P_L.

2.6 Power output constraints

Each generating unit is bounded by its least and extreme generation power edge as shown in (6):

$$P_{i,mu} \le P_i \le P_{i,mx} \quad (6)$$

where the least and extreme power limit of the ith generator are denoted by $P_{i,mu}$ and $P_{i,mx}$, respectively.

2.7 Transformation from multi-objective to single-objective function

The transformation of multi-objective into single-objective is done after taking PF $\left(h_{\max/\max}\right)$ [34] into consideration (as shown in (7)). The extended equation is shown in (8). The use of PF reduces the complexity of the equation and makes it easier to solve [35]. The value of $h_{i(\max/\max)}$ is max/max PF of the ith generating unit and is calculated in \$/kg as shown in (9).

$$\text{Minimize } F'(X) = \left(G(P) + h_{\max/\max} * J(P)\right) = \sum_{i=1}^{n_g} G_i(P_i) + h_{i(\max/\max)} \sum_{i=1}^{n_g} J_i(P_i) \quad (7)$$

$$= \sum_{I=1}^{n_g} G_i(P_i) + h_{i\left(\frac{\max}{\max}\right),\mathrm{SO_x}} J_{i,\mathrm{SO_x}}(P_i) + h_{i\left(\frac{\max}{\max}\right),\mathrm{NO_x}} J_{i,\mathrm{NO_x}}(P_i) + h_{i\left(\frac{\max}{\max}\right),\mathrm{CO_2}} J_{i,\mathrm{CO_2}}(P_i) \quad (8)$$

where

$$h_{i\left(\frac{\max}{\max}\right)} = \sum_{i=1}^{n_g} \frac{G_i\left(P_{i,mx}\right)}{J_i\left(P_{i,mx}\right)} \quad (9)$$

$J_{i,\mathrm{SO_x}}(P_i)$, $J_{i,\mathrm{NO_x}}(P_i)$ and $J_{i,\mathrm{CO_2}}(P_i)$ denote the emission of SO_x, NO_x and CO_2, respectively. The cubic equation governing these pollutants emission is shown in (10)–(12).

$$J_{i,\mathrm{SO_x}}(P_i) = \alpha_{0(\mathrm{SO_x}),i} + \alpha_{1(\mathrm{SO_x}),i}P_i + \alpha_{2(\mathrm{SO_x}),i}P_i^2 + \alpha_{3(\mathrm{SO_x}),i}P_i^3 \quad (10)$$

where the SO_x emission coefficient of the ith generator is denoted by $\alpha_{0(\mathrm{SO_x}),i}, \alpha_{1(\mathrm{SO_x}),i}, \alpha_{2(\mathrm{SO_x}),i}$ and $\alpha_{3(\mathrm{SO_x}),i}$.

$$J_{i,\mathrm{NO_x}}(P_i) = \alpha_{0(\mathrm{NO_x}),i} + \alpha_{1(\mathrm{NO_x}),i}P_i + \alpha_{2(\mathrm{NO_x}),i}P_i^2 + \alpha_{3(\mathrm{NO_x}),i}P_i^3 \quad (11)$$

where the NO_x emission coefficient of the ith generator is denoted by $\alpha_{0(\mathrm{NO_x}),i}, \alpha_{1(\mathrm{NO_x}),i}, \alpha_{2(\mathrm{NO_x}),i}$ and $\alpha_{3(\mathrm{NO_x}),i}$.

$$J_{i,\mathrm{CO_2}}(P_i) = \alpha_{0(\mathrm{CO_2}),i} + \alpha_{1(\mathrm{CO_2}),i}P_i + \alpha_{2(\mathrm{CO_2}),i}P_i^2 + \alpha_{3(\mathrm{CO_2}),i}P_i^3 \quad (12)$$

where the CO_x emission coefficient of the ith generator is denoted by $\alpha_{0(\mathrm{CO_x}),i}, \alpha_{1(\mathrm{CO_x}),i}, \alpha_{2(\mathrm{CO_x}),i}$ and $\alpha_{3(\mathrm{CO_x}),i}$.

The related terms involved are defined in (13)–(18).

$$h_{i\left(\frac{\max}{\max}\right),\mathrm{SO_x}} = \sum_{i=1}^{n_g} \frac{G_i\left(P_{i,mx}\right)}{J_{i,\mathrm{SO_x}}\left(P_{i,mx}\right)} \quad (13)$$

$$h_{i\left(\frac{\max}{\max}\right),\mathrm{SO_x}} = \sum_{i=1}^{n_g} \frac{\omega_{0,i} + \omega_{1,i}P_{i,mx} + \omega_{2,i}P_{i,mx}^2 + \omega_{3,i}P_{i,mx}^3}{\alpha_{0(\mathrm{SO_x}),i} + \alpha_{1(\mathrm{SO_x}),i}P_{i,mx} + \alpha_{2(\mathrm{SO_x}),i}P_{i,mx}^2 + \alpha_{3(\mathrm{SO_x}),i}P_{i,mx}^3} \quad (14)$$

$$h_{i\left(\frac{\max}{\max}\right),\mathrm{NO_x}} = \sum_{i=1}^{n_g} \frac{G_i\left(P_{i,mx}\right)}{J_{i,\mathrm{NO_x}}\left(P_{i,mx}\right)} \quad (15)$$

$$h_{i\left(\frac{\max}{\max}\right),\mathrm{NO_x}} = \sum_{i=1}^{n_g} \frac{\omega_{0,i} + \omega_{1,i}P_{i,mx} + \omega_{2,i}P_{i,mx}^2 + \omega_{3,i}P_{i,mx}^3}{\alpha_{0(\mathrm{NO_x}),i} + \alpha_{1(\mathrm{NO_x}),i}P_{i,mx} + \alpha_{2(\mathrm{NO_x}),i}P_{i,mx}^2 + \alpha_{3(\mathrm{NO_x}),i}P_{i,mx}^3} \quad (16)$$

$$h_{i\left(\frac{\max}{\max}\right),\mathrm{CO_2}} = \sum_{i=1}^{n_g} \frac{G_i\left(P_{i,mx}\right)}{J_{i,\mathrm{CO_2}}\left(P_{i,mx}\right)} \quad (17)$$

$$h_{i\left(\frac{\max}{\max}\right),\mathrm{CO_2}} = \sum_{i=1}^{n_g} \frac{\omega_{0,i} + \omega_{1,i}P_{i,mx} + \omega_{2,i}P_{i,mx}^2 + \omega_{3,i}P_{i,mx}^3}{\alpha_{0(\mathrm{CO_2}),i} + \alpha_{1(\mathrm{CO_2}),i}P_{i,mx} + \alpha_{2(\mathrm{CO_2}),i}P_{i,mx}^2 + \alpha_{3(\mathrm{CO_2}),i}P_{i,mx}^3} \quad (18)$$

The max to max penalty factor of the ith generator for SO_x, NO_x and CO_2 pollutants can be calculated numerically by (13), (15) and (17), respectively, which can be expanded as in (14), (16) and (18), respectively. The value comes out to

be in \$/Kg. The symbols involved have usual meaning and have been explained earlier.

3 TFA technique

The FA was introduced by Yang [36]. It was encouraged by the blinking action of fireflies and is employed in optimization. There are three major guidelines for the algorithm. These are presented below:

(a) The fireflies are unisex, and they will be attracted to each other irrespective of their sex.
(b) Attraction factor is associated with brightness factor. Less bright firefly will be attracted toward the brighter firefly. As the gap between fireflies increases, attraction factor as well as brightness factor decreases and vice versa.
(c) Objective function decides the brightness factor of the firefly.

Mathematically, the distance between any two fireflies i and j at positions x_i and x_j in the mth dimension can be formulated as Euclidean distance [37] using (19):

$$d_{ij} = \sqrt{\sum_{m=1}^{D} \left(x_{i,m} - x_{j,m}\right)^2} \tag{19}$$

where $i = 1, 2, 3, \ldots j, \ldots p, \ldots n_f$ and $m = 1, 2, \ldots D$; n_f are the number of fireflies and D is the number of dimensions; $x_{i,m}$ and $x_{j,m}$ are the position of the ith and the jth firefly in the mth dimension, respectively. The attraction or brightness factor when a firefly x_i is attracted to brighter firefly x_j is given by (20).

$$\beta\left(d_{ij}\right) = \beta_0 \times e^{-\gamma d_{ij}^2} \quad (m \geq 1) \tag{20}$$

where $\beta\left(d_{ij}\right)$ is the attraction factor and β_0 and γ are the random constants whose value lies between 0 and 1. If $\gamma = 0$ and upon selecting appropriate parameters, Eq. (21) will be converted into standard PSO algorithm. The equation dealing with the movement is governed by (21).

$$x_{i,m}^{t+\Delta t} = x_{i,m}^{t} + \beta\left(d_{ij}\right) * \left(x_{j,m}^{t} - x_{i,m}^{t}\right) + \alpha * \left(rand\left(-\frac{1}{2}\right)\right) \tag{21}$$

where $x_{i,m}^{t+1}$ and $x_{i,m}^{t}$ are the new and old positions of the ith firefly in the mth dimension, respectively, $x_{j,m}^{t}$ is the position of more brighter or the jth firefly in the mth dimension, α is the damping factor and lies in [0, 1] and *rand* is the random variable.

After the introduction of tidal force in place of attraction factor, the global minimum value of the objective function can be achieved within lesser number of generations. The TFA was proposed by Yelghi and Kose in [32]. The new modified equation dealing with the movement is given by (22).

$$x_{i,m}^{t+\Delta t} = x_{i,m}^{t} + f_i * \left(x_{j,m}^{t} - x_{i,m}^{t}\right) + \alpha * \left(rand\left(-\frac{1}{2}\right)\right) \tag{22}$$

where f_i is the tidal component of the ith firefly and is given by (23).

$$f_i = \alpha_0' e^{-\beta' \times \Delta f_{ij}} \tag{23}$$

where Δf_{ij} is the difference of near $(f_{near}^{x_i^t})$ to far $(f_{far}^{x_j^t})$ tidal force. Mathematically, it is described below:

$$\Delta f_{ij} = f_{near}^{x_i^t} - f_{far}^{x_j^t} \tag{23 - a}$$

where

$$f_{near}^{x_i^t} = \frac{G * m_1 * m_2}{(x_{i,m}^t) * (x_{i,m}^t)} \tag{23 - b}$$

$$f_{far}^{x_j^t} = \frac{G * m_1 * m_2}{(x_{j,m}^t) * (x_{j,m}^t)} \tag{23 - c}$$

where α_0' is the random constants in [0, 1], β' is the random constants in [0, 1], G is the gravitational constant, m_1 is the mass of one particle and m_2 is mass of the other particle.

4 Methodology

In the proposed QITFA, no particle is allowed to move under classical mechanics. It is assumed that every particle will follow the quantum mechanical rules, ignoring spin of the particles [38]. In the classical atmosphere, all the fireflies are attracted toward the brightest firefly (i.e., the jth firefly) through the optimization process. Hence, the global optimum is achieved. Now, in quantum mechanics, the basic equation dealing with the position and time is a time-dependent Schrodinger equation (TDSE) [39, 40] and is represented as shown in (24).

$$i\hbar \frac{\partial}{\partial t} \Psi(d,t) = \hat{H}(d) \Psi(d,t) \tag{24}$$

where $\Psi(d,t)$ is the position of the wave, $\hbar$ is the Planck's constant and the Hamiltonian operator $(\hat{H})$ is the sum of kinetic and potential energy of the system influenced by (25).

$$\hat{H}(d) = -\frac{\hbar^2}{2m} \nabla^2 + U(d) \tag{25}$$

where m is the mass of particle, $U(d)$ is the potential energy and ∇^2 is Laplacian in three-dimensional space and is defined in (26).

$$\nabla^2 = \frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} + \frac{\partial}{\partial z^2} \tag{26}$$

There is no physical significance of position $\Psi(d,t)$ of wave. But square of its amplitude gives the probability of finding the particle [see (27)].

$$\int\int\int_{-\infty}^{+\infty} |\Psi(d,t)|^2 \mathrm{d}x\mathrm{d}y\mathrm{d}z = 1 \tag{27}$$

In this proposed algorithm, square potential has been chosen to find the solution of TDSE. By applying Monte Carlo measurement method, the position of particles can be determined. The general equation governing the position is stated in (28) [41]:

$$x_{i,m}^{t+\Delta t} = x_{p,m}^{t} \pm L\cos^{-1}\left[\pm\sqrt{u}\right] \tag{28}$$

where $x_{p,m}^{t}$ is the best position of the brightest firefly in current time in the mth dimension and is determined by using tidal firefly movement equation. In (28), u is the random variable ranging from 0 to 1, L depends on factor g and its numerical value can be obtained from (29) and the value of g is calculated by (30) [41]:

$$L = \frac{1}{0.5 \times g}\left|x_{i,m}^{t} - x_{p,m}^{t}\right| \tag{29}$$

$$g = \frac{t_{\max}}{t_{\max} \times (e + 0.5) - e * t} \tag{30}$$

where the value of e decreases from 1 to 0.5, $t_{\max}$ is the maximum number of iterations and t is the current iteration. Local exploration is obtained with large value of g while global exploration is obtained with its small value but it makes the algorithm slow. Small value of g is considered and then tuned into its larger value. Hence, updated equation governing the movement of firefly involving quantum with tidal approach in FA is presented in (31).

$$x_{i,m}^{t+\Delta t} = x_{p,m}^{t} \pm \frac{1}{0.5 * g}\cos^{-1}\left[\pm\sqrt{u}\right]\left|x_{i,m}^{t} - x_{p,m}^{t}\right| \tag{31}$$

The flowchart of the proposed QITFA is illustrated in Fig. 1 while its underline concept is presented in Algorithm 1.

Algorithm 1: Proposed QITFA

Step 1 Initialize the basic parameters like maximum number of fireflies (n_f), number of dimensions (D) and maximum number of iterations (t_{max}) which is also a limiting factor to stop the algorithm. Also, particles are initially assigned random values in quantum space.

Step 2 Compute fitness function, i.e., objective function for each firefly. Also, calculate tidal factor intensity from (23) (mentioned subsequently in **Step 7**).

Step 3 Juxtapose tidal component of the ith firefly with the pth firefly (initially, assumed with high tidal component). If tidal component of the ith firefly is less than the pth firefly, then go to **Step 2**; otherwise, go to **Step 4**. The firefly with highest tidal component is the brightest firefly.

Step 4 Apply the formulas to compute L and g from (29) and (30). Update $x_{p,m}^{t}$, position of high tidal component firefly (say the pth one, that gives the best solution) in the mth dimension in the tth iteration or time.

Step 5 Update new position of firefly using (31).

Step 6 Store the result.

Step 7 Check for a maximum number of iterations. If the algorithm has reached it, **STOP**; otherwise, repeat from **Step 2** to **Step 5**.

5 Result and analysis

In this manuscript, QITFA has been analyzed on the basis of simulation results performed on IEEE 14-bus and IEEE 30-bus test system. For both the test cases, the line losses are not considered. The simulations are carried out using MATLAB R2015a on a HP pavilion × 360 Intel Core i3-6100 @ (2.3 GHz) (4 CPU) with 4 GB of RAM. The system configurations of the chosen electrical test systems are referred from the state-of-the-art literatures. The parameter settings for QITFA and other referred techniques are given in Table 1. However, parameter setting data for PSO and GA are not available in [42].

5.1 IEEE 14-bus

QITFA is applied on IEEE 14-bus test system having three generating units to solve MEED problem for a load of 2500 MW. For better comparative analysis, the QITFA is made to run for 500 iterations with 1500 fireflies and 6 dimensions and is compared with different existing methods. Table 2 shows the comparison of the obtained results of MEED from GA, PSO, QPSO and QITFA. It can be concluded from Table 2 that QITFA overtakes GA, PSO and QPSO in terms of total cost that is the combined fuel and emission penalty cost. Figure 2 shows the convergence characteristics of total cost for this system.

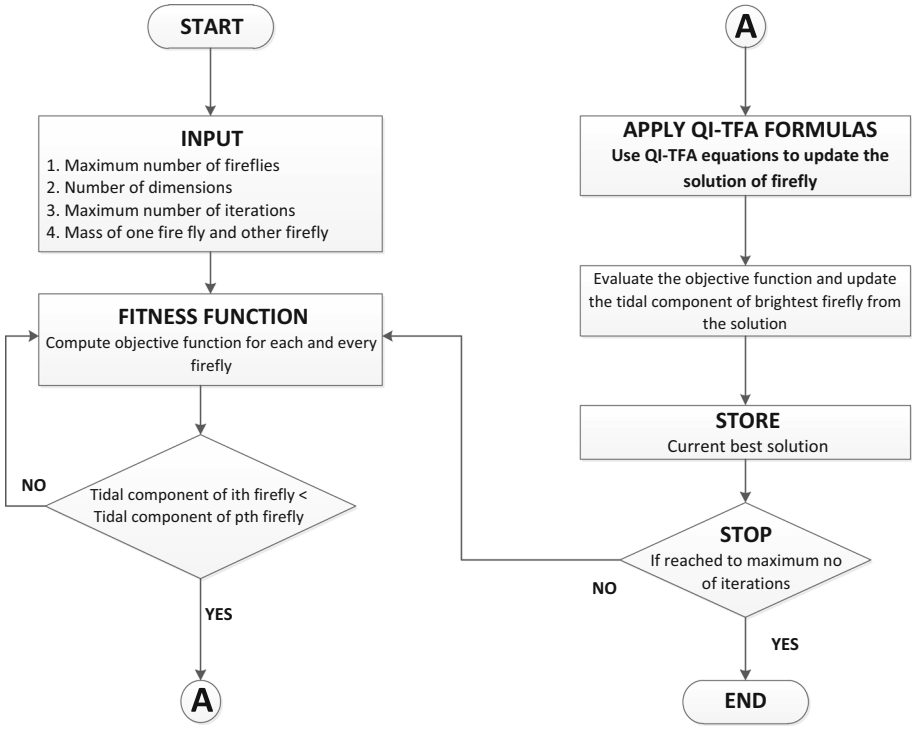


Fig. 1 Flowchart of the proposed QITFA

5.2 IEEE 30-bus

In this case, QITFA is implemented on IEEE 30-bus test system having six generating units to solve MEED problem for different loads (i.e., 150 MW, 175 MW, 200 MW and 225 MW). The implementation of QITFA is initialized with 500 iterations, 1500 fireflies and 6 dimensions, and its results are compared with different pre-existing techniques. In these techniques, cubic criterion function has been considered to model multiple economic and emission dispatch problem and transformed into single-objective one by taking max to max PFs into account. These techniques are limited to only LR, PSO, SA, QBA and QPSO. Table 3 shows the fuel cost coefficient and power generation limits for a six generator system. The emission penalty cost of pollutants like SO_x, NO_x and CO_2 have been obtained after taking max–max PF for respective pollutants and is shown in Tables 4, 5 and 6, respectively. Table 7 demonstrates the complete results of QITFA for IEEE 30-bus test system with emission (kg/h) and emission penalty cost ($/h) of pollutants such as SO_x, NO_x and CO_2, fuel cost ($/h) and total fuel cost ($/h) for various loading conditions.

Table 8 shows the comparison of fuel cost for IEEE 30-bus test system (6-unit) for various loads. It may be seen that QITFA is better than LR, PSO and SA methods except for the 150 MW case. While comparing with other quantum-infused algorithms, the results are quiet similar. In Fig. 3, the variation in fuel cost with different loads for all the methods is presented. Table 9 shows the value of CO_2 emission in kg/h for various loads. It may be inferred that the quantum-infused algorithms perform much better than PSO, LR and SA methods. For 150 MW and 225 MW load, QITFA offers better CO_2 emission reduction than QPSO and QBA. Table 10 shows the SO_x emission in Kg/h for various loads. In this case, the QITFA shows its dominance over all the other discussed methods. The proposed method has the least SO_x emission in all load conditions. Table 11 shows the NO_x emission in kg/h for various loads. The QITFA is much better than LR, PSO and SA methods for all the load cases. While comparing with QBA and QPSO, QITFA is found to yield reduced NO_x emission for 200 MW and 225 MW load cases. Table 12 shows the overall emission of the system for various loads and it may be deduced that QITFA is an ideal choice for a cleaner environment (Figs. 4, 5, 6).

Table 13 shows the comparison of total cost for IEEE 30-bus test system (6-unit) for various loads by using different techniques where QITFA has shown tremendous superiority and excellence, over all its counter parts, i.e., LR, PSO, SA, QBA and QPSO by the way of providing best solution which is the minimum total cost for different loads (150 MW, 175 MW, 200 MW and 225 MW). It must be noted that the proposed manuscript aims to minimize the overall cost which includes fuel cost and the penalty cost due to emission, with faster convergence profile (Figs. 7, 8).

Table 1 Parameter settings of referred optimization techniques

Sl. no.	Algorithm	Parameters	Values
1	QITFA	Number of dimensions, D	6
		Gravitational constant [32], G	0.667
		Mass of the particles [32], m_1, m_2	0.1 and 0.1
		g	$\frac{t_{\max}}{t_{\max}(e+0.5)-e*t}$
2	SA [31]	Initial temperature	10 °C
		Alpha	0.99
		Max tries	10,000
		Final temperature	1e–1 °C
3	QPSO [43]	Population size	1000
		Maximum iteration	100
		Dimension	6
		Contraction expansion factor	$\beta = (1-0.5)\left(\frac{t_{\max}-t}{t_{\max}}\right)+0.5$
4	QBA [44]	The maximal and minimal pulse rate	1 and 0, respectively
		The maximal and minimal frequency ($n_{\max}$ and $n_{\min}$)	1.5 and 0, respectively
		The maximal and minimal loudness	2 and 1, respectively
		Delta, δ	0.9
		Gamma, γ	0.9
		The frequency of updating the loudness and emission pulse rate, G	10
		The maximum and minimum probability of habitat selection	0.9 and 0.6, respectively
		The maximum and minimum compensation rate for Doppler effect ($C_{\max}$ and $C_{\min}$)	0.9 and 0.1, respectively
		The maximum and minimum contraction expansion coefficient ($\beta_{\max}$ and β_{min})	1 and 0.5, respectively
		The maximum and minimum inertia weight (w_{max} and $w_{\min}$)	0.9 and 0.5, respectively
5	LR [30]	Maximum iteration	1500
		Initial Lagrange multipliers value, λ	8

Table 2 Comparison of total cost for 3-unit IEEE 14-bus test system for a load demand of 2500 MW

Unit	GA [42]	PSO [42]	QPSO [43]	QITFA
1	725.02	724.99	726.866	727.757
2	910.19	910.15	908.4451	907.9965
3	864.88	864.85	864.6898	864.2465
Total cost ($/h)	22,730.14	22,729.35	22,728.52	22,727.92

Figure 9 represents convergence characteristics of QITFA to solve MEED problem for various load demands. From the graph, it may be seen that the curves are smooth and consistent. Also, the curves are converging in less than 15 iterations for all the cases of load variations. When compared with other methods, QITFA requires very less iterations to reach the global optimal point. LR [31], SA [31], QBA [44], PSO [31] and QPSO [43] require 1500, 10,000, 200, 500 and 20 iterations, respectively, to obtain its optimal point, as given in Table 14. Table 14 also represents the comparative analysis of all the employed algorithms. It may be seen from this table that standard deviation of the proposed algorithm is less than all the reported techniques which makes the algorithm more robust. The CPU time has not been reported by Mahdi et al. [43, 44]. However, it may be noted that the proposed technique takes 5.5224, 5.5889, 5.5118 and 5.5198 s for 150, 175, 200 and 225 MW load, respectively.

6 Conclusion and scope of future work

In this manuscript, TFA is infused with the concept of quantum mechanics and is proposed as QITFA. The proposed technique is applied to solve MEED problem of a three generator, IEEE 14-bus test system and a six generator IEEE 30-bus test system for various loads. QITFA and its application in solving MEED problem are proposed for the very first time in this work. The MEED problem is represented by a cubic polynomial with four objective functions, three for pollutants (SO_x, NO_x and CO_2) and one

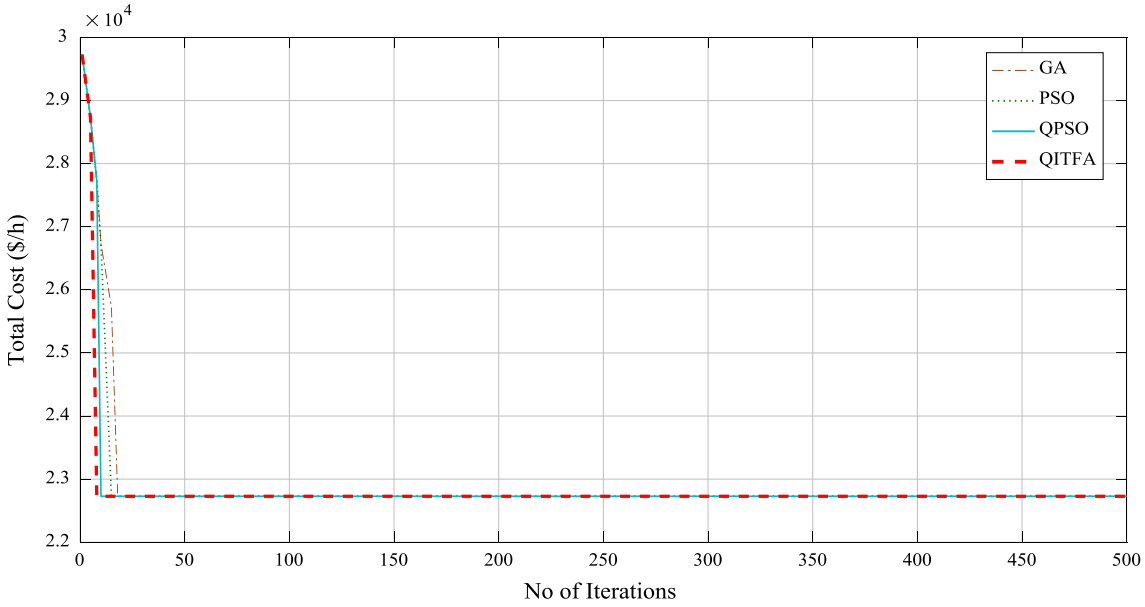


Fig. 2 Comparative convergence characteristics for 3-unit IEEE 14-bus test system for a load demand of 2500 MW

Table 3 Fuel cost coefficient and generator power limits for IEEE 30-bus test system

Generator	At bus	Generator limits		Cost coefficients of fuel			
		$P_{i,mu}$(MW)	$P_{i,mx}$ (MW)	$\omega_{0,i}$	$\omega_{1,i}$	$\omega_{2,i}$	$\omega_{3,i}$
1	1	50.00	200.00	− 136.00	14.50	0.0920	0.0010
2	2	20.00	80.00	− 3.50	22.00	0.0250	0.0004
3	5	15.00	50.00	− 81.00	23.00	0.0750	0.0006
4	8	10.00	50.00	− 14.50	13.50	0.1000	0.0002
5	11	10.00	50.00	− 9.75	11.50	0.1200	0.0013
6	13	12.00	40.00	75.60	12.50	0.0840	0.0004

Table 4 PF and emission coefficient of SO_x for IEEE 30-bus test system

Generator	At bus	Emission coefficients of SO_x				PF ($/kg)
		$\alpha_{0(SO_x),i}$	$\alpha_{1(SO_x),i}$	$\alpha_{2(SO_x),i}$	$\alpha_{3(SO_x),i}$	
1	1	− 90.00	17.00	0.150	0.0005	1.0851
2	2	− 30.50	12.00	0.055	0.0014	1.061
3	5	− 80.00	10.00	0.035	0.0010	2.1051
4	8	− 34.50	23.50	0.070	0.0020	0.5975
5	11	− 19.75	21.50	0.120	0.0013	0.6771
6	13	25.60	22.50	0.080	0.0021	0.6470

Table 5 PF and emission coefficient of NO_x for IEEE 30-bus test system

Generator	At bus	Emission coefficients of NO_x				PF ($/kg)
		$\alpha_{0(NO_x),i}$	$\alpha_{1(NO_x),i}$	$\alpha_{2(NO_x),i}$	$\alpha_{3(NO_x),i}$	
1	1	− 26.00	18.50	0.0520	0.0012	0.9407
2	2	− 35.00	12.00	0.0450	0.0004	1.4961
3	5	− 15.00	13.00	0.0500	0.0016	1.3869
4	8	− 74.50	17.50	0.0700	0.0012	0.8308
5	11	− 89.05	8.50	0.0400	0.0003	2.1705
6	13	− 75.00	15.50	0.0240	0.0014	1.0930

Table 6 PF and emission coefficient of CO_2 for IEEE 30-bus test system

Generator	At bus	Emission coefficients of CO_2				PF ($/kg)
		$\alpha_{0(CO_2),i}$	$\alpha_{1(CO_2),i}$	$\alpha_{2(CO_2),i}$	$\alpha_{3(CO_2),i}$	
1	1	− 16.00	14.00	0.0920	0.0015	0.7822
2	2	− 93.50	12.50	0.0250	0.0004	1.1895
3	5	− 85.00	13.50	0.0550	0.0016	1.4355
4	8	− 24.50	13.50	0.0100	0.0012	1.1332
5	11	− 59.00	21.00	0.0400	0.0023	0.7455
6	13	− 70.00	22.00	0.0800	0.0014	0.7158

Table 7 QITFA-based results of MEED problem for 6-unit IEEE 30-bus test system pertaining to different load demands

Attributes	150 MW	175 MW	200 MW	225 MW
Generation (MW)				
P_1	50.19	50.19	50.07	50.00
P_2	20.29	23.56	31.09	37.67
P_3	15.29	15.22	15.37	17.75
P_4	25.05	32.38	37.79	44.77
P_5	17.18	23.81	29.85	35.21
P_6	22.00	29.84	35.83	39.6
Total fuel cost ($/h)	2711.71	3189.91	3729.26	4316.34
Emission (kg/h)				
CO_2	2561.85	3133.60	3721.30	4322.39
SO_x	3087.92	3807.49	4528.28	5289.45
NO_x	2406.79	2856.22	3313.29	3815.71
Emission penalty cost[a] ($/h)				
CO_2	2245.4	2735.3	3265.3	3847.8
SO_x	2731.9	3212.4	3732	4311.6
NO_x	2565.9	3104	3687	4307.6
Total ($/h)	10,254.91	12,241.61	14,413.56	16,783.33

[a]Emission penalty cost ($/h) = PF × emission

for fuel cost. To convert these four functions into a single-objective function, max to max PF is used. The proposed algorithm is then applied and compared with other well-known optimization techniques like LR, GA, PSO, SA, QBA and QPSO. The effectiveness of the proposed QITFA may be clear from the results presented where it is defined that, for a three generator system, the total cost of fuel including PF when QITFA is employed is 22,727.92 $/h, which is much better than GA and PSO technique. For six generator system, the QITFA is compared with LR, PSO, SA, QBA and QPSO with different load demands (such as 150 MW, 175 MW, 200 MW and 225 MW) for fuel cost, total emission and total cost including PF minimization. As presented, QITFA is way better than the traditional LR, SA and PSO, and its supremacy in reduction in emission may be visualized where it surpasses every other technique taken into consideration. Overall reduction in total cost is obtained through QITFA. The obtained convergence characteristics are better than LR, PSO, SA, QBA and QPSO techniques. Therefore, the proposed QITFA optimization method is more economical with excellent convergence capability. The proposed algorithm may be applied in some large scale power system optimization application.

Table 8 Comparison of fuel cost results for 6-unit IEEE 30-bus test system corresponding to different load demands

Load demand (MW)	Total fuel cost ($/h)					
	LR [31]	PSO [31]	SA [31]	QBA [44]	QPSO [43]	QITFA
150	2729.35	2734.2	2705.21	2704.97	2704.89	2711.71
175	3475.41	3236.3	3220.51	3188.10	3187.93	3189.91
200	4210.30	3784.9	3735.73	3726.08	3727.61	3729.26
225	5130.53	4402.3	4321.52	4314.63	4315.03	4316.34

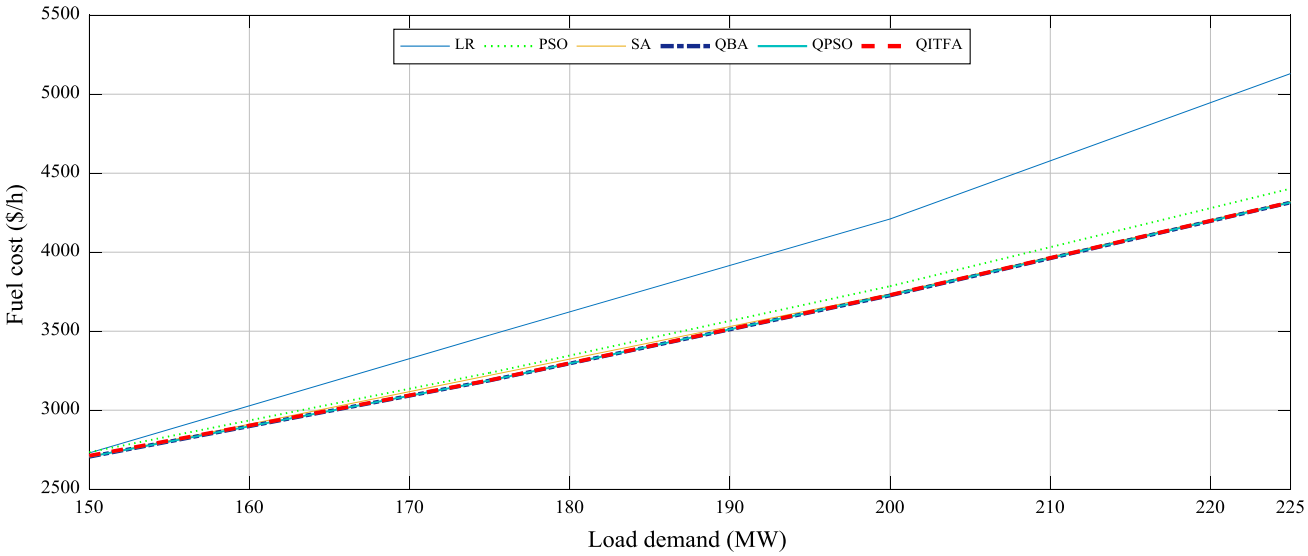


Fig. 3 Comparison of fuel cost for various load demands in 6-unit IEEE 30-bus test system

Table 9 Comparison of emission of CO_2 pollutant for 6-unit IEEE 30-bus test system pertaining to different load demands

Load demand (MW)	CO_2 emission (Kg/h)					
	LR [31]	PSO [31]	SA [31]	QBA [44]	QPSO [43]	QITFA
150	2537.122	2607.1	2568.946	2565.14	2564.82	2561.85
175	3613.531	3178	3094.688	3129.78	3129.91	3133.60
200	4473.369	3771.5	3714.333	3719.64	3715.45	3721.30
225	5502.522	4403	4324.3	4323.47	4322.77	4322.39

Table 10 Comparison of emission of SO_x pollutant for 6-unit IEEE 30-bus test system corresponding to different load demands

Load demand (MW)	SO_x emission (kg/h)					
	LR [31]	PSO [31]	SA [31]	QBA [44]	QPSO [43]	QITFA
150	3091.648	3193.6	3138.446	3147.38	3146.86	3087.92
175	4142.176	3904.9	3763.478	3858.96	3859.89	3807.49
200	5053.584	4670.6	4553.972	4598.20	4591.38	4528.28
225	6106.498	5426.1	5287.306	5344.75	5340.20	5289.45

Table 11 Comparison of emission of NO_x pollutant for 6-unit IEEE 30-bus test system corresponding to different load demands

Load demand (MW)	NO_x emission (kg/h)					
	LR [31]	PSO [31]	SA [31]	QBA [44]	QPSO [43]	QITFA
150	2448.218	2424.6	2379.35	2408.10	2406.37	2406.79
175	2604.886	2879.7	2789.92	2853.40	2854.22	2856.22
200	3102.077	3373.2	3285.647	3327.78	3324.62	3313.29
225	3798.383	3877.6	3781.191	3822.53	3820.14	3815.71

Table 12 Comparison of total emission of all pollutants for 6-unit IEEE 30-bus test system corresponding to different load demands

Load demand (MW)	Total emission (kg/h)					
	LR [31]	PSO [31]	SA [31]	QBA [44]	QPSO [43]	QITFA
150	8076.988	8225.3	8086.742	8120.62	8118.05	8056.56
175	10,360.59	9962.6	9648.086	9842.14	9844.02	9797.31
200	12,629.03	11,815.3	11,553.95	11,645.62	11,631.45	11,562.87
225	15,407.403	13,706.7	13,392.8	13,490.75	13,483.11	13,427.55

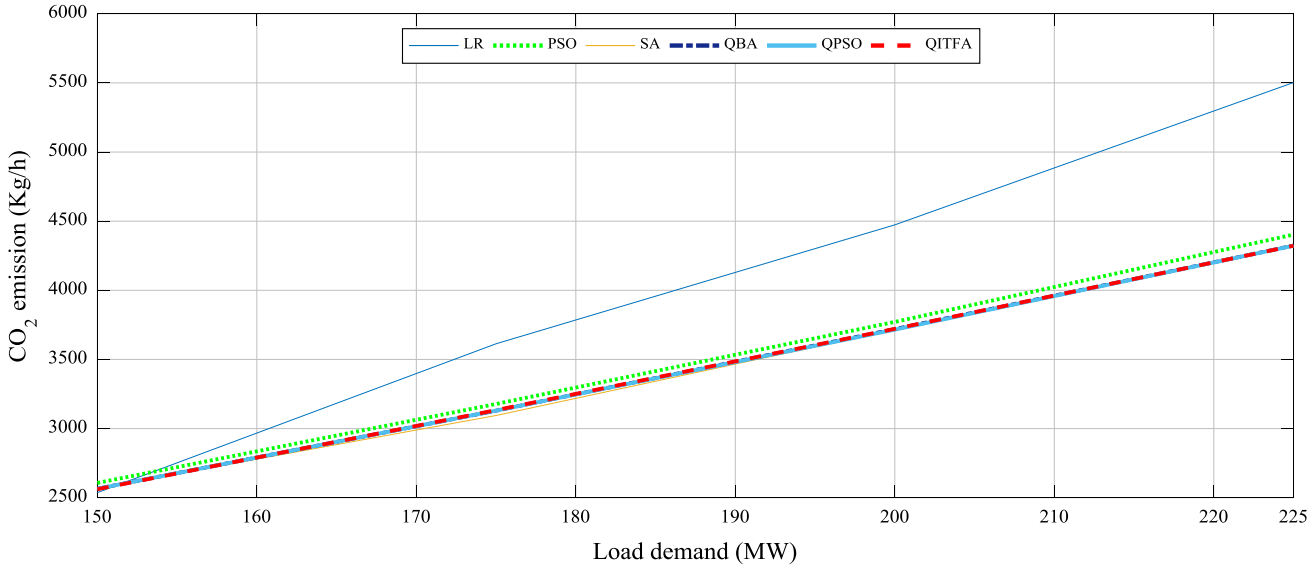


Fig. 4 Comparison of emission of CO_2 pollutant for various load demands in 6-unit IEEE 30-bus test system

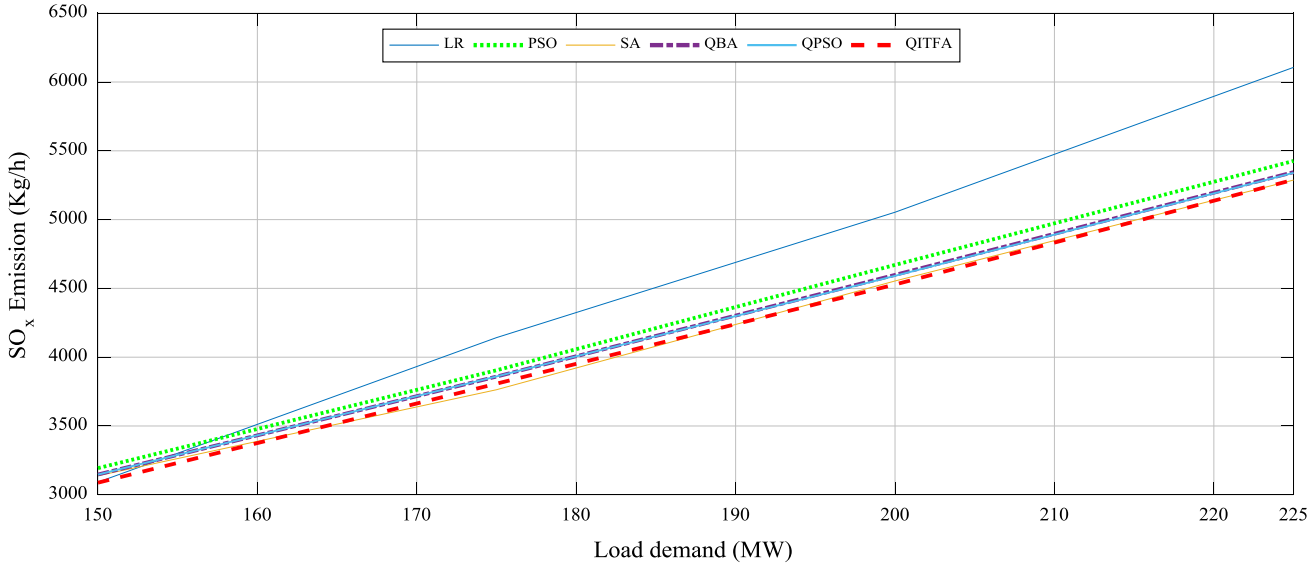


Fig. 5 Comparison of emission of SO_x pollutant for various load demands in 6-unit IEEE 30-bus test system

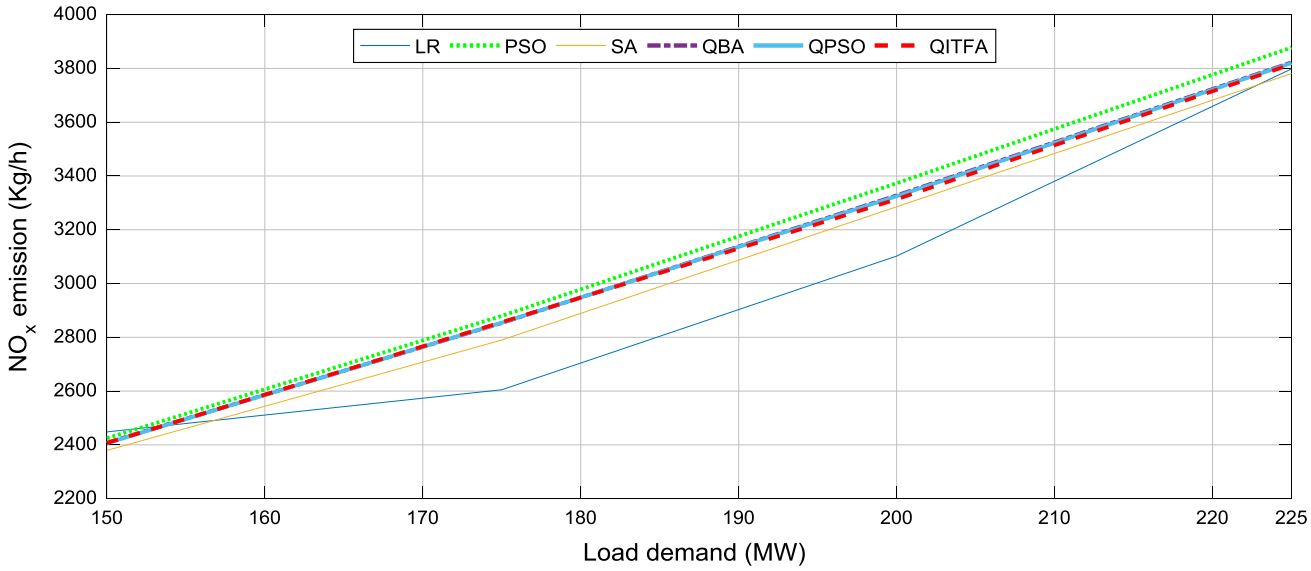


Fig. 6 Comparison of emission of NO_x pollutant for various load demands in 6-unit IEEE 30-bus test system

Table 13 Comparison of total cost for 6-unit IEEE 30-bus test system corresponding to different load demands

Load demand (MW)	Total cost ($/h)					
	LR [31]	PSO [31]	SA [31]	QBA [44]	QPSO [43]	QITFA
150	10,264.57	10,385	10,261.49	10,255.28	10,255.25	10,254.91
175	13,251.52	12,425	12,280.04	12,241.74	12,241.71	12,241.61
200	16,077.41	14,642	14,421.30	14,413.88	14,413.77	14,413.56
225	19,661.33	17,125	16,790.69	16,783.91	16,783.86	16,783.33

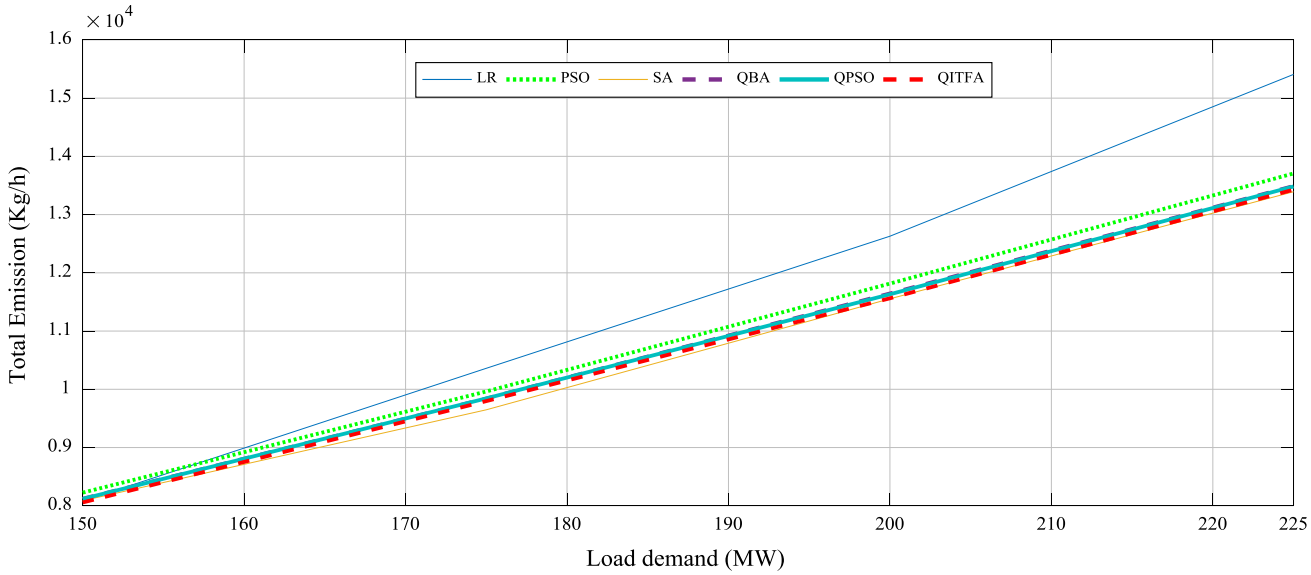


Fig. 7 Comparison of total emission of all pollutants for various load demands in 6-unit IEEE 30-bus test system

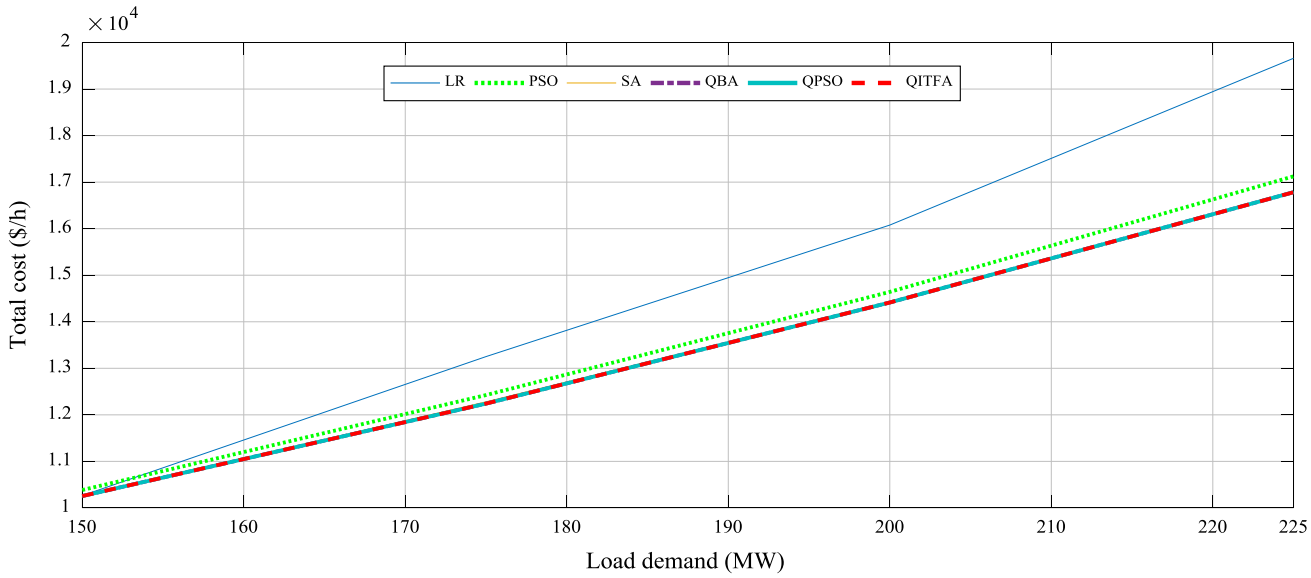


Fig. 8 Comparison of total cost for various load demands in 6-unit IEEE 30-bus test system

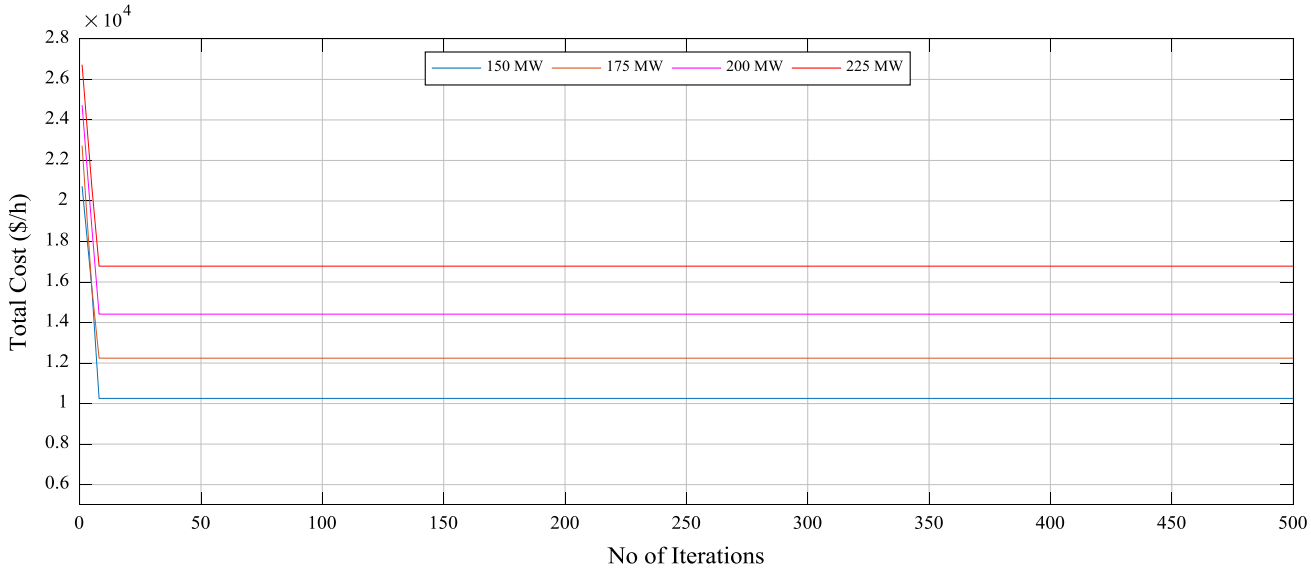


Fig. 9 QITFA-based convergence characteristics for various loads demands in 6-unit IEEE 30-bus test system

Table 14 Comparative results of all the compared algorithms

Load demand (MW)	Total cost ($/h)					
	LR [31]	PSO [31]	SA [31]	QBA [44]	QPSO [43]	QITFA
Standard deviation						
150	–	–	–	0.05393	0.00815	0.00792
175	–	–	–	0.02829	0.00614	0.00602
200	–	–	–	0.10581	0.00193	0.00098
225	–	–	–	0.04982	0.00020	0.00015
Average iterations						
150	1500	500	10,000	200	20	15
175	1500	500	10,000	200	20	15
200	1500	500	10,000	200	20	15
225	1500	500	10,000	200	20	15
CPU time (s)						
150	10.456	–	34.1570	–	–	5.5224
175	11.223	–	31.3873	–	–	5.5889
200	10.820	–	39.4865	–	–	5.5118
225	10.267	–	21.9241	–	–	5.5198

An entry "–" means not available in the referred literature

Compliance with ethical standards

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

References

1. Caldeira K, Kasting J (1993) Insensitivity of global warming potentials to carbon dioxide emission scenarios. Nature 366:251–252
2. Smith J, Wigley ML (2000) Global warming potentials: 1. Climatic implications of emissions reductions. Clim Change 44:445–457
3. Shang N, Guo L, Ding Y, Lin Y, Liu L, Shao C, Song Y (2017) Analysis of the influence of renewable energy generation on market power. J Eng 2017:1928–1933
4. Song YH et al (1997) Environmental/economic dispatch using fuzzy logic controlled genetic algorithms. IEEE Proc Gener Transm Distrib 144:377–382. https://doi.org/10.1049/ip-gtd:19971100
5. Nanda J, Kothari D, Lingamurthy K (1988) Economic-emission load dispatch though goal programming techniques. IEEE Trans Energy Convers 3:26–32. https://doi.org/10.1109/60.4195
6. Venkatesh P, Gnanadass R, Narayana PP (2003) Comparison and application of evolutionary programming techniques to combined economic emission dispatch with line flow constraints. IEEE Trans Power Syst 18:688–697. https://doi.org/10.1109/TPWRS.2003.811008
7. Abido MA (2003) Environmental/economic power dispatch using multiobjective evolutionary algorithms. IEEE Trans Power Syst 18:1529–1537. https://doi.org/10.1109/TPWRS.2003.818693
8. Wang L, Singh C (2007) Environmental/economic power dispatch using a fuzzified multi-objective particle swarm optimization algorithm. Electr Power Syst Res 77:1654–1664. https://doi.org/10.1016/j.epsr.2006.11.012
9. Raglend IJ et al (2010) Comparison of AI techniques to solve combined economic emission dispatch problem with line flow constraints. Int J Electr Power Energy Syst 32:592–598. https://doi.org/10.1016/j.ijepes.2009.11.015
10. Jiejin C et al (2010) A multi-objective chaotic ant swarm optimization for environmental/economic dispatch. Int J Electr Power Energy Syst 32:337–344. https://doi.org/10.1016/j.ijepes.20
11. Güvenç U et al (2012) Combined economic and emission dispatch solution using gravitational search algorithm. Sci Iran 19(6):1754–1762. https://doi.org/10.1016/j.scient.2012.02.030
12. Gopalakrishnan R, Krishnan A (2013) An efficient technique to solve combined economic and emission dispatch problem using modified Ant colony optimization. Sadhana 38(4):545–556. https://doi.org/10.1007/s12046-013-0153-1
13. Varma TD, Kumar V (2014) Multi-objective economic emission load dispatch using teacher-learning-based optimization technique. IFAC Proc Vol 47(1):819–826. https://doi.org/10.3182/20140313-3-IN-3024.00153
14. Doğan A et al (2014) Artificial bee colony algorithm with dynamic population size to combined economic and emission dispatch problem. Int J Electr Power Energy Syst 54:144–153. https://doi.org/10.1016/j.ijepes.2013.06.020
15. Hadji B et al (2015) Multi-objective economic emission dispatch solution using dance bee colony with dynamic step size. Energy Procedia 74:65–76. https://doi.org/10.1016/j.egypro.2015.07.524
16. Tamura K, Yasuda K (2017) The spiral optimization algorithm: convergence condition and settings. IEEE Trans Syst Man Cybern. https://doi.org/10.1109/TSMC.2017.2695577
17. Nasir ANK, Tokhi MO (2015) An improved spiral dynamic optimization algorithm with engineering application. IEEE Trans Syst Man Cybern Syst 45:943–954. https://doi.org/10.1109/TSMC.2014.2383995
18. Cruz-Duarte JM et al (2017) Primary study on the stochastic spiral optimization algorithm. In: IEEE international autumn meeting on power, electronics and computing (ROPEC 2017) Mexico. https://doi.org/10.1109/ROPEC.2017.8261609
19. Mahdi FP et al (2016) Emission dispatch problem with cubic function considering transmission loss using particle swarm optimization. J Telecommun Electron Comput Eng (JTEC) 8(12):17–21

20. Abbas G et al (2017) Solution of an economic dispatch problem though particle swarm optimization: a detailed survey—part I. IEEE Access 5:15105–15141. https://doi.org/10.1109/ACCESS.2017.2723862
21. Abbas G et al (2017) "Solution of an economic dispatch problem through particle swarm optimization: a detailed survey-part II. IEEE Access 5:24426–24445. https://doi.org/10.1109/ACCESS.2017.2768522
22. Sharafia Y, Khanesar MA, Teshnehla M (2016) COOA: competitive optimization algorithm. Swarm Evol Comput 30:39–63. https://doi.org/10.1016/j.swevo.2016.04.002
23. Roy PK, Sur A, Pradhan DK (2013) Optimal short-term hydro-thermal scheduling using quasi-oppositional teaching learning based optimization. Eng Appl Artif Intell 26:2516–2524. https://doi.org/10.1016/j.engappai.2013.08.002
24. Roy PK, Bhui S (2013) Multi-objective quasi-oppositional teaching learning based optimization for economic emission load dispatch problem. Int J Electr Power Energy Syst 53:937–948. https://doi.org/10.1016/j.ijepes.2013.06.015
25. Gupta S, Kumar S (2017) Artificial mountain ape optimization algorithm for maximum power point tracking under partial shading condition. In: International conference on energy, communication, data analytics and soft computing. https://doi.org/10.1109/icecds.2017.8389547
26. Bodha KD et al (2018) A levy flight based voltage particle swarm optimization for multiple-objective mixed cost-effective emission dispatch cloud computing. In: IEEE international conference on. data science & engineering (CONFLUENCE). https://doi.org/10.1109/confluence.2018.8442919
27. Yao F et al (2012) Quantum-inspired particle swarm optimization for power system operations considering wind power uncertainty and carbon tax in Australia. IEEE Trans Ind Inform 8(4):880–888. https://doi.org/10.1109/TII.2012.2210431
28. Babak J, Vahidinasab V (2014) A modified harmony search method for environmental/economic load dispatch of real-world power systems. Energy Convers Manag 78:661–675. https://doi.org/10.1016/j.enconman.2013.11.027
29. Muwaffaq IA, Oweis ZB (2018) Environmental economic dispatch using stochastic fractal search algorithm. Int Transactions on Electr Energy Syst 28:e2530. https://doi.org/10.1002/etep.2530
30. Krishnamurthy S, Tzoneva R (2012) Impact of price penalty factors on the solution of the combined economic emission dispatch problem using cubic criterion functions. In: Power and Energy Society General Meeting, 2012 IEEE. https://doi.org/10.1109/PESGM.2012.6345312
31. Ziane I, Benhamida F, Graa A (2017) Simulated annealing algorithm for combined economic and emission power dispatch using max/max price penalty factor. Neural Comput Appl 28:197–205. https://doi.org/10.1007/s00521-016-2335-3
32. Yelghi A, Köse C (2018) A modified firefly algorithm for global minimum optimization. Appl Soft Comput 62:29–44. https://doi.org/10.1016/j.asoc.2017.10.032
33. Abido MA (2006) Multiobjective evolutionary algorithms for electric power dispatch problem. IEEE Trans Evol Comput 10(3):315–329. https://doi.org/10.1109/TEVC.2005.857073
34. Pitono J, Soepriyanto A, Purnomo MH (2014) Advance optimization of economic emission dispatch by particle swarm optimization (PSO) using cubic criterion functions and various price penalty factors. Kursor (KL) 7(3):153–164
35. Qu BY et al (2017) Solving dynamic economic emission dispatch problem considering wind power by multi-objective differential evolution with ensemble of selection method. Nat Comput. https://doi.org/10.1007/s11047-016-9598-6
36. Yang XS (2010) Nature-inspired metaheuristic algorithms. Luniver Press, Bristol
37. Abdullah A et al (2012) A new hybrid firefly algorithm for complex and nonlinear problem. In: Omatu S, De Paz Santana J, González S, Molina J, Bernardos A, Rodríguez J (eds) Distributed computing and artificial intelligence. Advances in intelligent and soft computing, vol 151. Springer, Berlin
38. Zouache D, Nouioua F, Moussaoui A (2016) A Quantum-inspired firefly algorithm with particle swarm optimization for discrete optimization problems. Soft Comput 20(7):2781–2799. https://doi.org/10.1007/s00500-015-1681-x
39. Wong LA et al (2014) Novel quantum-inspired firefly algorithm for optimal power quality monitor placement. Front Energy 8(2):254–260. https://doi.org/10.1007/s11708-014-0302-1
40. Manju A, Nigam MJ (2014) Applications of quantum inspired computational intelligence: a survey. Artif Intell Rev 42:79–156. https://doi.org/10.1007/s10462-012-9330-6
41. Yumin D, Li Z (2014). Quantum behaved particle swarm optimization algorithm based on artificial fish swarm. In: Mathematical problems in engineering. http://dx.doi.org/10.1155/2014/592682
42. Adhinarayanan T, Sydulu M (2006) Particle swarm optimization for economic dispatch with cubic fuel cost function. In: TENCON 2006. 2006 IEEE Region 10 Conference, pp 1–4. https://doi.org/10.1109/TENCON.2006.344059
43. Mahdi FP et al (2017) A quantum inspired particle swarm optimization approach for environmental/economic power dispatch problem using cubic criterion function. Int Trans Electr Energy Syst 28:2497. https://doi.org/10.1002/etep.2497
44. Mahdi FP et al (2018) Quantum-behaved bat algorithm for many-objective combined economic emission dispatch problem using cubic criterion function. Neural Comput Appl. https://doi.org/10.1007/s00521-018-3399-z

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

Applied Physics A (2019) 125:600
https://doi.org/10.1007/s00339-019-2900-6

Applied Physics A
Materials Science & Processing

Impact of metal silicide source electrode on polarity gate induced source in junctionless TFET

Jaya Madan[1] · Rahul Pandey[1] · Rajnish Sharma[1] · Rishu Chaujar[2]

Received: 27 June 2019 / Accepted: 2 August 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract
Tunnel Field Effect Transistors (TFET) based on quantum mechanical band to band tunneling (BTBT) are promising alternatives for low power analog applications. Additionally, the concept of junctionless (JL) devices realized by the charge plasma concept offers added advantages in terms of simplified fabrication techniques. In n type JL-TFET, p+ source is induced using a polarity gate (PG) with suitable work function. However, retention of induced p+ source is not a sole contribution of PG, the source electrode (SE) metal silicide work function also plays a significant role in the retention of hole plasma (specially near the interface of SE/induced source). Thorough study regarding the combined influence of PG and SE metal silicide work function on induced p+ source is missing in the literatures. This work explores the performance of JL-TFET of different SE metal silicide such as $TiSi_2$ (4.53 eV), $CrSi_2$ (4.85 eV) and Pd_2Si (5.3 eV). It is perceived that for SE metal silicide with work function lower than p+ induced source i.e., $TiSi_2$ and $CrSi_2$ the depletion of hole plasma (formation of Schottky interface) appears near the SE/p+ induced source interface. The depletion of hole plasma is attributed to the combined electric field of SE metal silicide and the PG, the immediate consequence is the refrainment of current. Further, due to the formation of Schottky interface for $TiSi_2$ and $CrSi_2$, the performance of the device is examined by revoking and evoking the Universal Schottky Tunneling (UST) model. Results reveal the undervalued performance of the device without the inclusion of UST, primarily a lower drain current (and thereby the analog performance) of the device is obtained, since it ignores the Schottky tunneling at SE/p+ induced source interface. However, the inclusion of UST model emulates the performance of JL-TFET precisely, by incorporating the Schottky tunneling at the SE/p+ induced source interface. Thus, for the retention of hole plasma, appropriate SE work function i.e., $\Phi_{SE} > \Phi_{p+\text{ induced source}}$ is required, whereas for SE work function $\Phi_{SE} < \Phi_{p+\text{ induced source}}$ appropriate Schottky tunneling must be considered for accurate analysis of the device. The study also reveals that the depletion of hole plasma and hence the formation of Schottky interface can be avoided using SE with metal work function source for which consideration of UST is immaterial.

1 Introduction

Huge growth in applications such as 5G, Internet of Things, big data analytics, cloud computing and artificial intelligence have increased the demand of most prompt electronic devices like never before. These next generation devices are expected to respond and act favourably to the clock frequencies as high as few GHz. While conventional devices such as nMOS, pMOS, combination of the two i.e., CMOS, or other forms of the same such as High Electron Mobility Transistor (HEMT), BiCMOS and Heterojunction Bipolar Transistor (HBT) have been more or less meeting the requirements of these technological changes, but their performance falters at the sub 20 nm technology node. Though the continuous scaling down of CMOS technology from last few decades has enhanced the packing density and operation speed, however,

✉ Rahul Pandey
rahul.pandey@chitkara.edu.in

Jaya Madan
jaya.madan@chitkara.edu.in

Rajnish Sharma
rajnish.sharma@chitkarauniversity.edu.in

Rishu Chaujar
chaujar.rishu@dtu.ac.in

[1] VLSI Centre of Excellence, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

[2] Microelectronics Research Lab, Department of Applied Physics, Delhi Technological University, Delhi, India

Published online: 08 August 2019

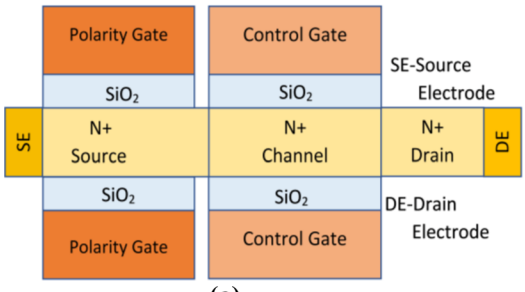
most of these devices encounter critical challenges in the fabrication of ultra-shallow junctions [1–3]. Primarily, the fabrication of ultra-shallow junctions with heavily doped source/drain region in sub-nm regime requires efficient ion implantation and annealing techniques [4]. In addition to that, heavily doped source/drain regions also suffer from the random dopant fluctuations [5]. Thus, to overcome these fabrication challenges, the junctionless devices are introduced that utilize the charge plasma concept to induce source and drain regions [3].

Besides these fabrication related challenges, the downscaling of CMOS transistors enhances the static power dissipations owing to the increased leakage current [6, 7]. This never-ending demand of low voltage and high-speed devices motivates to explore the devices with steeper swing. To meet both, the lower leakage current and steeper switching, tunnel field effect transistors (TFETs) have been proposed [8]. TFET remains one of the most researched devices so as to possibly fill these gaps. A large number of researchers have highlighted the strengths of this device such as low values of subthreshold swing, and extremely low OFF current. TFET employs the modulation of the tunneling barrier controlled by the gate bias and overwhelm the thermionic limit of 60 mV/decade at room temperature [9, 10]. The band to band tunneling (BTBT) utilized by TFET enables the scaling of supply voltage along with high I_{ON}/I_{OFF} current switching ratio [11]. However, there have been published reports about the difficulties faced in the realization of a TFET, which could meet the requirements of modern-day technological applications. Junction Less TFETs have drawn a huge attention of the researchers to overcome the challenges posed in the TFET fabrication processes. Junctionless devices are introduced that utilize the charge plasma concept to induce source and drain regions [3]. The concept of junctionless FETs does not require the formation of *pn* junctions within the device and thereby uproots the challenges related to the formation of sharp junctions. The JL-FETs can be realized (i) using a thin semiconductor layer with a thickness less than the Debye length to acquire full depletion of carriers, and (ii) heavily doped semiconductor so as to account for the sufficient amount of carriers to flow in ON state [12, 13]. In this work, the merits of TFET and JL-FET are integrated and the resulting device junctionless tunnel FET (JL-TFET) is explored. Most of the work in JL-TFET has been focused upon studying the influence of work function of the control/polarity gate for analysing/optimizing the performance of the JL-TFET [1, 3, 4, 14]. However, the energy band profile reported by several researchers shows that the depletion of hole plasma (HP) due to the source electrode (SE) metal near the induced source and SE region [1, 3, 15–17] has mostly been overlooked. The depleted HP forms a Schottky interface at the interface of SE/induced source region. Most of the researchers working in this field have not incorporated UST mechanism through the M/S interface [1, 3, 15–17]. That way, the performance of the device reported earlier is somehow undervalued due to the ignorance of UST at SE/source interface. To take into account the same, a thorough focus has been laid upon the importance of choice of SE metal in this work. Besides this, another ignored issue about due consideration of Universal Schottky Tunneling (UST) model in a JL-TFET has been deeply taken into account. Present piece of work may be extremely useful to put the research effort on JL-TFETs in a right direction and opening up new avenues about many other effects which might have not been taken into consideration so far.

This paper has been divided in four sections other than the references section provided at the end. Section 2 describing all the details about the device design, structural parameters and simulation methodology implemented for this research effort follows the present section of introduction. Results with their justifications are presented in Sect. 3. Further, the conclusion of the work done are summarised in the Sect. 4.

2 Device description and simulation methodology

Figure 1a show the cross-sectional view of the simulated device structure i.e., JL-TFET. The doping concentration and thickness of n+ doped silicon bar is 1×10^{19} cm^{-3}



(a)

Device parameter	Value
Polarity gate (PG) length	20nm
Control gate (CG) length	20nm
Oxide thickness	2nm
Work function of CG	4.2eV
Work function of PG	5.93eV
Gap Length between PG and CG	5nm

(b)

Fig. 1 **a** Schematic device structure of junctionless TFET (JL-TFET) used in the simulation. **b** Summary of device structural parameters

and 5 nm, respectively. The device structural parameters used in the simulations are tabulated in Fig. 1b. A polarity gate (PG) platinum (Pt) /control gate (CG) titanium (Ti) is utilized to induce the p+ -source/p-channel. The work function of PG and CG are chosen by tuning them for optimized analog performance (as also done in [18]). To analyze the influence of SE metal silicide on the performance of the JL-TFET, three metal silicides viz. $TiSi_2$ (4.53 eV), $CrSi_2$ (4.85 eV) and Pd_2Si (5.3 eV) are employed as SE [19]. To emulate the physics involved in the device, various models are incorporated in the simulation setup using SILVACO ATLAS device simulator. To account for the tunneling via trap states (having energy levels in between the top of valence band and bottom of conduction band), the TAT model given by Schenk is considered. Further, for accurate simulations, the quantum confinement effect is incorporated as given by Hansch quantum mechanical model. Details of these models can be found in [20]. Additionally, to include the BTBT at the source-channel junction, nonlocal BTBT model is evoked in the simulation setup. The nonlocal BTBT model considers the generation rate of carriers throughout the meshing at the tunneling interface [21, 22]. Additionally, for efficient nonlocal BTBT, the quantum tunneling meshing has also been done in the tunneling region. Besides this Shockley Read Hall, bandgap narrowing; Fermi Dirac statistics have also been incorporated in the simulation setup. To account for the design of a JL-TFET, with the work function of SE metal silicide lower than the source, the device performance is analyzed by including the UST model and the results thus obtained are simultaneously compared with the results obtained after the exclusion of UST. This is done to account for the Schottky tunneling at the source electrode and p+ induced source region, that is formed when the $\Phi_{SE} < \Phi_{p+\text{ induced source}}$. To the best of our knowledge, incorporation of UST model in the studies related to JL-TFETs has not been carried out earlier.

3 Results and discussions

In this section, the electrical/analog performance of conventional JL-TFET is inspected for three metal silicides viz., $TiSi_2$, $CrSi_2$ and Pd_2Si as SE.

Figure 2a shows the impact of SE metal silicide on the surface potential in the equilibrium state (i.e., at $V_{gs}=0$ V and $V_{ds}=0$ V). It is clearly evident that for the SE metal silicides, $TiSi_2$ and $CrSi_2$ which have lower metal work function (than the induced p+ source), the surface potential at the interface of SE and the induced p+ , depletes the induced HP (near the SE/p+ interface) formed by the PG. Whereas, for SE metal silicide Pd_2Si, the hole charge plasma formed by the PG is retained owing to its large work function than the p+ source. Observance of these kind of plots for the variation of surface potential with selection of electrodes with different work functions is very much in sync with the pre-existing theories related to this aspect of device physics [13]. Figure 2b–c shows the energy band profile with different SE metal silicides at $V_{gs}=0$ V and $V_{ds}=0$ V. It is clearly depicted from the band profile that for the SE metal silicides $TiSi_2$ and $CrSi_2$, the interface between SE and induced p+ source is Schottky, however, for SE metal silicide Pd_2Si which have higher work function than induced source, it becomes ohmic (which is in agreement with the results discussed for Fig. 2a). Band profile as shown in Fig. 2c illustrates the depletion of HP at the interface of the induced p+ source (i.e., induced p+ source region no longer remains p+ type at the SE/p+ interface). It clearly depicts the probability of electron movement towards the SE that reduces the effective charge carrier available for tunneling (the effective device conduction). However, for SE metal silicide Pd_2Si, the electron movement is only in the induced p+ source direction that is desired for effective conduction of the device shown in Fig. 2d. Further from Fig. 2b it is evident that for SE metal silicide Pd_2Si, the PG-induced HP is unaltered. Thus, the HP formed beneath the PG is not solely

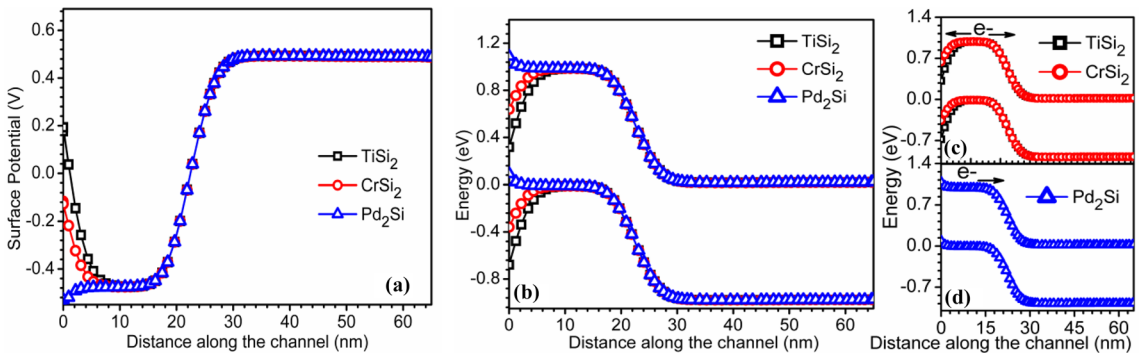


Fig. 2 **a** Influence of SE metal silicide on surface potential at no applied biased condition. **b** Energy bands of JL-TFET for $TiSi_2$, $CrSi_2$ and Pd_2Si at no applied biased condition. **c** Formation of Schottky interface with $TiSi_2$ and $CrSi_2$. **d** Energy bands for Pd_2Si

contributed by the PG, SE metal also plays a critical role in retaining that HP. Matsuzawa et al. have also worked on the transformation of type of contact from Schottky to Ohmic with the selection of materials with different work functions [23]. However, the kind of transformation of type of contact in the present scenario under the consideration of JL-TFETs opens up altogether new avenues of research on this device.

The direction of the electric field is obtained under the ON state ($V_{gs} = V_{ds} = 1.0$ V) of the device and is shown in Fig. 3a–c, by the vector representation, for different SE metal silicides $TiSi_2$, $CrSi_2$ and Pd_2Si. The electric field vector representation is shown only near the SE and p+ induced source interface owing to the depletion of HP by the SE to Debye length. It is clearly depicted that for SE metal silicide $TiSi_2$ and $CrSi_2$, the formation of electric field due to the depletion of HP near SE and induced source interface, refrain the current injection in the channel. Thereby, the overall current will be reduced for SE of $TiSi_2$ and $CrSi_2$, and there is no such refrainment for Pd_2Si. The electric field vector shows the motion of electrons from source to channel, as the direction of flow of electrons is opposite to the direction of electric field.

As shown in Fig. 3a, b, the movement of electrons get refrain towards the SE region. Same is confirmed by the considerably large/small electron/hole concentration near the SE for $TiSi_2$ and $CrSi_2$ depicted in Fig. 4a. This enhanced electron concentration (near SE) reduces the electron concentration at the source-channel junction and thus reduces the nonlocal BTBT rate of electrons as depicted in Fig. 4b. The reduced BTBT rate of electrons decreases the overall conduction of the device. This also reduces marginally the number of electrons available for BTBT at the source channel junction leading to the reduction in the value of the drain current. Owing to the presence of the Schottky interface for $TiSi_2$ and $CrSi_2$, the inclusion of Universal Schottky Tunneling (UST) model is demanded to emulate the actual performance of the device accurately. Inclusion of UST considers the tunneling phenomenon at the interface of SE and induced source. Same is shown for $TiSi_2$, $CrSi_2$ and Pd_2Si as SE in Fig. 4b and as a contour plot in Fig. 5a–c. For $CrSi_2$, the Schottky tunneling is large due to the low Schottky potential barrier and width shown in Fig. 2b. Whereas, the Schottky potential and width are comparatively large for $TiSi_2$ that leads to the lower Schottky tunneling. Further, for Pd_2Si, the absence of Schottky interface is observed due to the appropriate metal silicide (workfunction is higher than that of induced p+ source), and hence there is no Schottky tunneling (the UST model does not affect the device performance). The contour plot depicted in Fig. 5a–c shows the Schottky tunneling rate for SE as $TiSi_2$, $CrSi_2$ and Pd_2Si. It can be clearly observed that the Schottky tunneling rate is considerably high at the SE/p+ source interface for SE $TiSi_2$ and $CrSi_2$ and gets vanished for Pd_2Si. Figure 4b simultaneously shows the BTBT at the source-channel interface for different SE metal silicides. A marginal variation in the BTBT rate of electrons is observed for SE metal silicide $TiSi_2$, owing to the reduced electron concentration available for tunneling.

The transfer characteristics of JL-TFET are plotted in Fig. 6a for different SE metal silicide ($TiSi_2$, $CrSi_2$ and Pd_2Si). The results include the impact of UST model on the simulations. The major contribution of the UST model

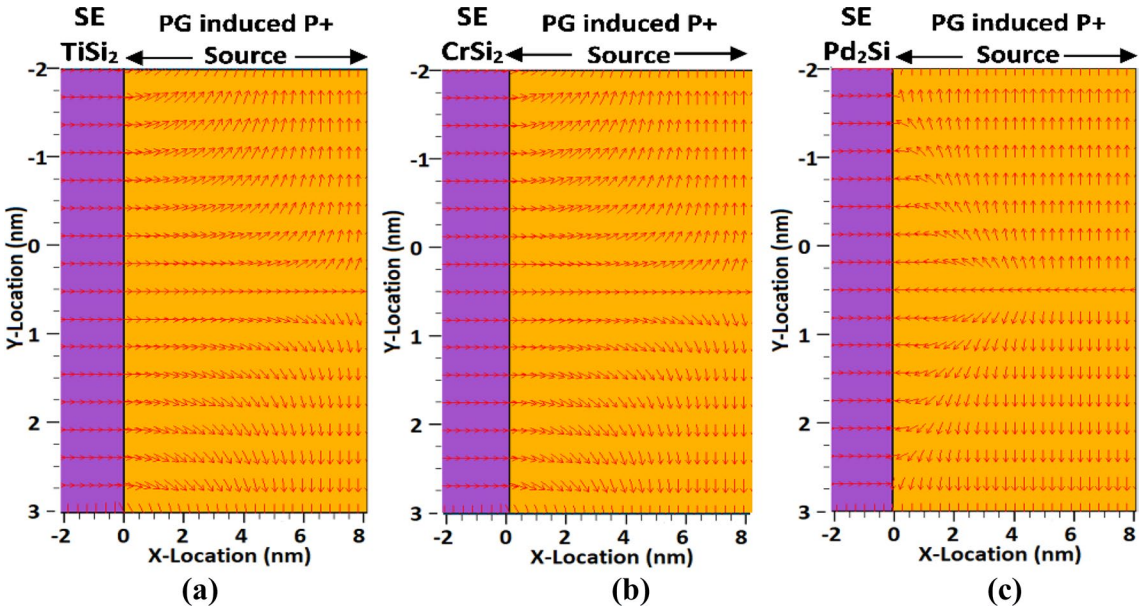


Fig. 3 Vector representation of electric field with SE of **a** $TiSi_2$, **b** $CrSi_2$ and **c** $PdSi_2$ in the ON-state

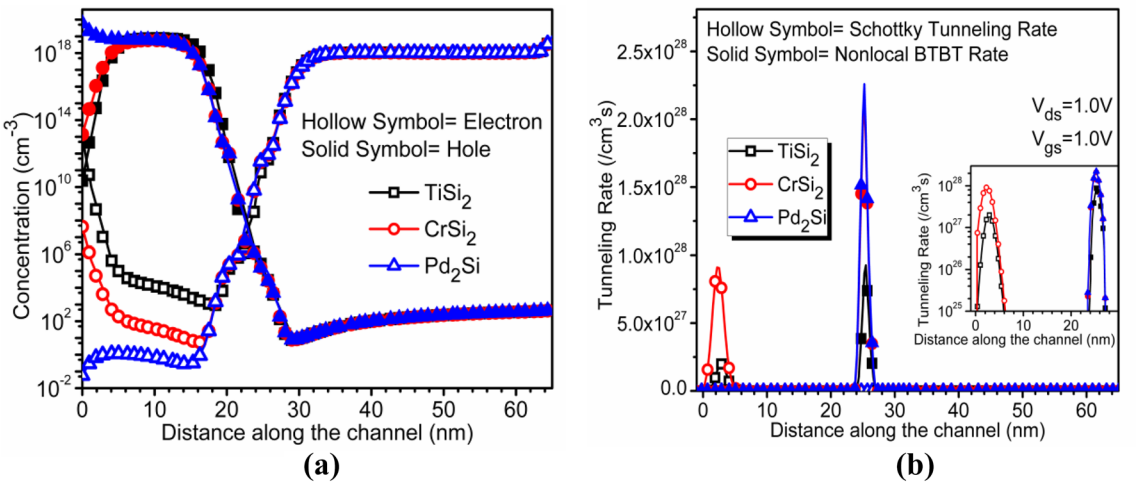


Fig. 4 **a** Influence of SE metal silicide on electron and hole concentration at $V_{gs} = V_{ds} = 1.0$ V. **b** Schottky tunneling rate and band to band tunneling rate along the device length for different SE metal silicide

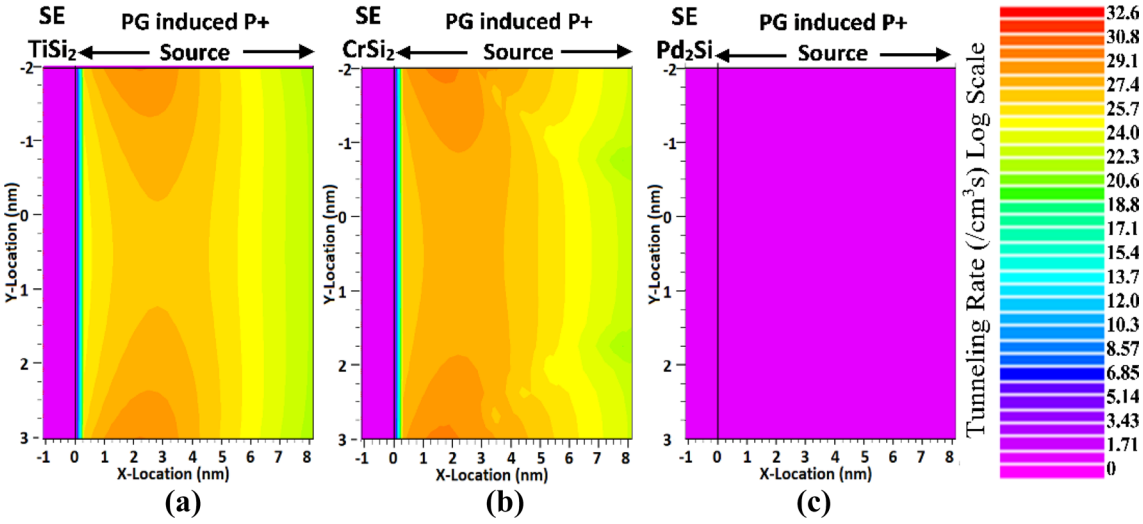


Fig. 5 Contour plot of Schottky tunneling rate **a** $TiSi_2$, **b** $CrSi_2$, **c** Pd_2Si at $V_{gs} = V_{ds} = 1.0$ V

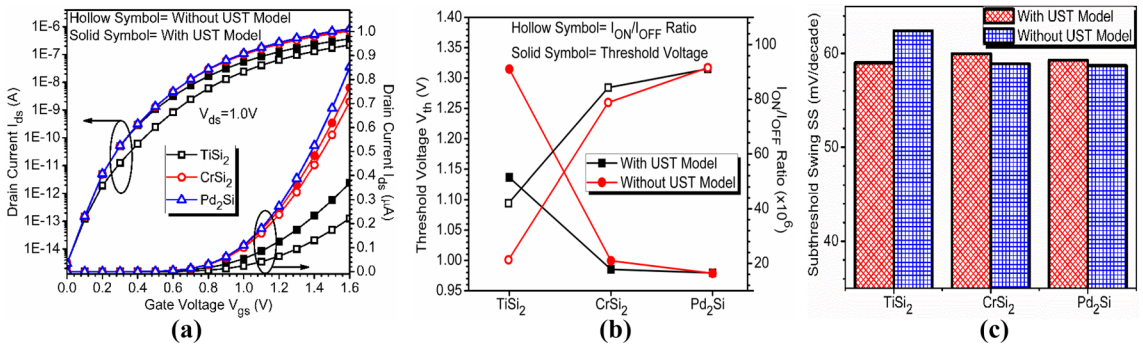


Fig. 6 Influence of SE on **a** drain current log scale and **b** Current switching ratio (I_{ON}/I_{OFF}), V_{th} and Subthreshold swing of JL-TFET

is obtained for the case when a metal depletes the HP at the SE and p+ induced source i.e., for $TiSi_2$ and $CrSi_2$. It is evident from the transfer characteristics that the drain current is maximum for the case when ohmic contact is formed at the SE and induced source interface i.e., for Pd_2Si. Further, for Pd_2Si the drain current is exactly same for with and without the inclusion of UST model. This is due to the fact that the Pd_2Si forms an ohmic contact with the p+ induced source. However, for the case of $TiSi_2$ and $CrSi_2$ as SE, the depleted HP creates a Schottky interface and the UST model includes the Schottky tunneling at the SE/induced source interface and accurately calculate the drain current, which is undervalued otherwise. For $TiSi_2$ and $CrSi_2$, the presence of the Schottky interface refrains the electrons towards the SE and thereby reduces the overall current in the device. Thus, apart from the tunneling at the source channel junction, a very crucial role is being played by the SE metal in the conduction of JL-TFET. It is evaluated at $V_{gs} = 1.2$ V that drain current for $TiSi_2$/$CrSi_2$ is 1.96/1.09 times higher with UST model, owing to the inclusion of Schottky tunneling at the interface of SE/induced source as shown in Fig. 4b. Further, it is observed that the significant contribution of the SE metal is for the ON-state of the device and not in the OFF state. In the OFF state, the formation of large tunneling barrier at the source channel junction blocks the tunneling of electrons, and thereby oversee the effect of SE metal. Thus, it is primarily the tunneling at the source channel junction which contributes to the drain current in the device with obligatory sustenance of retained hole charge plasma by the SE. The consequences of depleted HP by the SE near the SE/induced source region are also analyzed on the current switching ratio, the threshold voltage (V_{th}), and subthreshold swing (SS), shown in Fig. 6b, c. It is evaluated that the depleted HP by SE at the SE/ source interface, reduces the current switching ratio of the JL-TFET. This reduced I_{ON}/I_{OFF} ratio is primarily due to the reduced ON-state current (as depicted in Fig. 6a) without any variation on OFF-state current. Further, it is examined that after the inclusion of UST model, the I_{ON}/I_{OFF} ratio for $TiSi_2$/$CrSi_2$ increases by 1.97/1.06 times owing to the enhancement in the I_{ON}. The V_{th} in this work is evaluated by the constant current method (V_{gs}@ $I_{DS} = 10^{-7}$A). For SE metal silicide with work function greater than that of induced source, the V_{th} reduces by 13.81%. This reduction in V_{th} is attributed to the improvement in the ON current of the device due to absentia of refrainment of electrons near the SE/source interface. Further, it must be noticed that the results are examined for constant control gate metal that results in the same tunneling rate at induced p+ source and channel junction for all the cases. Thus, it is primarily the conductivity offered by the SE metal silicide that is reducing the V_{th}. Further, while comparing the results obtained by UST and without UST model, it is evaluated that for $TiSi_2$ /$CrSi_2$ the V_{th} reduces by 1.15/1.01 times for with UST model w.r.t. without UST model. This is due to the undervalued drain current without UST model in the simulations. Additionally, the SS of the JL-TFET reduces by 5.95% with the formation of ohmic contact at the source electrode. The marginal reduction in SS is achieved with inclusion of UST model as shown in Fig. 6c. However, the device under consideration is the conventional JL-TFET (i.e., with no device/material engineering) thereby the SS is 59 mV/decade. After device/material engineering, the SS can be further reduced. Observance of this kind of variations of threshold voltage and an ability to control the same through selection of different type of SE may have far reaching consequences in the selection of a right device for a particular application.

4 Conclusion

This work investigates a detailed study of the impact of SE metal silicide with different work function and the importance of Universal Schottky tunneling (UST) on the performance of JL-TFET. The study reports the surface potential, electric field, energy bands, conduction current density inside the device with different metal silicides as source electrodes. It is explored that in addition to the tunneling rate at source-channel junction, SE metal also plays a crucial role in determining the overall performance of a JL-TFET. This work reveals that the SE metal silicide with lower work function than the WF of p+ induced source (controlled by the polarity gate), results in the depletion of hole plasma near the SE/source interface that eventually reduces the overall conduction current of the device. Further, the Schottky contact enhances the probability of movement of electrons towards the SE. This undesired movement of electrons towards SE reduces the number of electrons moving towards the channel that contributed to the effective conduction, i.e., the drain current. Thus, the Schottky contact (at the SE and induce p+ source) inhibits the conduction of electrons, even if there is efficient tunneling at the source channel junction. Thus, in assistance with the efficient BTBT controlled by the control gate the formation of an ohmic contact at the SE and the induced source is requisite concurrently. Thus, the metal silicide used to form SE must have work function higher than that of the induced p+ source in the n-type JL-TFET. The higher work function of the source electrode will form the ohmic contact and thereby facilitate the efficient conduction of electrons. Further for numerical simulations, to consider the Schottky tunneling at the SE and the induced p+ source interface UST model must be invoked, though a metal silicide with work function higher than the work function of the p+ induced source should be preferably chosen while fabricating JL-TFET.

Acknowledgements Authors would like to acknowledge the members of VLSI center of excellence, Chitkara University, Punjab, India and Department of Applied Physics, Delhi Technological University, Delhi for continuous support during the preparation of the manuscript.

References

1. Y. Goswami, B. Ghosh, P.K. Asthana, Analog performance of Si junctionless tunnel field effect transistor and its improvisation using III–V semiconductor. RSC Adv. **4**, 10761–10765 (2014)
2. P.K. Asthana, B. Ghosh, Y. Goswami, B.M.M. Tripathi, High-speed and low-power ultradeep-submicrometer III-V hetero-junctionless tunnel field-effect transistor. IEEE Trans. Electron. Devices **61**, 479–486 (2014)
3. B. Ghosh, M.W. Akram, Junctionless tunnel field effect transistor. IEEE Electron. Device Lett. **34**, 584–586 (2013)
4. K. Nigam, P. Kondekar, D. Sharma, B.R. Raad, A new approach for design and investigation of junction-less tunnel FET using electrically doped mechanism. Superlattices Microstruct. **98**, 1–7 (2016)
5. N. Damrongplasit, S.H. Kim, T.-J.K. Liu, Study of random dopant fluctuation induced variability in the raised-Ge-source TFET. IEEE Electron. Device Lett. **34**, 184–186 (2013)
6. S. Tirkey, D.S. Yadav, D. Sharma, Controlling ambipolar current of dopingless tunnel field-effect transistor. Appl. Phys. A **124**, 809 (2018)
7. M. Vadizadeh, Performance estimation of junctionless field effect diode. Appl. Phys. A **125**, 495 (2019)
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. J. Madan, R. Chaujar, Gate drain-overlapped-asymmetric gate dielectric-GAA-TFET: a solution for suppressed ambipolarity and enhanced ON state behavior. Appl. Phys. A **122**, 973 (2016)
10. S. Shekhar, J. Madan, R. Chaujar, Source/gate material-engineered double gate TFET for improved RF and linearity performance: a numerical simulation. Appl. Phys. A **124**, 739 (2018)
11. J. Madan, R. Chaujar, Temperature associated reliability issues of heterogeneous gate dielectric—gate all around—tunnel FET. IEEE Trans. Nanotechnol. **17**, 41–48 (2018)
12. M.S. Parihar, D. Ghosh, G.A. Armstrong, R. Yu, P. Razavi, A. Kranti, Bipolar effects in unipolar junctionless transistors. Appl. Phys. Lett. **101**, 093507 (2012)
13. R.J. Hueting, B. Rajasekharan, C. Salm, J. Schmitz, The charge plasma PN diode. IEEE Electron. Device Lett. **29**, 1367–1369 (2008)
14. S. Gupta, K. Nigam, S. Pandey, D. Sharma, P.N. Kondekar, Effect of interface trap charges on performance variation of heterogeneous gate dielectric junctionless-TFET. IEEE Trans. Electron. Devices **64**, 4731–4737 (2017)
15. S.B. Rahi, B. Ghosh, High-k double gate junctionless tunnel FET with a tunable bandgap. RSC Adv. **5**, 54544–54550 (2015)
16. S. Tirkey, K. Nigam, S. Pandey, D. Sharma, P. Kondekar, Investigation of gate material engineering in junctionless TFET to overcome the trade-off between ambipolarity and RF/linearity metrics. Superlattices Microstruct. **109**, 307–315 (2017)
17. H. Aghandeh, S.A.S. Ziabari, Gate engineered heterostructure junctionless TFET with Gaussian doping profile for ambipolar suppression and electrical performance improvement. Superlattices Microstruct. **111**, 103–114 (2017)
18. P.K. Asthana, Y. Goswami, S. Basak, S.B. Rahi, B. Ghosh, Improved performance of a junctionless tunnel field effect transistor with a Si and SiGe heterostructure for ultra low power applications. RSC Adv. **5**, 48779–48785 (2015)
19. T.J. Drummond, *Work functions of the transition metals and metal silicides* (Albuquerque, Sandia National Labs, 1999)
20. Atlas Users Manual 2010 Silvaco Santa Clara CA
21. A. Biswas, S.S. Dan, C. Le Royer, W. Grabinski, A.M. Ionescu, TCAD simulation of SOI TFETs and calibration of non-local band-to-band tunneling model. Microelectron Eng **98**, 334–337 (2012)
22. J. Madan, R. Chaujar, Gate drain underlapped-PNIN-GAA-TFET for comprehensively upgraded analog/RF performance. Superlattices Microstruct. **102**, 17–26 (2017)
23. K. Matsuzawa, K. Uchida, A. Nishiyama, A unified simulation of Schottky and ohmic contacts. IEEE Trans. Electron. Devices **47**, 103–108 (2000)

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

arXiv:1907.13385v1 [math.CV] 31 Jul 2019

INVERSE COEFFICIENT BOUNDS FOR CERTAIN CLASSES OF CLOSE TO CONVEX FUNCTIONS

PRIYANKA GOEL AND S. SIVAPRASAD KUMAR

Abstract. Let $\mathcal{K}(g)$ be the class of close-to-convex functions with respect to a starlike function g. Let $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ be the subclasses of $\mathcal{K}(g)$, obtained by choosing $g(z)$ as $z/(1-z)$, $z/(1-z^2)$ and $z/(1-z+z^2)$ respectively. In this paper, we obtain sharp upper bounds of first five initial inverse coefficients for functions in each of these classes $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$.

1. Introduction

Let $\mathcal{A}$ denote the class of analytic functions of the form

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n = z + a_2 z^2 + a_3 z^3 + \dots, \tag{1.1}$$

which are analytic in the open unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. Let $\mathcal{S}$ be a subclass of $\mathcal{A}$ consisting of all univalent (one-to-one) functions in $\mathcal{A}$. A function $f \in \mathcal{A}$ is said to be starlike if f maps $\mathbb{D}$ onto a domain which is starlike with respect to origin. The class of starlike functions in $\mathcal{S}$ is denoted by $\mathcal{S}^*$ and is analytically characterized as $f \in \mathcal{S}^*$ if and only if $\mathrm{Re}(zf'(z)/f(z)) > 0$ in $\mathbb{D}$. Similarly, a function $f \in \mathcal{A}$ is said to be convex if it maps $\mathbb{D}$ onto a domain which is convex. The class of convex functions in $\mathcal{S}$ is denoted by $\mathcal{C}$. We say that $f \in \mathcal{C}$ if and only if $\mathrm{Re}((1+zf''(z))/f'(z)) > 0$ in $\mathbb{D}$. Further, Alexander theorem establishes a two way passage between $\mathcal{S}^*$ and $\mathcal{C}$ namely, $f \in \mathcal{C}$ if and only if $zf'(z) \in \mathcal{S}^*$. A function f defined on $\mathbb{D}$ is said to be close to convex with respect to a starlike function g and with argument $\alpha \in (-\pi/2, \pi/2)$, if $\mathrm{Re}(e^{i\alpha} zf'(z)/g(z)) > 0$ and the class of all such functions is denoted by $\mathcal{K}_\alpha(g)$. Let $\widetilde{\mathcal{K}}(g)$ be the class of close to convex functions with respect to g and $\widetilde{\mathcal{K}}_\alpha$ be the class of close to convex functions with argument α and are given by

$$\widetilde{\mathcal{K}}(g) := \bigcup_{\alpha \in (-\pi/2, \pi/2)} \mathcal{K}_\alpha(g) \qquad \text{and} \qquad \widetilde{\mathcal{K}}_\alpha := \bigcup_{g \in \mathcal{S}^*} \mathcal{K}_\alpha(g).$$

The class $\mathcal{K}$ defined as

$$\mathcal{K} := \bigcup_{\alpha \in (-\pi/2, \pi/2)} \widetilde{\mathcal{K}}_\alpha = \bigcup_{g \in \mathcal{S}^*} \widetilde{\mathcal{K}}(g),$$

denote the class of close to convex functions. Let $\mathcal{K}(g) := \mathcal{K}_0(g)$ be the class of close to convex functions with respect to g and argument $\alpha = 0$. Kaplan [2] proved that every close to convex function is univalent in $\mathbb{D}$. In geometrical terms, it means that if $f \in \mathcal{K}$, then the complement of the image of $\mathbb{D}$ under f is the union of non-intersecting half lines. Let $\mathcal{P}$ denote the class of analytic functions of the form

$$p(z) = 1 + \sum_{n=1}^{\infty} c_n z^n = 1 + c_1 z + c_2 z^2 + c_3 z^3 + \cdots \quad (z \in \mathbb{D}), \tag{1.2}$$

2010 *Mathematics Subject Classification.* 30C45,30C50, 30C80.

Key words and phrases. Inverse Coefficients, Close-to-convex functions, Univalent functions, Starlike functions.

The first author is supported by The Council of Scientific and Industrial Research(CSIR). Ref.No.:08/133(0018)/2017-EMR-I..

such that $\operatorname{Re} p(z) > 0$ and this class is known as Caratheódory class. To prove our main results, we need the following results pertaining to the class $\mathcal{P}$:

Lemma 1.1. [7, p.41] *If $p(z)$ is in $\mathcal{P}$ and is given by (1.2), then $|c_n| \leq 2$ for each n.*

Lemma 1.2. [7] *Let $p \in \mathcal{P}$ and is given by (1.2). Then*

$$\left| c_2 - \frac{c_1^2}{2} \right| \leq 2 - \frac{|c_1|^2}{2}.$$

This inequality is sharp for the functions $P_{t,\vartheta}(z)$ given by

$$P_{t,\vartheta}(z) = t\left(\frac{1+e^{i\vartheta}z}{1-e^{i\vartheta}z}\right) + (1-t)\left(\frac{1+e^{i2\vartheta}z^2}{1-e^{i2\vartheta}z^2}\right),$$

$0 \leq t \leq 1$ and $0 \leq \vartheta \leq 2\pi$.

Lemma 1.3. [4] *Let $p \in \mathcal{P}$ and is given by (1.2). Then*

$$\begin{aligned} 2c_2 &= c_1^2 + x(4-c_1^2) \\ 4c_3 &= c_1^3 + 2(4-c_1^2)c_1x - c_1(4-c_1^2)x^2 + 2(4-c_1^2)(1-|x|^2)y, \\ 8c_4 &= c_1^4 + (4-c_1^2)x(c_1^2(x^2-3x+3)+4x) \\ &\quad -4(4-c_1^2)(1-|x|^2)(c_1(x-1)y + \bar{x}y^2 - (1-|y|^2)z). \end{aligned}$$

where $x, y, z \in \mathbb{C}$ such that $|x| \leq 1$, $|y| \leq 1$ and $|z| \leq 1$.

Lemma 1.4. [6] *Let $p \in \mathcal{P}$ and is given by (1.2). Then*

$$|c_3 - 2c_1c_2 + c_1^3| \leq 2$$

and

$$|c_1^4 - 3c_1^2c_2 + c_2^2 + 2c_1c_3 - c_4| \leq 2.$$

This result is sharp for the function $p(z) = (1+z)/(1-z)$.

It may be noted that the function given by

$$H_{t,\beta}(z) := (1-2t)\left(\frac{1+z}{1-z}\right) + t\left(\frac{1+\beta z}{1-\beta z}\right) + t\left(\frac{1+\bar{\beta}z}{1-\bar{\beta}z}\right) \tag{1.3}$$

is a Caratheódary function if $t \in [0, 1/2]$ and $|\beta| = 1$, we make use of it to find the extremal functions needed for sharp bounds. Let the inverse of $f(z)$ defined on a disk of radius atleast $1/4$ has a series expansion of the form

$$f^{-1}(w) = w + \sum_{n=2}^{\infty} \delta_n w^n = w + \delta_2 w^2 + \delta_3 w^3 + \cdots. \tag{1.4}$$

Since $z = f^{-1}(w)$, we have from (1.1)

$$w = f^{-1}(w) + a_2(f^{-1}(w))^2 + a_3(f^{-1}(w))^3 + \cdots. \tag{1.5}$$

Comparing the coefficients on both sides of (1.5) after replacing the value of $f^{-1}(w)$ given in (1.4), we obtain the following relations:

$$\delta_2 = -a_2 \tag{1.6}$$

$$\delta_3 = 2a_2^2 - a_3 \tag{1.7}$$

$$\delta_4 = 5a_2a_3 - 5a_2^3 - a_4 \tag{1.8}$$

$$\delta_5 = 14a_2^4 - 21a_2^2a_3 + 6a_2a_4 + 3a_3^2 - a_5. \tag{1.9}$$

In 1982, Libera and Złotkiewicz [6] derived the bounds for the magnitude of the first seven inverse coefficients δ_k $(k = 2, 3...7)$ of the functions in $\mathcal{C}$. Later in 1984, they estimated the bounds for

first six inverse coefficients of the functions, whose derivative belongs to $\mathcal{P}$ [5]. In 1989, Silverman estimated the bounds of inverse coefficients for starlike functions (see [12]). Similar work has been carried out for various other classes such as class of starlike functions of positive order [3], spirallike functions [14], starlike functions represented by symmetric gap series [13] and strongly starlike functions [1]. In the present paper, we consider three extensively studied ([9], [10], [8]) subclasses of close to convex functions defined as follows:

$$\begin{aligned}\mathcal{F}_1 &:= \{f \in \mathcal{A} : \mathrm{Re}(1-z)f'(z) > 0 \text{ for } z \in \mathbb{D}\}\\ \mathcal{F}_2 &:= \{f \in \mathcal{A} : \mathrm{Re}(1-z^2)f'(z) > 0 \text{ for } z \in \mathbb{D}\}\\ \mathcal{F}_3 &:= \{f \in \mathcal{A} : \mathrm{Re}(1-z+z^2)f'(z) > 0 \text{ for } z \in \mathbb{D}\}.\end{aligned}$$

It is pertinent to mention here that Ponnusamy et al. discussed the region of variability for these three classes in [9] and [10]. Further Ponnusamy et al. [8] generalized the classes $\mathcal{F}_1$ and $\mathcal{F}_2$ to a class of harmonic close to convex functions defined on $\mathbb{D}$. In [11], Kumar and Vasudevarao estimated the bounds of first three logarithmic coefficients of functions belonging to these three classes. However, in the present investigation, we try to find sharp upper bounds of the first five consecutive inverse coefficients for functions in each of these classes $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$. Although the fifth inverse coefficient bound obtained here was not sharp for functions in $\mathcal{F}_2$, but the range in which the sharp bound lies is also pointed out.

2. Main Results

Let $f(z)$ be given by (1.1) and belongs to $\widetilde{\mathcal{K}}_0$. Then we have

$$\mathrm{Re}\left(\frac{zf'(z)}{g(z)}\right) > 0, \tag{2.1}$$

where g is a starlike function defined as

$$g(z) = z + \sum_{n=2}^{\infty} b_n z^n. \tag{2.2}$$

In view of (2.1), there exists $p \in \mathcal{P}$, with power series representation given by (1.2), such that

$$zf'(z) = g(z)p(z), \tag{2.3}$$

which implies

$$z + \sum_{n=2}^{\infty} n a_n z^n = \left(z + \sum_{n=2}^{\infty} b_n z^n\right)\left(1 + \sum_{n=1}^{\infty} c_n z^n\right). \tag{2.4}$$

Upon equating the like term coefficients on either side of (2.4), we get

$$2a_2 = b_2 + c_1 \tag{2.5}$$

$$3a_3 = b_3 + b_2c_1 + c_2 \tag{2.6}$$

$$4a_4 = b_4 + c_1b_3 + c_2b_2 + c_3 \tag{2.7}$$

$$5a_5 = b_5 + b_4c_1 + b_3c_2 + b_2c_3 + c_4. \tag{2.8}$$

Using (2.5)-(2.8) in (1.6)-(1.9), we get the following δ_i's in terms of b_i's and c_i's:

$$\begin{aligned}
\delta_2 &= -\frac{1}{2}(b_2+c_1) && (2.9)\\
\delta_3 &= \frac{1}{6}(3b_2^2+3c_1^2+4b_2c_1-2b_3-2c_2) && (2.10)\\
\delta_4 &= \frac{1}{24}(20b_2b_3-25b_2^2c_1+14b_2c_2+14c_1b_3-25b_2c_1^2+20c_1c_2\\
&\quad -15b_2^3-15c_1^3-6b_4-6c_3) && (2.11)\\
\delta_5 &= \frac{7}{8}b_2^4-\frac{7}{4}b_2^2b_3+\frac{1}{3}b_3^2+\frac{3}{4}b_4b_2-\frac{1}{5}b_5+\frac{7}{4}b_2^3c_1-\frac{25}{12}b_2b_3c_1\\
&\quad +\frac{11}{20}b_4c_1+\frac{25}{12}b_2^2c_1^2-b_3c_1^2+\frac{7}{4}b_2c_1^3+\frac{7}{8}c_1^4-b_2^2c_2+\frac{7}{15}b_3c_2\\
&\quad -\frac{25}{12}b_2c_1c_2-\frac{7}{4}c_1^2c_2+\frac{1}{3}c_2^2+\frac{11}{20}b_2c_3+\frac{3}{4}c_1c_3-\frac{1}{5}c_4. && (2.12)
\end{aligned}$$

Applying triangle inequality in (2.9), we obtain

$$2|\delta_2|\leq |b_2|+|c_1|. \tag{2.13}$$

In a similar way, using triangle inequality in (2.10) and applying Lemma.1.2, we get

$$3|\delta_3|\leq 2-\frac{|c_1|^2}{2}+\left|(b_3-\frac{1}{2}b_2^2)-(c_1+b_2)^2\right|. \tag{2.14}$$

Let $c_1=pe^{i\alpha}$ and $q=\cos\alpha$ such that $0\leq p\leq 2$ and $0\leq\alpha\leq 2\pi$. Now, we rewrite (2.14) in terms of p and q as follows:

$$3|\delta_3|\leq 2-\frac{p^2}{2}+\left|(b_3-\frac{1}{2}b_2^2)-(pq+ip\sqrt{1-q^2}+b_2)^2\right|. \tag{2.15}$$

An application of Lemma.1.3 to (2.11), yields

$$\begin{aligned}
48|\delta_4|=&|6y(1-|x|^2)(4-c_1^2)+13c_1^3+(50b_2^2-28b_3)c_1+(30b_2^3+12b_4\\
&-40b_2b_3)-x(4-c_1^2)(14b_2+14c_1+3c_1x)+36b_2c_1^2|.
\end{aligned} \tag{2.16}$$

Assume that b_i's are all real and replace c_1 by c, where $0\leq c\leq 2$. Further, let $x=re^{i\theta}$ with $d=\cos\theta$ such that $0\leq r\leq 1$ and $0\leq\theta\leq 2\pi$. Now, applying triangle inequality in (2.16), we obtain

$$48|\delta_4|\leq 6(1-r^2)(4-c^2)+|\psi(c,r,d)|, \tag{2.17}$$

where

$$\begin{aligned}
\psi(c,r,d)=&13c^3+(50b_2^2-28b_3)c+(30b_2^3+12b_4-40b_2b_3)+36b_2c^2\\
&-r(d+i\sqrt{1-d^2})(4-c^2)(14b_2+14c+3cr(d+i\sqrt{1-d^2})).
\end{aligned}$$

Theorem 2.1. *Let $f(z)=z+a_2z^2+a_3z^3+\ldots\in\mathcal{F}_1$. Then*

(*i*) $|\delta_2|\leq 3/2$
(*ii*) $|\delta_3|\leq 17/6$
(*iii*) $|\delta_4|\leq 49/8$
(*iv*) $|\delta_5|\leq 1729/120$.

These bounds are sharp.

Proof. (i) Let $f \in \mathcal{F}_1$. Since f is close to convex function with respect to the starlike function $z/(1-z)$, from (2.3) we have

$$zf'(z) = \frac{z}{1-z}p(z). \tag{2.18}$$

For $g(z) = z/(1-z)$, we have $b_i = 1$ for all i. Thus in view of Lemma 1.1, (2.13) reduces to

$$2|\delta_2| \leq 1 + |c_1| \leq 3. \tag{2.19}$$

We know that this inequality is sharp, whenever $|c_1| = 2$ which is true for the function $P_{1,\vartheta}(z)$ $(0 \leq \vartheta \leq 2\pi)$ given in Lemma 1.2. The inequality $|\delta_2| \leq 3/2$ is sharp since there exists an extremal function $\hat{f}_1 \in \mathcal{F}_1$, which is a solution of $z\hat{f}'_1(z) = z(1-z)^{-1}P_{1,\vartheta}(z)$.

(ii) By taking $b_2 = b_3 = 1$ in (2.15), we get

$$3|\delta_3| \leq 2 - \frac{p^2}{2} + \left|\frac{1}{2} - (1 + pq + ip\sqrt{1-q^2})^2\right|,$$

which can be written as

$$3|\delta_3| \leq 2 - \frac{p^2}{2} + \sqrt{(p^2 + \frac{1}{2} + 2pq)^2 + 2p^2(1-q^2)} =: \phi_1(p,q). \tag{2.20}$$

In the domain $D = \{(p,q) : 0 \leq p \leq 2, -1 \leq q \leq 1\}$, we need to find the points where $\phi_1(p,q)$ attains its maximum. A simple computation shows that there exists no solution of

$$\frac{\partial \phi_1(p,q)}{\partial p} = 0 \quad \text{and} \quad \frac{\partial \phi_1(p,q)}{\partial q} = 0$$

in $\mathbb{R} \times \mathbb{R}$. Thus, the maximum value of $\phi_1(p,q)$ is not attained inside D. Now, we consider the edges of D in order to find the maximum of $\phi_1(p,q)$. On the line segment $p = 0$, $\phi_1(0,q) = 5/2$. On the line segment $p = 2$, $\phi_1(2,q) = \sqrt{8q^2 + 36q + 113/4}$, which is an increasing function for $q \in [-1,1]$. Thus $\max_{q\in[-1,1]} \phi_1(2,q) = \phi_1(2,1) = 17/2$. On the line segment $q = -1$, $\phi_1(p,-1) = 2 - p^2/2 + |1/2 - 2p + p^2|$ which further reduces into two cases. By using elementary calculus, we obtain that $\max_{p\in[0,2]} \phi_1(p,-1) = \phi_1(0,-1) = 5/2$. On the line segment $q = 1$, $\phi_1(p,1) = 5/2 + p^2/2 + 2p$, which is an increasing function for $p \in [0,2]$. Thus $\max_{p\in[0,2]} \phi_1(p,1) = \phi_1(2,1) = 17/2$. Hence the maximum value of $\phi_1(p,q)$ is attained at $(2,1)$ and is equal to $17/2$. Thus $|\delta_3| \leq 17/6$. The inequality $|\delta_3| \leq 17/6$ is sharp since there exists an extremal function $\tilde{f}_1 \in \mathcal{F}_1$, which is a solution of $z\tilde{f}'_1(z) = z(1-z)^{-1}P_{1,0}(z)$.

(iii) Upon substituting $b_2 = b_3 = b_4 = 1$ in (2.17), we get

$$48|\delta_4| \leq 6(4-c^2)(1-r^2) + \sqrt{\psi_1(c,r,d)} =: Q(c,r,d), \tag{2.21}$$

where

$$\begin{aligned}\psi_1(c,r,d) =& (13c^3 + 36c^2 + 22c + 2)^2 + (4-c^2)^2r^2(84c^2dr + 9c^2r^2 \\ &+ 196c^2 + 84crd + 392c + 196) \\ &- 2(4-c^2)r(13c^3 + 36c^2 + 22c + 2)(-3cr + 6cd^2r + 14cd + 14d).\end{aligned}$$

It is sufficient to find the points in the rectangular cube $K := \{(c,r,d) : 0 \leq c \leq 2, 0 \leq r \leq 1 \text{ and } -1 \leq d \leq 1\}$, where the maximum value of $Q(c,r,d)$ is attained. In order to find maximum in the interior of K, we try to find the points where

$$\frac{\partial Q(c,r,d)}{\partial c} = \frac{\partial Q(c,r,d)}{\partial r} = \frac{\partial Q(c,r,d)}{\partial d} = 0 \tag{2.22}$$

It is easy to note that $\partial Q(c,r,d)/\partial d = 0$ implies

$$d = \frac{2 + 22c + 36c^2 - 182c^3 - 182c^4 + 168cr^2 - 42c^3r^2 + 168c^2r^2 - 42c^4r^2}{12c(2 + 22c + 36c^2 + 13c^3)r}.$$

Upon substituting the above value of d in $\partial Q(c,r,d)/\partial r = 0$, we arrive at no real valued solution of the equations stated in (2.22). That is, there exist no solution inside the cuboid K. Now, we find the maximum of $Q(c,r,d)$ on the six faces of K. On the rectangular face $c = 0$, we have

$$Q(0,r,d) = 24(1 - r^2) + \sqrt{4 - 224d + 3136r^2}.$$

A simple calculation shows that $(0,0)$ is the only solution of the system

$$\frac{\partial Q(0,r,d)}{\partial r} = 0 \quad \text{and} \quad \frac{\partial Q(0,r,d)}{\partial d} = 0.$$

Thus, we can say that the maximum value is not attained in the interior of the face $c = 0$. Now we consider the face $c = 2$. Then we have $Q(2,r,d) = 294$ and clearly

$$\max_{r\in(0,1),d\in(-1,1)} Q(2,r,d) = 294.$$

On the square face $r = 0$, we have $Q(c,0,d) = 13c^3 + 30c^2 + 22c + 26$ which is clearly an increasing function of c, where $c \in (0,2)$. Thus, we have

$$\max_{c\in(0,2),d\in(-1,1)} Q(c,0,d) = Q(2,0,d) = 294.$$

On the square face $r = 1$, we have

$$Q(c,1,d) = \sqrt{\rho_1(c,d) + 2\left(13c^3 + 36c^2 + 22c + 2\right)\left(c^2 - 4\right)\left(6cd^2 + 14cd - 3c + 14d\right)},$$

where

$$\rho_1(c,d) = \left(13c^3 + 36c^2 + 22c + 2\right)^2 + \left(4 - c^2\right)^2\left(84c^2d + 205c^2 + 84cd + 392c + 196\right).$$

It is easy to verify that $\partial Q(c,1,d)/\partial d = 0$ implies

$$d = -\frac{56(c+1)\left(8c^3 + 18c^2 + 5c + 1\right)}{24c\left(13c^3 + 36c^2 + 22c + 2\right)}. \tag{2.23}$$

A simple calculation reveals that $\partial Q(c,1,d)/\partial c = 0$ implies

$$\begin{aligned}&\left(1776c^5 + 5560c^4 + 2416c^3 - 1944c^2 + 5736c + 6408\right)\\ &+ \left(2688c^5 + 7280c^4 - 2016c^3 - 16464c^2 - 10192c - 1344\right)d\\ &+ \left(936c^5 + 2160c^4 - 1440c^3 - 5112c^2 - 2112c - 96\right)d^2 = 0.\end{aligned} \tag{2.24}$$

Upon substituting (2.23) in (2.24) and simplifying, we have

$$\frac{-8(5c^3 + 18c^2 + 17c + 1)\xi(c)}{3c^2(13c^3 + 36c^2 + 22c + 2)^2} = 0, \tag{2.25}$$

where

$$\begin{aligned}\xi(c) =& 1950c^{10} + 9355c^9 - 13790c^8 - 180902c^7 - 487347c^6 - 620475c^5\\ &- 399657c^4 - 117081c^3 - 9367c^2 + 980c + 196.\end{aligned}$$

The numerical solution of (2.25) in the range $(0,2)$ is $c \approx 0.105407$. Substituting this value in (2.23), we obtain $d \approx -8.97501$, but $d \in (-1,1)$. Therefore, there exists no solution of

$$\frac{\partial Q(c,1,d)}{\partial d} = 0 \qquad \text{and} \qquad \frac{\partial Q(c,1,d)}{\partial c} = 0$$

in the square face $r = 1$, that is $\{(c, d) : 0 < c < 2, -1 < d < 1\}$. Now we shall consider the rectangular face $d = -1$, then we have

$$\begin{aligned} Q(c, r, -1) =& 3c^3r^2 - 14c^3r + 13c^3 + 6c^2r^2 - 14c^2r + 30c^2 - 12cr^2 + 56cr \\ &+ 22c - 24r^2 + 56r + 26. \end{aligned} \tag{2.26}$$

A computation shows that $\partial Q(c, r, -1)/\partial r = 0$ implies

$$r = \frac{7(c+1)}{3(c+2)}. \tag{2.27}$$

It is easy to see that $\partial Q(c, r, -1)/\partial c = 0$ gives

$$9c^2r^2 - 42c^2r + 39c^2 + 12cr^2 - 28cr + 60c - 12r^2 + 56r + 22 = 0. \tag{2.28}$$

Upon substituting (2.27) in (2.28), we get $10c^2 - 60c - 71 = 0$. Clearly, this equation has no solution in the range $(0, 2)$. Finally considering the rectangular face $d = 1$, we have

$$\begin{aligned} Q(c, r, 1) =& 6c^2r^2 - 6c^2 - 24r^2 + 24 + |3c^3r^2 + 14c^3r + 13c^3 + 14c^2r \\ &+ 36c^2 - 12cr^2 - 56cr + 22c - 56r + 2|. \end{aligned} \tag{2.29}$$

If the expression in the modulus in (2.29) is positive, then $\partial Q(c, r, 1)/\partial r = 0$ implies

$$r = -\frac{7(c+1)}{3(c+2)}, \tag{2.30}$$

which in turn confines the value of r to be negative, whenever $c \in (0, 2)$ but $0 < r < 1$. Therefore, there exist no solution for this case. If the expression in the modulus of the equation (2.29) is negative, then $\partial Q(c, r, 1)/\partial r = 0$ implies

$$r = \frac{7(c+1)}{3(2-c)}. \tag{2.31}$$

On substituting (2.31) in $\partial Q(c, r, 1)/\partial c = 0$, we get $(179 + 140c + 30c^2)/3 = 0$. Obviously, this equation has no real solution. Thus in either case, there exists no solution of the system of equations

$$\frac{\partial Q(c, r, 1)}{\partial r} = 0 \qquad \text{and} \qquad \frac{\partial Q(c, r, 1)}{\partial c} = 0.$$

Now we are left with the edges of K to find where $Q(c, r, d)$ attains its maximum value. On the line segment $c = 0; r = 0$, we have $Q(0, 0, d) = 26$ and clearly $\max_{d\in[-1,1]} Q(0, 0, d) = 26$. On the line segment $c = 0; r = 1$, we have $Q(0, 1, d) = \sqrt{3140 - 224d}$, which is clearly a decreasing function of d and therefore, $\max_{d\in[-1,1]} Q(0, 1, d) = Q(0, 1, -1) = 58$. On the line segment $c = 0; d = -1$, we obtain $Q(0, r, -1) = 26 - 24r^2 + 56r$, which is an increasing function of r and therefore, $\max_{r\in[0,1]} Q(0, r, -1) = Q(0, 1, -1) = 58$. On the line segment $c = 0; d = 1$, we observe that $Q(0, r, 1) = 24(1 - r^2) + 2|28r - 1|$. A computation shows that $Q(0, r, 1)$ attains its maximum at $r = 1$ and thus $\max_{r\in[0,1]} Q(0, r, 1) = Q(0, 1, 1) = 58$. Also, we know that $Q(2, r, d) = 294$, which implies

$$\max_{d\in[-1,1]} Q(2, 0, d) = \max_{d\in[-1,1]} Q(2, 1, d) = \max_{r\in[0,1]} Q(2, r, \pm 1) = 294.$$

On the line segments $r = 0; d = \pm 1$, we have $Q(c, 0, -1) = Q(c, 0, 1) = 13c^3 + 30c^2 + 22c + 26$ which is clearly an increasing function of c and thus

$$\max_{c\in[0,2]} Q(c, 0, -1) = \max_{c\in[0,2]} Q(c, 0, 1) = 294.$$

On the line segment $r = 1; d = -1$, we obtain $Q(c, 1, -1) = 2c^3 + 22c^2 + 66c + 58$ and it is easy to verify that $\max_{c\in[0,2]} Q(c, 1, -1) = Q(2, 1, -1) = 294$. Finally on the line segment $r = 1; d = 1$, we have $Q(c, 1, 1) = |D|$, where $D = 30c^3 + 50c^2 - 46c - 54$. If $D > 0$ then $Q(c, 1, 1)$ is an increasing

function of c and thus $\max_{c\in[0,2]} Q(c,1,1) = Q(2,1,1) = 294$. If $D < 0$ then $Q(c,1,1)$ attains its maximum at $\left(2\sqrt{415}-25\right)/45$ and $Q(\left(2\sqrt{415}-25\right)/45,1,1) \approx 62.6888$. As $Q(c,1,1)$ is uniformly continuous on $[0,2]\times\{1\}\times\{1\}$, we have $|Q(\left(2\sqrt{415}-25\right)/45,1,1) - 62.6888| < \epsilon << 1$. When ϵ approaches to 0, we have $\max_{c\in[0,2]} Q(c,1,1) = 62.6888$. In view of all the above cases, we have

$$\max_{(c,r,d)\in K} Q(c,r,d) = 294. \tag{2.32}$$

Thus we obtain $|\delta_4| \leq 49/8$. The function $\tilde{k}_1 \in \mathcal{F}_1$, obtained by solving (2.18) with $P_{1,0}(z)$ in place of $p(z)$ acts as an extremal function for the inequality $|\delta_4| \leq 49/8$ and hence is sharp.

(iv) Upon substituting of $b_2 = b_3 = b_4 = b_5 = 1$ in (2.12), we get

$$\begin{aligned}\delta_5 =& \frac{1}{120} + \frac{13}{60}c_1 + \frac{13}{12}c_1^2 + \frac{7}{4}c_1^3 + \frac{7}{8}c_1^4 - \frac{8}{15}c_2 - \frac{25}{12}c_1c_2 - \frac{7}{4}c_1^2c_2 \\ &+ \frac{1}{3}c_2^2 + \frac{11}{20}c_3 + \frac{3}{4}c_1c_3 - \frac{1}{5}c_4,\end{aligned} \tag{2.33}$$

which can be written as

$$\delta_5 = \frac{1}{5}A + \frac{11}{20}B + \varrho(c_1,c_2,c_3) + \varsigma(c_1,c_2), \tag{2.34}$$

where

$$\begin{aligned}A &= c_1^4 - 3c_1^2c_2 + c_2^2 + 2c_1c_3 - c_4 \\ B &= c_3 - 2c_1c_2 + c_1^3 \\ \varrho(c_1,c_2,c_3) &= \frac{1}{120} + \frac{13}{60}c_1 + \frac{2}{15}c_2^2 + \frac{7}{20}c_1c_3 \qquad \text{and} \\ \varsigma(c_1,c_2) &= \frac{8}{15}\left(\frac{65}{32}c_1^2 - c_2\right) + \frac{59}{60}c_1\left(\frac{72}{59}c_1^2 - c_2\right) + \frac{23}{20}c_1^2\left(\frac{27}{46}c_1^2 - c_2\right).\end{aligned}$$

In view of Lemma 1.4, we have $|A| \leq 2$ and $|B| \leq 2$. Evidently $\varrho(c_1,c_2,c_3)$ attains its maximum at $(2,2,2)$. Thus, $|\varrho(c_1,c_2,c_3)| \leq 19/8$. Now, let us consider $\varsigma(c_1,c_2)$ which can be again written in terms of c_1 and x by using Lemma 1.3 as follows:

$$\varsigma_1(c_1,x) = \frac{49}{60}c_1^2 + \frac{17}{24}c_1^3 + \frac{1}{10}c_1^4 - x(4-c_1^2)\left(\frac{8}{30} + \frac{59}{120}c_1 + \frac{23}{40}c_1^2\right). \tag{2.35}$$

Let us write $c_1 = c$ $(0 \leq c \leq 2)$ and $x = re^{i\theta}$, with $d = \cos\theta$ such that $0 \leq r \leq 1$ and $0 \leq \theta \leq 2\pi$. Then (2.35) reduces to:

$$\varsigma_2(c,r,d) = \frac{49}{60}c^2 + \frac{17}{24}c^3 + \frac{1}{10}c^4 - (4-c^2)\left(\frac{8}{30} + \frac{59}{120}c + \frac{23}{40}c^2\right)(rd + ir\sqrt{1-d^2}). \tag{2.36}$$

Now let us assume

$$|\varsigma_2(c,r,d)| = \sqrt{(\operatorname{Re}\varsigma_2(c,r,d))^2 + (\operatorname{Im}\varsigma_2(c,r,d))^2} =: Q_1(c,r,d), \tag{2.37}$$

where

$$\operatorname{Re}\varsigma_2(c,r,d) = \frac{49}{60}c^2 + \frac{17}{24}c^3 + \frac{1}{10}c^4 - (4-c^2)rd\left(\frac{8}{30} + \frac{59}{120}c + \frac{23}{40}c^2\right)$$

and

$$\operatorname{Im}\varsigma_2(c,r,d) = (4-c^2)r\sqrt{(1-d^2)}\left(\frac{8}{30} + \frac{59}{120}c + \frac{23}{40}c^2\right).$$

In order to maximize $Q_1(c,r,d)$, we use the similar method as used in part(iii) and observe that

$$\max_{c\in[0,2],r\in[0,1],d\in[-1,1]} Q_1(c,r,d) = Q_1(2,r,d) = \frac{158}{15}.$$

Now applying triangle inequality on (2.34), we have

$$\begin{aligned} |\delta_5| &\leq \frac{1}{5}|A|+\frac{11}{20}|B|+|\varrho(c_1,c_2,c_3)|+|\varsigma(c_1,c_2)| \\ &\leq \frac{1}{5}(2)+\frac{11}{20}(2)+\frac{19}{8}+\frac{158}{15} \\ &= \frac{1729}{120}. \end{aligned}$$

The above inequality is sharp since there exists an extremal function $\hat{k}_1 \in \mathcal{F}_1$, which is a solution of $z\hat{k}'_1(z) = z(1-z)^{-1}P_{1,0}(z)$. ∎

Theorem 2.2. *Let $f(z) = z + a_2z^2 + a_3z^3 + \ldots \in \mathcal{F}_2$. Then*

(*i*) $|\delta_2| \leq 1$
(*ii*) $|\delta_3| \leq 1$
(*iii*) $|\delta_4| \leq 16/3\sqrt{15}$
(*iv*) $|\delta_5| \leq 2.947584$.

Except (iv) rest all above bounds are sharp.

Proof. (i) Let $f \in \mathcal{F}_2$. Since f is close to convex function with respect to the starlike function $z/(1-z^2)$, we have

$$zf'(z) = \frac{z}{1-z^2}p(z). \tag{2.38}$$

If $g(z) = z/(1-z^2)$, then we have $b_2 = 0$, $b_3 = 1$, $b_4 = 0$ and $b_5 = 1$. Thus, in view of Lemma 1.1, (2.13) reduces to

$$|\delta_2| \leq \frac{|c_1|}{2} \leq 1. \tag{2.39}$$

We know that this inequality is sharp, whenever $|c_1| = 2$ which is true for the function $P_{1,\vartheta}(z)$ $(0 \leq \vartheta < 2\pi)$ given in Lemma 1.2. The upper bound of $|\delta_2|$ is sharp since there exists an extremal function $\hat{f}_2 \in \mathcal{F}_2$, which is the solution of $z\hat{f}'_2(z) = z(1-z^2)^{-1}P_{1,\vartheta}(z)$ $(0 \leq \vartheta < 2\pi)$.

(ii) By taking $b_2 = 0$ and $b_3 = 1$ in (2.10), we get

$$\delta_3 = \frac{1}{6}(3c_1^2 - 2 - 2c_2).$$

Now using Lemma 1.3 and applying the triangle inequality, we obtain

$$|\delta_3| \leq \frac{1}{6}(|2c_1^2 - 2| + |x|(4 - c_1^2)).$$

For the sake of convenience, we shall take $c_1 = c$ and $|x| = r$. Now we have to find the maximum of $\phi_2(c,r) := |2c^2 - 2| + r(4 - c^2)$, whenever $0 \leq c \leq 2$ and $0 \leq r \leq 1$. If $0 \leq c < 1$, then $\phi_2(c,r) = 2 - 2c^2 + r(4 - c^2)$ attains its maximum value 6 at $(0,1)$. Further, if $1 < c \leq 2$, then $\phi_2(c,r) = 2c^2 - 2 + r(4 - c^2)$ attains its maximum value 6 at $(2,r)$. Clearly at $c = 1$, we have $\phi_2(1,r) = 3r$, which obviously attains its maximum value 3 at $r = 1$. Hence $\max_{r\in[0,1],c\in[0,2]} \phi_2(c,r) = 6$ and therefore $|\delta_3| \leq 1$. The upper bound of $|\delta_3|$ is sharp since there exists an extremal function $\tilde{f}_2 \in \mathcal{F}_2$, which is the solution of $z\tilde{f}'_2(z) = (z/(1-z^2))P_{0,0}(z)$.

(iii) Upon substituting $b_2 = 0, b_3 = 1$ and $b_4 = 0$ in (2.17), we get

$$48|\delta_4| \leq 6(4 - c^2)(1 - r^2) + c\sqrt{\psi_2(c,r,d)} =: G(c,r,d), \tag{2.40}$$

where

$$\begin{aligned}\psi_2(c,r,d) =&(13c^2-28)^2+(4-c^2)^2r^2(9r^2+84rd+196)\\ &-2(4-c^2)r(13c^2-28)(14d-3r+6rd^2).\end{aligned}$$

Now, it is enough to find the points in the rectangular cube $K := \{(c,r,d) : 0 \le c \le 2, 0 \le r \le 1 \text{ and } -1 \le d \le 1\}$, where the maximum of $G(c,r,d)$ is attained. So we shall first find the critical points by solving the following equations:

$$\frac{\partial G(c,r,d)}{\partial c} = \frac{\partial G(c,r,d)}{\partial r} = \frac{\partial G(c,r,d)}{\partial d} = 0. \tag{2.41}$$

It is easy to note that $\partial G(c,r,d)/\partial d = 0$ implies

$$d = \frac{168r^2 - 42r^2c^2 - 182c^2 + 392}{12(13c^2-28)r}.$$

Upon substituting the above value of d in $\partial G(c,r,d)/\partial r = 0$, we arrive at no real valued solution of the equations stated in (2.41). That is, there exist no solution inside the cuboid K. Now, we proceed to find the maximum of $G(c,r,d)$ on the six faces of K. On the rectangular face $c = 0$, we have $G(0,r,d) = 24(1-r^2)$, which decreases as r increases from 0 to 1. Thus $\max_{r\in(0,1),d\in(-1,1)} G(0,r,d) = 24$. Now we consider the face $c = 2$. Then we have $G(2,r,d) = 48$ and clearly

$$\max_{r\in(0,1),d\in(-1,1)} G(2,r,d) = 48.$$

On the square face $r = 0$, we have $G(c,0,d) = 24 - 6c^2 + c|13c^2 - 28|$. When $c \in (0, \sqrt{28/13})$, we have $G(c,0,d) = 24 - 6c^2 - 13c^3 + 28c$ which is independent of d. A simple calculation shows that $G(c,0,d)$ attains its maximum at $c = 2\left(\sqrt{282}-3\right)/39$ and $G(2\left(\sqrt{282}-3\right)/39, 0, d) = 16\left(94\sqrt{282}+1863\right)/1521 \approx 36.2028$. When $c \in (\sqrt{28/13}, 2)$, we observe that $G(c,0,d) = 24 - 6c^2 + 13c^3 - 28c$, which is an increasing function of c $(0 < c < 2)$. Clearly, the maximum value 48 of $G(c,0,d)$ is attained at $c = 2$. At $c = \sqrt{28/13}$, we have $G(c,0,d) = G(\sqrt{28/13}, 0, d) = 144/13 \approx 11.0769$. Thus, we have

$$\max_{c\in(0,2),d\in(-1,1)} G(c,0,d) = G(2,0,d) = 48.$$

On the square face $r = 1$, we have

$$G(c,1,d) = c\sqrt{\rho_2(c,d) + 2(13c^2-28)(c^2-4)(14d+6d^2-3)},$$

where

$$\rho_2(c,d) = \left(13c^2-28\right)^2 + (4-c^2)^2(9+84d+196).$$

It is easy to verify that $\partial G(c,1,d)/\partial d = 0$ implies

$$d = -\frac{-56c^2+140}{39c^2-84}. \tag{2.42}$$

Upon simplifying $\partial G(c,1,d)/\partial c = 0$, yields

$$3c^4\left(39d^2+112d+74\right) - 16c^2\left(30d^2+91d+59\right) + 16\left(21d^2+70d+53\right) = 0. \tag{2.43}$$

Upon substituting (2.42) in (2.43) and simplifying, we have

$$4875c^8 - 60800c^6 + 213600c^4 - 268800c^2 + 100352 = 0. \tag{2.44}$$

The numerical solution of (2.44) in the range $(0,2)$, is $c \approx 0.80673$ and $c \approx 1.60961$. Substituting these values in (2.42), we obtain $d \approx -1.76659$ and $d \approx -0.298498$ respectively. Since $d \in (-1,1)$, the only solution of

$$\frac{\partial G(c,1,d)}{\partial d} = 0 \qquad \text{and} \qquad \frac{\partial G(c,1,d)}{\partial c} = 0$$

in the square $r = 1$, is $(c, d) = (1.60961, -0.298498)$ and the value of G is $G(1.60961, 1, -0.298498) \approx 35.8455$. As $G(c, 1, d)$ is uniformly continuous on $[0, 2]\times\{1\}\times[-1, 1]$, we have $|G(1.60961, 1, -0.298498) - 35.8455| < \epsilon << 1$. When ϵ approaches to 0, we have $\max_{c\in(0,2),d\in(-1,1)} G(c, 1, d) = 35.8455$. Now we shall consider the rectangular face $d = -1$, then we have

$$G(c, r, -1) = 24 - 6c^2 - 24r^2 + 6c^2r^2 + c|(13c^2 - 28) - r(4 - c^2)(3r - 14)|. \tag{2.45}$$

If the expression in the modulus of the equation (2.45) is positive, we have

$$G(c, r, -1) = 24 - 6c^2 - 24r^2 + 6c^2r^2 + 13c^3 - 28c - 12r^2c + 56rc + 3r^2c^3 - 14c^3r.$$

A computation shows that $\partial G(c, r, -1)/\partial r = 0$ implies

$$r = \frac{7c}{3(c + 2)}. \tag{2.46}$$

It is easy to see that $\partial G(c, r, -1)/\partial c = 0$ gives

$$-12c + 12cr^2 + 39c^2 - 28 - 12r^2 + 56r + 9c^2r^2 - 42c^2r = 0. \tag{2.47}$$

Upon Substituting (2.46) in (2.47), we get $-30c^2 + 160c - 84 = 0$. The solution of this equation in the range $(0, 2)$ is $c = \left(40 - \sqrt{970}\right)/15$. From (2.46), we have $r = (-7)\left(\sqrt{970} - 61\right)/393$ and the value of the function G at this point is $G(\left(40 - \sqrt{970}\right)/15, \left(\sqrt{970} - 61\right)/393, -1) \approx 16.0781$. If the expression in the modulus of the equation (2.45) is negative, then we have

$$G(c, r, -1) = 24 - 6c^2 - 24r^2 + 6c^2r^2 - 13c^3 + 28c + 12r^2c - 56rc - 3r^2c^3 + 14c^3r.$$

A computation shows that $\partial G(c, r, -1)/\partial r = 0$ implies

$$r = \frac{7c}{3(c - 2)}, \tag{2.48}$$

which in turn restricts the value of r to be negative, whenever $c \in (0, 2)$, but $r \in (0, 1)$. Therefore, there exist no solution of the system of equations $\partial G(c, r, -1)/\partial r = 0$ and $\partial G(c, r, -1)/\partial c = 0$ in this case. Thus

$$\max_{c\in(0,2),r\in(0,1)} G(c, r, -1) = 16.0781.$$

Finally considering the rectangular face $d = 1$, we have

$$G(c, r, 1) = 24 - 6c^2 - 24r^2 + 6c^2r^2 + c|13c^2 - 28 - (4 - c^2)r(3r + 14)|. \tag{2.49}$$

If the expression in the modulus is positive, then $\partial G(c, r, 1)/\partial r = 0$ implies

$$r = -\frac{7c}{3(c + 2)}, \tag{2.50}$$

which forces r to be negative, whenever $c \in (0, 2)$, but $r \in (0, 1)$. Therefore, there exist no solution in this case. If the expression in the modulus is negative, then $\partial G(c, r, 1)/\partial r = 0$ implies

$$r = \frac{7c}{3(2 - c)}. \tag{2.51}$$

On substituting (2.51) in $\partial G(c, r, 1)/\partial c = 0$, we get $84 + 160c + 30c^2 = 0$. This equation has no solution in $(0, 2)$. Thus in either case, there is no solution of the following equations:

$$\frac{\partial G(c, r, 1)}{\partial r} = 0 \qquad \text{and} \qquad \frac{\partial G(c, r, 1)}{\partial c} = 0.$$

Now we are left with the edges of K to find where $G(c, r, d)$ attains its maximum value. On the line segment $c = 0; r = 0$, we have $G(0, 0, d) = 24$ and clearly

$$\max_{d\in[-1,1]} G(0, 0, d) = 24.$$

On the line segment $c = 0; r = 1$, we have $G(0, 1, d) = 0$ and therefore,

$$\max_{d\in[-1,1]} G(0, 1, d) = 0.$$

On the line segments $c = 0; d = \pm 1$, we obtain $G(0, r, -1) = G(0, r, 1) = 24(1 - r^2)$, which is a decreasing function of r and therefore,

$$\max_{r\in[0,1]} G(0, r, -1) = \max_{r\in[0,1]} G(0, 0, 1) = 24.$$

Also, we know that $G(2, r, d) = 48$ which implies

$$\max_{d\in[-1,1]} G(2, 0, d) = \max_{d\in[-1,1]} G(2, 1, d) = \max_{r\in[0,1]} G(2, r, \pm 1) = 48.$$

On the line segments $r = 0; d = \pm 1$, we have $G(c, 0, -1) = G(c, 0, 1) = 24-6c^2+c|13c^2-28|$. If $c \in [0, \sqrt{28/13})$, we have $G(c, 0, \pm 1) = 24-6c^2-13c^3+28c$. It is easy to verify that $\partial G(c, 0, \pm 1)/\partial c = 0$ implies $-39c^2 - 12c + 28 = 0$, whose solution in the range $[0, \sqrt{28/13})$ is $c = 2\left(\sqrt{282} - 3\right)/29$ and we have

$$G(2\left(\sqrt{282} - 3\right)/29, 0, \pm 1) = \frac{16}{1521}\left(94\sqrt{282} + 1863\right) \approx 36.2028.$$

If $c \in (\sqrt{28/13}, 2]$, then we have $G(c, 0, \pm 1) = 24 - 6c^2 + 13c^3 - 28c$, which is an increasing function of c and therefore, the maximum value is attained at $c = 2$. At $c = \sqrt{28/13}$, we have $G(c, 0, \pm 1) = G(\sqrt{28/13}, 0, \pm 1) = 144/13 \approx 11.0769$. Combining these three cases, we obtain

$$\max_{c\in(0,2)} G(c, 0, \pm 1) = G(2, 0, \pm 1) = 48.$$

On the line segment $r = 1; d = -1$, we have $G(c, 1, -1) = 2c^3 + 16c$ and it is easy to verify that $\max_{c\in[0,2]} G(c, 1, -1) = G(2, 1, -1) = 48$. Finally considering the line segment $r = 1; d = 1$, we have $G(c, 1, 1) = c|30c^2 - 96|$. If $c^2 > 16/5$, then $G(c, 1, 1)$ is an increasing function of c and thus $\max_{c\in[0,2]} G(c, 1, 1) = G(2, 1, 1) = 48$. If $c^2 < 16/5$, we have $G(c, 1, 1) = c(96 - 30c^2)$ attains its maximum at $\sqrt{96/90}$ and $G(\sqrt{96/90}, 1, 1) = 256/\sqrt{15} \approx 66.0989$. Thus we have

$$\max_{(c,r,d)\in K} G(c, r, d) = \frac{256}{\sqrt{15}}. \tag{2.52}$$

In view of all above, we obtain $|\delta_4| \leq 16/3\sqrt{15}$. It may be verified that a function $h(z) \in \mathcal{P}$ for which $c_1 = \sqrt{96/90}$, $c_2 = 2$ and $c_3 = \sqrt{96/90}$, then it can be written as $h(z) = H_{t_0,-1}$ where $t_0 = (15 - 2\sqrt{15})/60$. The function $\tilde{k}_2(z) \in \mathcal{F}_2$, obtained by solving $z\tilde{k}_2'(z) = (z/(1 - z^2))H_{t_0,-1}(z)$ acts as an extremal function for the inequality $|\delta_4| \leq 16/3\sqrt{15}$ and hence qualifies it to be sharp.

(iv) Upon substitution of $b_2 = 0$, $b_3 = 1$, $b_4 = 0$ and $b_5 = 1$ in (2.12), we get

$$\delta_5 = \frac{2}{5} - c_1^2 + \frac{7}{8}c_1^4 + \frac{7}{15}c_2 - \frac{7}{4}c_1^2c_2 + \frac{1}{3}c_2^2 + \frac{3}{4}c_1c_3 - \frac{1}{5}c_4. \tag{2.53}$$

In view of Lemma 1.3 with $c_1 = c$ $(0 \leq c \leq 2)$, we have

$$\begin{aligned}\delta_5 =& \frac{2}{15} - \frac{23}{30}c^2 + \frac{59}{240}c^4 + \frac{7}{30}x(4 - c^2) - \frac{49}{120}xc^2(4 - c^2) - \frac{9}{80}x^2c^2(4 - c^2)\\ &- \frac{1}{40}x^3c^2(4 - c^2) - \frac{1}{10}x(4 - c^2) + \frac{1}{12}x^2(4 - c^2)^2\\ &+ \left(\frac{11}{40} + \frac{1}{10}x\right)(4 - c^2)(1 - |x|^2)cy + \frac{1}{10}(4 - c^2)(1 - |x|^2)\bar{x}y^2\\ &- \frac{1}{10}(4 - c^2)(1 - |x|^2)(1 - |y|^2)z.\end{aligned} \tag{2.54}$$

Now applying triangle inequality, we get

$$|\delta_5|\leq |\gamma_1(c,x)|+|\gamma_2(c,x)||y|+|\gamma_3(c,x)||y|^2+|\gamma_4(c,x,y)||z|, \tag{2.55}$$

where

$$\begin{aligned}
\gamma_1(c,x) = \quad & \frac{2}{15}-\frac{23}{30}c^2+\frac{59}{240}c^4+\frac{7}{30}x(4-c^2)-\frac{49}{120}xc^2(4-c^2)\\
& -\frac{9}{80}x^2c^2(4-c^2)-\frac{1}{40}x^3c^2(4-c^2)-\frac{1}{10}x^2(4-c^2)+\frac{1}{12}x^2(4-c^2)^2\\
\gamma_2(c,x) = \quad & \left(\frac{11}{40}+\frac{1}{10}x\right)(4-c^2)(1-|x|^2)c\\
\gamma_3(c,x) = \quad & \frac{1}{10}(4-c^2)(1-|x|^2)\bar{x}\\
\gamma_4(c,x,y) = \quad & -\frac{1}{10}(4-c^2)(1-|x|^2)(1-|y|^2).
\end{aligned}$$

Taking $x=re^{i\theta}$ with $d=\cos\theta$ such that $0\leq r\leq 1$ and $0\leq\theta\leq 2\pi$, we may write

$$G_1(c,r,d) := |\gamma_1(c,x)|^2. \tag{2.56}$$

Then

$$\begin{aligned}
G_1(c,r,d) =& \frac{c^8r^6}{1600}+\frac{47c^8r^5d}{4800}+\frac{49c^8r^4d^2}{1200}+\frac{1033c^8r^4}{57600}+\frac{59c^8r^3d^3}{1200}+\frac{443c^8r^3d}{3600}\\
&+\frac{2773c^8r^2d^2}{14400}+\frac{2029c^8r^2}{28800}+\frac{2891c^8rd}{14400}+\frac{3481c^8}{57600}-\frac{c^6r^6}{200}\\
&-\frac{9}{100}c^6r^5d-\frac{7}{20}c^6r^4d^2-\frac{1607c^6r^4}{7200}-\frac{7}{20}c^6r^3d^3-\frac{1169}{900}c^6r^3d\\
&-\frac{5761c^6r^2d^2}{3600}-\frac{1043c^6r^2}{1440}-\frac{2779c^6rd}{1800}-\frac{1357c^6}{3600}+\frac{c^4r^6}{100}+\frac{1}{4}c^4r^5d\\
&+\frac{21}{25}c^4r^4d^2+\frac{47c^4r^4}{48}+\frac{16}{25}c^4r^3d^3+\frac{1333}{300}c^4r^3d+\frac{207}{50}c^4r^2d^2\\
&+\frac{653c^4r^2}{300}+\frac{343}{100}c^4rd+\frac{49c^4}{75}-\frac{14}{75}c^2r^5d-\frac{28}{75}c^2r^4d^2-\frac{77c^2r^4}{45}\\
&-\frac{8}{75}c^2r^3d^3-\frac{1193}{225}c^2r^3d-\frac{766}{225}c^2r^2d^2-\frac{401c^2r^2}{225}-\frac{434}{225}c^2rd\\
&-\frac{46c^2}{225}+\frac{196r^4}{225}+\frac{392r^3d}{225}+\frac{112r^2d^2}{225}+\frac{28r^2}{45}+\frac{56rd}{225}+\frac{4}{225}.
\end{aligned}$$

Using the similar method as in (iii), we observe that the maximum of $G_1(c,r,d)$ on the cuboid $\{(c,r,d): 0\leq c\leq 2, 0\leq r\leq 1, -1\leq d\leq 1\}$ is attained at the point $(\sqrt{15/7},1,0)$ and is equal to $791/392$. Next, let us write $|x|=r$ and $|y|=q$ in

$$G_2(c,r,q) := |\gamma_2(c,x)||y|+|\gamma_3(c,x)||y|^2+|\gamma_4(c,x,y)| \tag{2.57}$$

Again using the similar method, we find the maximum of $G_2(c,r,q)$ on the cuboid $\{(c,r,q): 0\leq c\leq 2, 0\leq r\leq 1, 0\leq q\leq 1\}$ and observe that the maximum value is attained at $(c_0,r_0,1)$, where $c_0\approx 1.12539$ is the smallest root of $5552+1936c-6200c^2-2776c^3+1479c^4+945c^5=0$ and $r_0=11(4-3c_0^2)/(4(3c_0^2+2c_0-4))\approx 0.268895$. Hence $G_2(c_0,r_0,1)\approx 0.929727$. From (2.55), (2.56) and (2.57), we obtain

$$|\delta_5|\leq \max_{c\in[0,2],r\in[0,1],s\in[-1,1]}\sqrt{G_1(c,r,s)}+\max_{c\in[0,2],r\in[0,1],q\in[0,1]}G_2(c,r,q). \tag{2.58}$$

Hence $|\delta_5|\leq 791/392+0.929727\approx 2.947584$. In fact, this bound is not sharp but we may conclude that the sharp bound lies in the range $[791/392, 791/392+0.929727]$ as there exists a function

$\hat{k}_2(z) \in \mathcal{F}_2$ for which the fifth inverse coefficient is equal to 791/392. The function $\hat{k}_2(z)$ can be obtained by solving $z\hat{k}_2(z) = (z/(1-z^2))H_{t_1,\beta_1}(z)$, where $t_1 = (1/56)\left(14 - \sqrt{105}\right)$ and $\beta = -1$. ∎

Theorem 2.3. *Let $f(z) = z + a_2z^2 + a_3z^3 + \ldots \in \mathcal{F}_3$. Then*

(*i*) $|\delta_2| \leq 3/2$
(*ii*) $|\delta_3| \leq 19/6$
(*iii*) $|\delta_4| \leq 61/8$
(*iv*) $|\delta_5| \leq 2371/120$.

These all bounds are sharp.

Proof. (i) Let $f \in \mathcal{F}_3$. Since f is close to convex function with respect to the starlike function $z/(1 - z + z^2)$, from (2.3) we have

$$zf'(z) = \frac{z}{1 - z + z^2}p(z). \tag{2.59}$$

For $g(z) = z/(1 - z + z^2)$, we have $b_2 = 1,\ b_3 = 0,\ b_4 = -1$ and $b_5 = -1$. Thus, in view of Lemma 1.1, (2.13) reduces to

$$2|\delta_2| \leq 1 + |c_1| \leq 3. \tag{2.60}$$

We know that this inequality is sharp whenever $|c_1| = 2$, which is true for the function $P_{1,\vartheta}(z)$ $(0 \leq \vartheta < 2\pi)$, defined in Lemma 1.2. The inequality $|\delta_2| \leq 3/2$ is sharp since there exists an extremal function $\hat{f}_3 \in \mathcal{F}_3$, which is a solution of $z\hat{f}'_3(z) = z(1 - z)^{-1}P_{1,\vartheta}(z)$.

(ii) By taking $b_2 = 1$ and $b_3 = 0$ in (2.15), we get

$$3|\delta_3| \leq 2 - \frac{p^2}{2} + \left|\frac{1}{2} - (1 + pq + ip\sqrt{1 - q^2})^2\right|,$$

which can be written as

$$3|\delta_3| \leq 2 - \frac{p^2}{2} + \sqrt{(p^2 + \frac{3}{2} + 2pq)^2 - 2p^2(1 - q^2)} =: \phi_3(p, q). \tag{2.61}$$

In the domain $D = \{(p, q) : 0 \leq p \leq 2, -1 \leq q \leq 1\}$, we need to find the points where $\phi_3(p, q)$ attains its maximum. A simple computation shows that the only solution of

$$\frac{\partial\phi_3(p,q)}{\partial p} = 0 \quad \text{and} \quad \frac{\partial\phi_3(p,q)}{\partial q} = 0$$

in $\mathbb{R} \times \mathbb{R}$ is (0,0). Thus, the maximum value of $\phi_3(p, q)$ is not attained inside D. Now, we consider the edges of D to find the maximum of $\phi_3(p, q)$. On the line segment $p = 0$, $\phi_3(0, q) = 7/2$. On the line segment $p = 2$, $\phi_3(2, q) = \sqrt{(11/2 + 4q)^2 - 8(1 - q^2)}$ which is a decreasing function of q, whenever $-1 \leq q < -11/12$ and thus maximum value is attained at $q = -1$. Similarly, $\phi_3(2, q)$ is an increasing function of q, whenever $-11/12 < q \leq 1$ and thus maximum value is attained at $q = 1$. Now, we have $\phi_3(2, -1) = 3/2$ and $\phi_3(2, 1) = 19/2$. Thus, $\max_{q\in[-1,1]} \phi_3(2, q) = \phi_3(2, 1) = 19/2$. On the line segment $q = -1$, $\phi_3(p, -1) = 7/2 + p^2/2 - 2p$ which is a decreasing function of p whenever $0 \leq p \leq 2$. Thus, $\max_{p\in[0,2]} \phi_3(p, -1) = \phi_3(0, -1) = 7/2$. On the line segment $q = 1$, $\phi_3(p, 1) = 7/2 + p^2/2 + 2p$ which is an increasing function of p whenever $0 \leq p \leq 2$. Thus, $\max_{p\in[0,2]} \phi_3(p, 1) = \phi_3(2, 1) = 19/2$. It is easy to note that maximum value of $\phi_3(p, q)$ is attained at $(2, 1)$ and is equal to $19/2$. Thus $|\delta_3| \leq 19/6$. The extremal function $\tilde{f}_3 \in \mathcal{F}_3$ for which the upper bound of $|\delta_3|$ is sharp, can be obtained by solving (2.59) with $P_{1,0}(z)$ in place of $p(z)$.

(iii) Upon substituting $b_2 = 1, b_3 = 0$ and $b_4 = -1$ in (2.17), we get

$$48|\delta_4| \leq 6(4 - c^2)(1 - r^2) + \sqrt{\psi_3(c, r, d)} =: N(c, r, d), \tag{2.62}$$

where

$$\begin{aligned}\psi_3(c,r,d) =&(13c^3+36c^2+50c+18)^2+(4-c^2)^2r^2(84c^2dr+9c^2r^2\\&+196c^2+84crd+392c+196)\\&-2(4-c^2)r(13c^3+36c^2+50c+18)(-3cr+6cd^2r+14cd+14d).\end{aligned}$$

Now we have to find the points in the rectangular cube $K := \{(c,r,d) : 0 \leq c \leq 2, 0 \leq r \leq 1 \text{ and } -1 \leq d \leq 1\}$, where $N(c,r,d)$ attains its maximum. First we try to explore the interior of K for the critical points by solving:

$$\frac{\partial N(c,r,d)}{\partial c} = \frac{\partial N(c,r,d)}{\partial r} = \frac{\partial N(c,r,d)}{\partial d} = 0. \tag{2.63}$$

It is easy to note that $\partial N(c,r,d)/\partial d = 0$ implies

$$d = \frac{-21c^4r^2-91c^4-21c^3r^2-343c^3+84c^2r^2-602c^2+84cr^2-476c-126}{6c\,(13c^3+36c^2+50c+18)\,r}.$$

Upon substituting the above value of d in $\partial N(c,r,d)/\partial r = 0$, we arrive at no real solution of the equations stated in (2.63). That is, there exist no solution inside the cuboid K. Thus, we proceed further to find the maximum of $N(c,r,d)$ on the six faces of K. On the rectangular face $c = 0$, $N(r,c,d)$ reduces to

$$N(0,r,d) = 24(1-r^2) + \sqrt{324-2016rd+3136r^2}.$$

A simple calculation shows that $(0,0)$ is the only solution of the system

$$\frac{\partial N(0,r,d)}{\partial r} = 0 \quad \text{and} \quad \frac{\partial N(0,r,d)}{\partial d} = 0.$$

Thus, we can say that the maximum value is not attained in the interior of the face $c = 0$. Now we consider the face $c = 2$. Then we have $N(2,r,d) = 366$ and clearly

$$\max_{r\in(0,1),d\in(-1,1)} N(2,r,d) = 366.$$

On the square face $r = 0$, we have $N(c,0,d) = 13c^3+30c^2+50c+18$, which is clearly an increasing function of c, where $c \in (0,2)$. Thus, we have

$$\max_{c\in(0,2),d\in(-1,1)} N(c,0,d) = N(2,0,d) = 366.$$

On the square face $r = 1$, we have

$$N(c,1,d) = \sqrt{\rho_3(c,d) + 2\,(13c^3+36c^2+50c+18)\,(c^2-4)\,(6cd^2+14cd-3c+14d)},$$

where

$$\begin{aligned}\rho_3(c,d) =& \left(13c^3+36c^2+50c+18\right)^2+\left(4-c^2\right)^2(84c^2d+205c^2+84cd+392c\\&+196).\end{aligned}$$

It is easy to verify that $\partial N(c,1,d)/\partial d = 0$ implies

$$d = -\frac{14(c+1)\left(16c^3+36c^2+38c+18\right)}{12c\,(13c^3+36c^2+50c+18)}. \tag{2.64}$$

A simple calculation reveals that $\partial N(c,1,d)/\partial c = 0$ implies

$$\begin{aligned}&\left(1776c^5+5560c^4+4656c^3+5064c^2+13416c+8504\right)\\&+\left(2688c^5+7280c^4+1120c^3-12768c^2-15568c-6272\right)d\\&+\left(936c^5+2160c^4-96c^3-4536c^2-4800c-864\right)d^2=0.\end{aligned} \tag{2.65}$$

Upon substituting (2.64) in (2.65) and simplifying, we have

$$\frac{-8(5c^3+18c^2+31c+9)\tau(c)}{3c^2(13c^3+36c^2+50c+18)^2}=0, \tag{2.66}$$

where

$$\begin{aligned}\tau(c) =&1950c^{10}+9355c^9-17290c^8-222568c^7-729631c^6-1304461c^5\\ &-1390549c^4-805631c^3-172575c^2+33516c+15876.\end{aligned}$$

The numerical solution of (2.66) in the range $(0,2)$ is $c\approx 0.229261$. Substituting this value in (2.64), we obtain $d\approx -5.71653$, but $d\in(-1,1)$. Therefore, there exists no solution of

$$\frac{\partial N(c,1,d)}{\partial d}=0 \qquad \text{and} \qquad \frac{\partial N(c,1,d)}{\partial c}=0.$$

Now we shall consider the rectangular face $d=-1$, then we have

$$\begin{aligned}N(c,r,-1) =&3c^3r^2-14c^3r+13c^3+6c^2r^2-14c^2r+30c^2-12cr^2+56cr\\ &+50c-24r^2+56r+42.\end{aligned} \tag{2.67}$$

A computation shows that $\partial N(c,r,-1)/\partial r=0$ implies

$$r=\frac{7(c+1)}{3(c+2)}. \tag{2.68}$$

It is easy to see that $\partial N(c,r,-1)/\partial c=0$ gives

$$9c^2r^2-42c^2r+39c^2+12cr^2-28cr+60c-12r^2+56r+50=0 \tag{2.69}$$

Upon substituting (2.68) in (2.69), we get $10c^2-60c-99=0$. Clearly, this equation has no solution in $(0,2)$. Finally considering the rectangular face $d=1$, we have

$$\begin{aligned}N(c,r,1) =&6c^2r^2-6c^2-24r^2+24+|(13c^3+36c^2+50c+18)\\ &-(4-c^2)(14r+14cr+3cr^2)|.\end{aligned} \tag{2.70}$$

If the expression in the modulus in (2.70) is positive, then $\partial N(c,r,1)/\partial r=0$ implies

$$r=-\frac{7(c+1)}{3(c+2)}, \tag{2.71}$$

which compels r to be negative, whenever $c\in(0,2)$, but $0<r<1$. Thus, there exist no solution in this case. If the expression in the modulus of the equation (2.70) is negative, then $\partial N(c,r,1)/\partial r=0$ implies

$$r=\frac{7(c+1)}{3(2-c)}. \tag{2.72}$$

On substituting (2.72) in $\partial N(c,r,1)/\partial c=0$, we get $5(6c^2+28c+19)/3=0$. Obviously, this equation has no solution in $(0,2)$. Thus in either case, there exists no solution of

$$\frac{\partial N(c,r,1)}{\partial r}=0 \qquad \text{and} \qquad \frac{\partial N(c,r,1)}{\partial c}=0.$$

Now the edges of K need to be explored for the critical points, where $N(c,r,d)$ attains its maximum value. On the line segment $c=0;r=0$, we have $N(0,0,d)=42$ and clearly $\max_{d\in[-1,1]}N(0,0,d)=42$. On the line segment $c=0;r=1$, we have $N(0,1,d)=\sqrt{3460-2016d}$, which is clearly a decreasing function of d and therefore, $\max_{d\in[-1,1]}N(0,1,d)=N(0,1,-1)=74$. On the line segment $c=0;d=-1$, we obtain $N(0,r,-1)=42-24r^2+56r$, which is an increasing function of r and therefore, $\max_{r\in[0,1]}N(0,r,-1)=Q(0,1,-1)=74$. On the line segment $c=0;d=1$, we observe that $N(0,r,1)=24(1-r^2)+|18-56r|$, which is a decreasing function of r, whenever $r\in(0,9/28)$ and therefore, attains its maximum value 42 at $r=0$. Similarly $N(0,r,1)=6+56r-24r^2$, which is an increasing function of r, whenever $r\in(9/28,1)$ and therefore, attains its maximum

value 38 at $r = 1$. Clearly at $r = 9/28$, we have $N(0, 9/28, 1) = 2109/98$. Combining the above three cases, we have $\max_{r\in[0,1]} N(0, r, 1) = N(0, 0, 1) = 42$. Also, we know that $N(2, r, d) = 366$, which implies

$$\max_{d\in[-1,1]} N(2, 0, d) = \max_{d\in[-1,1]} N(2, 1, d) = \max_{r\in[0,1]} N(2, r, \pm 1) = 366.$$

On the line segments $r = 0; d = \pm 1$, we have $N(c, 0, -1) = N(c, 0, 1) = 13c^3 + 30c^2 + 50c + 42$, which is clearly an increasing function of c and thus

$$\max_{c\in[0,2]} N(c, 0, -1) = \max_{c\in[0,2]} N(c, 0, 1) = 366.$$

On the line segment $r = 1; d = -1$, we obtain $N(c, 1, -1) = 2c^3 + 22c^2 + 94c + 74$ and it is easy to verify that $\max_{c\in[0,2]} N(c, 1, -1) = N(2, 1, -1) = 366$. Finally on the line segment $r = 1; d = 1$, we have $N(c, 1, -1) = |M|$, where $M = 30c^3+50c^2-18c-38$. If $M > 0$ then $N(c, 1, -1)$ is an increasing function of c and therefore, $\max_{c\in[0,2]} N(c, 1, 1) = N(2, 1, 1) = 366$. If $M < 0$ then $N(c, 1, -1)$ attains its maximum at $\left(\sqrt{1030} - 25\right)/45$ and $N(\left(\sqrt{1030} - 25\right)/45, 1, 1) \approx 19.7387$. As $N(c, 1, 1)$ is uniformly continuous on $[0, 2]\times\{1\}\times\{1\}$, we have $|N(\left(\sqrt{1030} - 25\right)/45, 1, 1)-19.7387| < \epsilon << 1$. When ϵ approaches to 0, we have $\max_{c\in[0,2]} N(c, 1, 1) = 19.7387$. In view of all the above cases, we have

$$\max_{(c,r,d)\in K} N(c, r, d) = 366. \tag{2.73}$$

Thus we obtain $|\delta_4| \leq 61/8$. The function $\tilde{k}_3 \in \mathcal{F}_3$ obtained by solving $z\tilde{k}_3'(z) = z(1 - z + z^2)^{-1}P_{1,0}(z)$ acts as an extremal function for the inequality $|\delta_4| \leq 61/8$ and hence the inequality becomes sharp.

(iv) Upon substituting $b_2 = 1,\ b_3 = 0,\ b_4 = -1$ and $b_5 = -1$ in (2.12), we get

$$\begin{aligned}\delta_5 =& \frac{13}{40} + \frac{6}{5}c_1 + \frac{25}{12}c_1^2 + \frac{7}{4}c_1^3 + \frac{7}{8}c_1^4 - c_2 - \frac{25}{12}c_1c_2 - \frac{7}{4}c_1^2c_2 \\ &+ \frac{1}{3}c_2^2 + \frac{11}{20}c_3 + \frac{3}{4}c_1c_3 - \frac{1}{5}c_4,\end{aligned} \tag{2.74}$$

which can be written as

$$\delta_5 = \frac{1}{5}A + \frac{11}{20}B + \varrho^*(c_1, c_2, c_3) + \varsigma^*(c_1, c_2), \tag{2.75}$$

where

$$\begin{aligned}A &= c_1^4 - 3c_1^2c_2 + c_2^2 + 2c_1c_3 - c_4 \\ B &= c_3 - 2c_1c_2 + c_1^3 \\ \varrho^*(c_1, c_2, c_3) &= \frac{13}{40} + \frac{6}{5}c_1 + \frac{2}{15}c_2^2 + \frac{7}{20}c_1c_3 \qquad \text{and} \\ \varsigma^*(c_1, c_2) &= \left(\frac{25}{12}c_1^2 - c_2\right) + \frac{59}{60}c_1\left(\frac{72}{59}c_1^2 - c_2\right) + \frac{23}{20}c_1^2\left(\frac{27}{46}c_1^2 - c_2\right).\end{aligned}$$

Using Lemma 1.4, we have $|A| \leq 2$ and $|B| \leq 2$. Evidently $\varrho^*(c_1, c_2, c_3)$ attains its maximum at $(2, 2, 2)$. Thus, $|\varrho^*(c_1, c_2, c_3)| \leq 559/120$. Now, let us consider $\varsigma^*(c_1, c_2)$, which can be again written in terms of c_1 and x by using Lemma 1.3 as follows:

$$\varsigma_1^*(c_1, x) = \frac{19}{12}c_1^2 + \frac{17}{24}c_1^3 + \frac{1}{10}c_1^4 - x(4 - c_1^2)\left(\frac{1}{2} + \frac{59}{120}c_1 + \frac{23}{40}c_1^2\right). \tag{2.76}$$

We shall aptly write $c_1 = c$ $(0 \leq c \leq 2)$ and $x = re^{i\theta}$ with $d = \cos\theta$ such that $0 \leq r \leq 1$ and $0 \leq \theta \leq 2\pi$. Then (2.76) reduces to

$$\varsigma_2^*(c,r,d) = \frac{19}{12}c^2 + \frac{17}{24}c^3 + \frac{1}{10}c^4 - (4-c^2)\left(\frac{1}{2} + \frac{59}{120}c + \frac{23}{40}c^2\right)(rd + ir\sqrt{1-d^2}). \tag{2.77}$$

Thus, we may write

$$|\varsigma_2^*(c,r,d)| = \sqrt{(\operatorname{Re}\varsigma_2^*(c,r,d))^2 + (\operatorname{Im}\varsigma_2^*(c,r,d))^2} =: N_1(c,r,d), \tag{2.78}$$

where

$$\operatorname{Re}\varsigma_2^*(c,r,d) = \frac{19}{12}c^2 + \frac{17}{24}c^3 + \frac{1}{10}c^4 - (4-c^2)rd\left(\frac{1}{2} + \frac{59}{120}c + \frac{23}{40}c^2\right)$$

and

$$\operatorname{Im}\varsigma_2(c,r,d) = (4-c^2)r\sqrt{(1-d^2)}\left(\frac{8}{30} + \frac{59}{120}c + \frac{23}{40}c^2\right).$$

In order to maximize $N_1(c,r,d)$, we use the similar method as in part(iii) and obtain

$$\max_{c\in[0,2],r\in[0,1],d\in[-1,1]} N_1(c,r,d) = N_1(2,r,d) = \frac{68}{5}.$$

Applying triangle inequality on (2.75), we have

$$\begin{aligned} |\delta_5| &\leq \frac{1}{5}|A| + \frac{11}{20}|B| + |\varrho^*(c_1,c_2,c_3)| + |\varsigma^*(c_1,c_2)| \\ &\leq \frac{1}{5}(2) + \frac{11}{20}(2) + \frac{559}{120} + \frac{68}{5} \\ &= \frac{2371}{120}. \end{aligned}$$

The inequality $|\delta_5| \leq 2371/120$ is sharp since there exists an extremal function $\hat{k}_3 \in \mathcal{F}_3$, which is a solution of $z\hat{k}_3'(z) = z(1-z)^{-1}P_{1,0}(z)$. ∎

References

[1] R. M. Ali, Coefficients of the inverse of strongly starlike functions, Bull. Malays. Math. Sci. Soc. (2) **26** (2003), no. 1, 63–71.

[2] W. Kaplan, Close-to-convex schlicht functions. Mich. Math. J. 1,(1952), 169-185

[3] G. P. Kapoor and A. K. Mishra, Coefficient estimates for inverses of starlike functions of positive order, J. Math. Anal. Appl. **329** (2007), no. 2, 922–934.

[4] O. S. Kwon, A. Lecko and Y. J. Sim, On the fourth coefficient of functions in the Carathéodory class, Comput. Methods Funct. Theory **18** (2018), no. 2, 307–314.

[5] R. J. Libera and E. J. Złotkiewicz, Coefficient bounds for the inverse of a function with derivative in $\mathcal{P}$. II, Proc. Amer. Math. Soc. **92** (1984), no. 1, 58–60

[6] R. J. Libera and E. J. Złotkiewicz, Early coefficients of the inverse of a regular convex function, Proc. Amer. Math. Soc. **85** (1982), no. 2, 225–230.

[7] C. Pommerenke, Univalent functions, Vandenhoeck & Ruprecht, Göttingen, 1975.

[8] S. Ponnusamy, A. Rasila and A. Sairam Kaliraj, Harmonic close-to-convex functions and minimal surfaces, Complex Var. Elliptic Equ. **59** (2014), no. 7, 986–1002.

[9] S. Ponnusamy, A. Vasudevarao and H. Yanagihara, Region of variability for close-to-convex functions, Complex Var. Elliptic Equ. **53** (2008), no. 8, 709–716.

[10] S. Ponnusamy, A. Vasudevarao and H. Yanagihara, Region of variability for close-to-convex functions. II, Appl. Math. Comput. **215** (2009), no. 3, 901–915.

[11] U. Pranav Kumar and A. Vasudevarao, Logarithmic coefficients for certain subclasses of close-to-convex functions, Monatsh. Math. **187** (2018), no. 3, 543–563.

[12] H. Silverman, Coefficient bounds for inverses of classes of starlike functions, Complex Variables Theory Appl. **12** (1989), no. 1-4, 23–31.

[13] H. M. Srivastava, A. K. Mishra and S. N. Kund, Coefficient estimates for the inverses of starlike functions represented by symmetric gap series, PanAmer. Math. J. **21** (2011), no. 4, 105–123.

[14] Q.-H. Xu, C.-B. Lv and H. M. Srivastava, Coefficient estimates for the inverses of a certain general class of spirallike functions, Appl. Math. Comput. **219** (2013), no. 12, 7000–7011.

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
E-mail address: priyanka.goel0707@gmail.com

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
E-mail address: spkumar@dce.ac.in

Investigating Political Herd Mentality: A Community Sentiment Based Approach

Anjali Bhavan*
Delhi Technological University
anjalibhavan98@gmail.com

Rohan Mishra*
Delhi Technological University
rohan.mishra1997@gmail.com

Pradyumna Prakhar Sinha*
Delhi Technological University
pradyumna014@gmail.com

Ramit Sawhney
Netaji Subhash Institute of Technology
ramits.co@nsit.net.in

Rajiv Ratn Shah
MIDAS, IIIT-Delhi
rajivratn@iiitd.ac.in

Abstract

Analyzing polarities and sentiments inherent in political speeches and debates poses an important problem today. This experiment aims to address this issue by analyzing publicly-available Hansard transcripts of the debates conducted in the UK Parliament. Our proposed approach, which uses community-based graph information to augment hand-crafted features based on topic modeling and emotion detection on debate transcripts currently surpasses the benchmark results on the same dataset. Such sentiment classification systems could prove to be of great use in today's politically turbulent times, for public knowledge of politicians stands on various relevant issues proves vital for good governance and citizenship. The experiments also demonstrate that continuous feature representations learned from graphs can improve performance on sentiment classification tasks significantly.

1 Introduction

One of the key aspects of a functional, free society is being presented with comprehensive options in electing government representatives. The decision is aided by the positions politicians take on relevant issues like water, housing, etc. Hence, it becomes important to relay political standings to the general public in a comprehensible manner. The Hansard transcripts of speeches delivered in the UK Parliament are one such source of information. However, owing to the voluminous quantity, esoteric language and opaque procedural jargon of Parliament, it is tougher for the non-expert citizen to assess the standings of their elected representative. Therefore, conducting stance classification studies on such data is a challenging task with potential benefits. However, the documents tend to be highly tedious and difficult to comprehend, and thus become a barrier to information about political issues and leanings.

Sentiment analysis of data from various relevant sources (social media, newspapers, transcripts, etc.) has often given several insights about public opinion, issues of contention, general trends and so on (Carvalho et al., 2011; Loukis et al., 2014). Such techniques have even been used for purposes like predicting election outcomes and the reception of newly-launched products. Since these insights have wide-ranging consequences, it becomes imperative to develop rigorous standards and state-of-the-art techniques for them.

One aspect that helps with analyzing such patterns and sentiments is studying about the interconnections and networks underlying such data. Homophily, or the tendency of people to associate with like-minded individuals, is the fundamental aspect of depicting relationships between users of a social network (for instance). Constructing graphs to map such complex relationships and attributes in data could help one arrive at ready insights and conclusions. This comes particularly useful when studying parliamentary debates and sessions; connecting speakers according to factors like party or position affiliations pro-

* Indicates equal contribution.

vides information on how a speaker is likely to respond to an issue being presented. Attempts to analyze social media data based on such approaches have been made (Deitrick and Hu, 2013).

2 Related Work

The analysis of political content and parliamentary debates have opened an exciting line of research in recent years and has shown promising results in tasks of stance classification (Hasan and Ng, 2013) and opinion mining (Karami et al., 2018). A large part of the work initially concentrated on legislative speeches, but the focus has shifted to social media content analysis in recent times. This shift in focus has been particularly rapid with the proliferation of social media data and research (**?**Shah and Zimmermann, 2017; Shah et al., 2016b; Mahata et al., 2018; Shah et al., 2016c,a).

Lauderdale and Herzog (2016) presented their method of determining political positions from legislative speech. The datasets were sourced from Irish and US Senate debates. Rheault et al. (2016) examined the emotional polarity variations in speeches delivered in the British parliament over a hundred years. They observed a correlation between the variations in emotional states of a particular period of time and the national economic situation. Thomas et al. (2006) studied the relationships between segments of speeches delivered in the Congress and the overall tone: of opposition or support. A significant amount of research exists on the political temperament across social media websites like Facebook and Twitter. Stieglitz and Dang-Xuan (2012) studied the relationship between the inherent sentiment of politically relevant tweets and the retweet activity. Ceron et al. (2014) proposed methods for determining the political alignments of citizens and tested them on French and Italian-context datasets. Many new findings based on the contemporary political landscape continue to be developed and presented. Wang and Liu (2018) analyzed US President Donald Trump's speeches delivered during his 2016 election campaign. Rudkowsky et al. (2018) proposed the usage of word embeddings in the place of the traditional Bag-of-Words (BOW) approach for text analysis, and demonstrated experiments on Austrian parliamentary speeches. There have been some approaches to model interactions among members of a network to help in the task of sentiment analysis. Moreover, there have been applications that extract information about each user by representing them as a node in the social graph and creating low dimensional representation usually induced by neural architectures (Grover and Leskovec, 2016). Mishra et al. (2018) and Qian et al. (2018) use such social graph-based features to gain considerable improvement in the task of abuse detection in social media. However there has been no work done to model the interaction between the members of the Parliament for the task of stance classification.

For studying transcripts of speeches delivered in the House of Commons in the UK Parliament, Abercrombie and Batista-Navarro (2018b) curated a dataset consisting of parliamentary motions and debates as provided in the Hansard transcripts, along with other information like party affiliations and polarities of the motions being discussed. This was followed by carrying out studies on the dataset and developing a sentiment analysis model which also demonstrated the results of motion-independent (one-step) and motion-dependent classification of polarities Abercrombie and Batista-Navarro (2018a). This dataset is used for further analysis in our experiments.

3 Dataset

In the UK, transcripts of parliamentary debates are publicly available along with information related to *division votes* as well as manually annotated sentiment labels. To investigate the effectiveness of our pipeline, experiments were conducted using the *HanDeSeT* dataset as created by (Abercrombie and Batista-Navarro, 2018b). The dataset consists of 607 politicians and their speeches over various motions, with a total of 1251 samples. The speeches are divided into five utterances, and other features such as *Debate ID, Debate title, Motion subject with polarities: manual annotation and ruling-opposing-based, Motion and Speaker party affiliations, Speech Polarities: manual and vote-based, Rebellion percentage.*

Sentiment polarity is present in both speeches and motions. Hence labels are provided for motion polarities as well. Two label types are provided for motions: a manually-annotated one predicting positive or negative polarity, and a government/opposition one decided as follows: if the

speaker who proposes the motion belongs to the ruling government, the polarity is positive; if the speaker belongs to the opposition then the polarity is negative. Two label types are provided for speeches as well: one manually-annotated, and the other a speaker-vote label extracted from the division related to the corresponding debate.

4 Methodology

The models described in Abercrombie and Batista-Navarro (2018a) extracted n-gram features (uni-grams, bi-grams, tri-grams, and their combinations) from the utterances for sentiment classification. The stance-based relationships between the members are modeled, and their effectiveness is analyzed. This study aims to develop on the limitations of using only text-based features and by doing so present a sound, coherent model for sentiment classification for parliamentary speeches.
The methodology consists of the following subsections: *preprocessing*, to describe the initial data preprocessing methods undertaken; *feature extraction*, which discusses the feature sets used for our model, and *model description* and training, to elaborate on our model and training procedures.

4.1 Preprocessing

The dataset was preprocessed for further analysis. This was required so unnecessary words; characters etc. could be removed and not add further noise to the dataset. The text was lower-cased, and all punctuation marks and other special characters were removed. Following this, stopword removal was done using NLTK. Finally, a few custom stopwords specific to the parliamentary procedure were removed. These were taken from Abercrombie and Batista-Navarro (2018b). Finally, the utterances were concatenated and prepared for feature extraction and model training.

4.2 Feature Extraction

4.2.1 Textual Features

Various textual features were extracted for classification and normalized using the L2 norm. These are listed below.

- *TF-IDF*: Term Frequency-inverse Document Frequency (TF-IDF) features were extracted from n-grams (upto 3) in the text. N-gram features are immensely useful for factoring in contextual information surrounding the components of a text (whether characters or words) and are widely used for text analysis, language processing, etc.

- LDA-based topic modeling: Topic modeling is used to derive information related to the underlying "topics" contained in a text. In order to extract such topic-based features from the utterances, the Latent Dirichlet Allocation (LDA) (Blei et al., 2003) model was used. The probability distribution over the most commonly occurring 30 topics was used as features for each speech.

- *NRC Emotion*: The NRC Emotion Lexicon (Mohammad and Turney, 2013) is a publicly available lexicon that contains commonly occurring words along with their affect category (anger, fear, anticipation, trust, surprise, sadness, joy, or disgust) and two polarities (negative or positive). The score along these 10 features was computed for the utterances.

4.2.2 Graph-based features

For our analysis, two graphs were constructed from the dataset. The graph consists of nodes that represent the members who participate in the proceedings of the Parliament. The edges among the members are conditioned upon their accord or discord on debates regarding policies. Two members of the same or varying political parties either agree on a policy or differ on it. Therefore, the two graphs are constructed.

- *simGraph*: In order to model the similarity on stances among members, $G_{sim}(v, e)$ is a weighted undirected graph induced on the dataset with vertices v corresponding to the members m of political parties where an edge e between two vertices v and u is defined as $weight(e) = \mid f(v) \cap f(u) \mid$ where $f(v)$ is the set of stances taken by the member that is represented by node v.

- *oppGraph*: Similarly, to model the differences among the members, $G_{opp}(v, e)$ is induced on the dataset such that an edge e between two vertices v and u is defined as $weight(e) = \mid (f(v) \setminus f(u)) \cap (f(u) \setminus f(v)) \mid$ where $f(v)$ is the set of stances taken by the member that is represented by node v.

node2vec: To obtain community based embeddings, feature representations were generated using *node2vec* (Grover and Leskovec, 2016).

Table 1: Statistical properties of constructed graphs

Properties	Values	
	simGraph	oppGraph
Number of nodes	607	607
Number of edges	5,431	2,893
Density of graphs	0.0295	0.0157
Average weight	1.047	1.037

node2vec is similar to word2vec (Mikolov et al., 2013b) and uses the same loss function to assign similar representations to nodes that are in the context of each other. To obtain the context of a node, node2vec samples a neighborhood for each of the nodes by constructing a fixed number of random walks of constant length. The traversal strategy for these random walks is determined by the hyper-parameters Return Parameter p and In-out Parameter q which have the ability to moderate the sampling between a depth-first strategy and a breadth-first strategy. The return parameter p controls the likelihood of immediately revisiting a node in the walk, while the in-out parameter q allows the search to differentiate between inward and outward nodes.

Formally, given a graph $G = (V, E)$, we learn a function $f : V \rightarrow \mathbb{R}^d$ that maps nodes to feature representations where d is the dimension of the representation. In order to do so, for every node $u \in V$, we define a neighbourhood $N_S(u) \subseteq V$ is generated using the sampling strategy S.

The skip-gram model (Mikolov et al., 2013a) is then employed to maximize the following objective function:

$$\max_f \sum logPr(N_s(u)|f(u)). \quad (1)$$

Combining Graph Embeddings: To combine embeddings generated for each member in the two graphs, a dense neural network was used. The embeddings were projected onto a linear layer and fine-tuned upon the classification task. The penultimate layer of the model was used as the graph embedding corresponding to each user.
The network consisted of two input layers for the two embedding sets, followed by single dense layers with hidden layer size 16 and activation ReLU. These two layers were then combined, and the resultant combination passed through two dense layers (layer size 16, activation ReLU), before being passed through a final dense softmax layer. The network was optimized using Adam, and trained over 20 epochs with batch size 64.

4.2.3 Other features

Of all the feature sets explored in Abercrombie and Batista-Navarro (2018a), the feature set all the meta-features had the best results consistently across all the three models. Hence, we used these in addition to our textual and community-based graph features. The meta-features consisted of speaker party affiliation, debate IDs and motion party affiliation.

4.3 Baseline models

The original experiments consisted of 3 models for classification: a one-step model and two two-step models. We consider the two-step models as our baselines, which are described below.

- manAnnot: a two-step model in which motion polarity classification is first performed based on manually-annotated positive or negative sentiments, corresponding to model 2a in the original experiments;
- govAnnot: a two-step model in which motion polarity classification is first performed based on government or opposition labeling, corresponding to model 2b in the original experiments.

In the case of the two two-step models, the dataset is divided into two parts based on the predicted polarities. These two divided datasets are then used for training and classification separately. Two classifiers were used in both the steps: Support Vector Machine (SVM) with the linear kernel and Multi-Layer Perceptron (MLP) with 1 hidden layer containing 100 neurons.

4.4 Proposed model

In the original experiment, the best results were obtained from the two-step models with the MLP classifier. A similar two-step approach is followed here as well, with MLP as the chosen classifier. The network consists of 1 hidden layer with 100 neurons.

5 Experiments

Experiments on two models are presented:

1. manAnnot: here, the dataset is divided into two parts based on predicted motion polarity from manually annotated labels.

Table 2: Observations for manModel

Feature Combinations	without graph-based features				with graph-based features			
	Acc.(%)	**Prec.**	**Recall**	**F1**	**Acc.(%)**	**Prec.**	**Recall**	**F1**
TF-IDF+meta	89.38	0.897	0.887	0.884	92.26	0.920	0.917	0.917
LDA+meta	86.34	0.875	0.839	0.850	92.34	0.930	0.902	0.915
NRC+meta	86.43	0.860	0.859	0.858	92.25	0.932	0.903	0.916
TF-IDF+LDA+meta	88.59	0.885	0.867	0.874	91.70	0.915	0.910	0.912
TF-IDF+NRC+meta	88.73	0.896	0.867	0.879	91.94	0.918	0.914	0.914
LDA+NRC+meta	85.86	0.861	0.828	0.842	**92.66**	**0.938**	**0.905**	**0.920**
TF-IDF+LDA+NRC+meta	90.89	0.908	0.900	0.908	91.78	0.917	0.909	0.919

Table 3: Observations for govModel

Feature Combinations	without graph-based features				with graph-based features			
	Acc.(%)	**Prec.**	**Recall**	**F1**	**Acc.(%)**	**Prec.**	**Recall**	**F1**
TF-IDF+meta	90.25	0.902	0.900	0.898	**92.72**	**0.927**	**0.924**	**0.923**
LDA+meta	88.42	0.879	0.880	0.877	92.09	0.917	0.923	0.918
NRC+meta	86.91	0.875	0.848	0.858	92.73	0.927	0.919	0.921
TF-IDF+LDA+meta	89.06	0.887	0.882	0.883	92.33	0.920	0.925	0.920
TF-IDF+NRC+meta	88.97	0.892	0.883	0.885	92.80	0.923	0.911	0.914
LDA+NRC+meta	86.35	0.864	0.851	0.855	92.41	0.923	0.922	0.920
TF-IDF+LDA+NRC+meta	89.22	0.896	0.876	0.883	92.33	0.920	0.922	0.918

2. govAnnot: Here, the dataset is separated into two parts based on the speaker's affiliation: if the speaker presenting the motion belongs to the ruling government, then the motion polarity is positive, or otherwise negative.

The hyperparameters (for each of the feature sets and the classifier) were tuned using grid search. L-BFGS (Liu and Nocedal, 1989) was used for optimization in the neural network. Model training and evaluation was carried out using stratified 10-fold cross-validation. Stratification was performed to account for the slight imbalance in the dataset. Two types of labels are presented in the dataset: vote-based and manually-annotated. We use the manually-annotated labels for our experiments.
For the graph-based features, a grid search was performed which yielded the following parameters for generating embeddings:

- simGraph: p = 10, q = 1, walk length = 15, number of walks = 15, window size = 10. The feature vector obtained from these parameters yielded an accuracy of 79.51%.
- oppGraph: p = 0.1, q = 10, walk length = 5, number of walks = 10, window size = 10. The feature vector obtained from these parameters yielded an accuracy of 69.53%.

6 Results and Discussion

Table 2 and Table 3 present the results on the two models respectively. The values of accuracy, precision, recall, and F1-score are presented on feature sets with and without graph-based features.
In the case of both models, the usage of graph-based features outperforms the results obtained without using them. The difference is large in the case of the feature set comprising of LDA, NRC, and meta-features in the model with manually-annotated labels: the F1 scores obtained with and without graph features differ by 7.8%.
It can be observed that by using graph-based features The baselines for both have been surpassed by using graph-based features along with the other textual and meta-features. Our best results for manAnnot are obtained by using the combination of LDA, NRC, and graph-based features along with meta-features. The best results for govAnnot are obtained by using the combination of TF-IDF and meta-features along with graph-based features.

7 Conclusion

We presented a method for sentiment analysis of parliamentary debate transcripts, which could go a long way in helping determine the position an elected representative might assume on issues of great importance to the general public. The experiments were carried out on the Hansard parliamentary debates dataset (Abercrombie and Batista-Navarro, 2018b). We performed experiments on a variety of textual analysis methods (e.g. topic modeling, emotion classification, n-grams), and

combined them with community-based graph features obtained by representational learning on the dataset using node2vec. Our results surpass the state-of-the-art results using both govAnnot and manAnnot. Also, the F1 and accuracy values of the models using graph-based features are higher than those without graph-based features, the difference being considerable in some cases. This gives sufficient demonstration for the ability of representational learning to enhance performances on tasks like sentiment analysis.

8 Future Work

Future work in this area could involve the following aspects:

- Application of the proposed approach to tasks other than sentiment classification, for instance analysis of mental health and suicide ideation on social media.
- Constructing different graphs and analyzing other training and feature extraction methods for enhancing performance and deriving better inferences.
- Application of the proposed approach for analyzing data in different contexts; an example could be the analysis of the recently-conducted elections in India.
- Extend the proposed methodology to other problems (Mahata and Talburt, 2015; Mahata et al., 2015a,b) based on social media.

References

Gavin Abercrombie and Riza Batista-Navarro. 2018a. 'aye'or'no'? speech-level sentiment analysis of hansard uk parliamentary debate transcripts. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*, pages 33–40.

Gavin Abercrombie and Riza Theresa Batista-Navarro. 2018b. Identifying opinion-topics and polarity of parliamentary debate motions. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 280–285.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Paula Carvalho, Luís Sarmento, Jorge Teixeira, and Mário J Silva. 2011. Liars and saviors in a sentiment annotated corpus of comments to political debates. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 564–568. Association for Computational Linguistics.

Andrea Ceron, Luigi Curini, Stefano M Iacus, and Giuseppe Porro. 2014. Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens political preferences with an application to italy and france. *New media & society*, 16(2):340–358.

William Deitrick and Wei Hu. 2013. Mutually enhancing community detection and sentiment analysis on twitter networks. *Journal of Data Analysis and Information Processing*, 1(03):19.

Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864. ACM.

Kazi Saidul Hasan and Vincent Ng. 2013. Stance classification of ideological debates: Data, models, features, and constraints. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 1348–1356.

Amir Karami, London S Bennett, and Xiaoyun He. 2018. Mining public opinion about economic issues: Twitter and the us presidential election. *International Journal of Strategic Decision Sciences (IJSDS)*, 9(1):18–28.

Benjamin E Lauderdale and Alexander Herzog. 2016. Measuring political positions from legislative speech. *Political Analysis*, 24(3):374–394.

Dong C Liu and Jorge Nocedal. 1989. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528.

Euripides Loukis, Yannis Charalabidis, and Aggeliki Androutsopoulou. 2014. An analysis of multiple social media consultations in the european parliament from a public policy perspective.

Debanjan Mahata, Jasper Friedrichs, Rajiv Ratn Shah, and Jing Jiang. 2018. Detecting personal intake of medicine from twitter. *IEEE Intelligent Systems*, 33(4):87–95.

Debanjan Mahata and John R Talburt. 2015. A framework for collecting, extracting and managing event identity information from twitter. In *ICIQ*.

Debanjan Mahata, John R Talburt, and Vivek Kumar Singh. 2015a. From chirps to whistles: Discovering event-specific informative content from twitter. In *Proceedings of the ACM web science conference*, page 17. ACM.

Debanjan Mahata, John R Talburt, and Vivek Kumar Singh. 2015b. Identification and ranking of event-specific entity-centric informative content from twitter. In *International Conference on Applications of Natural Language to Information Systems*, pages 275–281. Springer.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Pushkar Mishra, Marco Del Tredici, Helen Yannakoudakis, and Ekaterina Shutova. 2018. Author profiling for abuse detection. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1088–1098.

Saif M Mohammad and Peter D Turney. 2013. Nrc emotion lexicon. *National Research Council, Canada*.

Jing Qian, Mai ElSherief, Elizabeth M Belding, and William Yang Wang. 2018. Leveraging intra-user and inter-user representation learning for automated hate speech detection. *arXiv preprint arXiv:1804.03124*.

Ludovic Rheault, Kaspar Beelen, Christopher Cochrane, and Graeme Hirst. 2016. Measuring emotion in parliamentary debates with automated textual analysis. *PloS one*, 11(12):e0168843.

Elena Rudkowsky, Martin Haselmayer, Matthias Wastian, Marcelo Jenny, Štefan Emrich, and Michael Sedlmair. 2018. More than bags of words: Sentiment analysis with word embeddings. *Communication Methods and Measures*, 12(2-3):140–157.

Rajiv Shah and Roger Zimmermann. 2017. *Multimodal analysis of user-generated multimedia content*. Springer.

Rajiv Ratn Shah, Anupam Samanta, Deepak Gupta, Yi Yu, Suhua Tang, and Roger Zimmermann. 2016a. Prompt: Personalized user tag recommendation for social media photos leveraging personal and social contexts. In *2016 IEEE International Symposium on Multimedia (ISM)*, pages 486–492. IEEE.

Rajiv Ratn Shah, Yi Yu, Suhua Tang, Shin'ichi Satoh, Akshay Verma, and Roger Zimmermann. 2016b. Concept-level multimodal ranking of flickr photo tags via recall based weighting. In *Proceedings of the 2016 ACM Workshop on Multimedia COMMONS*, pages 19–26. ACM.

Rajiv Ratn Shah, Yi Yu, Akshay Verma, Suhua Tang, Anwar Dilawar Shaikh, and Roger Zimmermann. 2016c. Leveraging multimodal information for event summarization and concept-level sentiment analysis. *Knowledge-Based Systems*, 108:102–109.

Stefan Stieglitz and Linh Dang-Xuan. 2012. Political communication and influence through microblogging – an empirical analysis of sentiment in twitter messages and retweet behavior. In *2012 45th Hawaii International Conference on System Sciences*, pages 3500–3509. IEEE.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 327–335. Association for Computational Linguistics.

Yaqin Wang and Haitao Liu. 2018. Is trump always rambling like a fourth-grade student? an analysis of stylistic features of donald trumps political discourse during the 2016 election. *Discourse & Society*, 29(3):299–323.

Investigating the effect of different carrier gases on plasma-assisted multi-walled carbon nanotube growth

U. Sharma[1], S. C. Sharma[1]

[1]*Department of Applied Physics, Delhi Technological University (DTU)*

Shahbad Daulatpur, Bawana Road, Delhi 110042, India.

E-mail: umangsharma2810@gmail.com, suresh321sharma@gmail.com

ABSTRACT

A model is developed consolidating the carbon nanotube (CNT) charging rate, the energy of all the plasma species, and the multi-walled carbon nanotube (MWCNT) growth rate owing to the surface and bulk diffusions and accumulation of particles on the catalyst nanoparticle surface. Using acetylene as the hydrocarbon source, we have compared the influence of distinctive carrier gases on the structure of CNT. The conclusion drawn from the results obtained were that while argon favors the growth of carbon nanotubes, ammonia and nitrogen contributes to diminishing the growth of carbon nanotubes. The work can be thought useful to serve to a better understanding of the growth of MWCNT in the plasma and to study their field emission applications.

I. INTRODUCTION

The plasma-aided CNT fabrication via plasma-enhanced chemical vapor deposition (PECVD) has many advantages over other techniques namely low-temperature growth, vertical and uniformly aligned growth, single-walled CNT (SWCNT) or multi-walled CNT (MWCNT), etc. [1-7]. Several groups have experimentally observed the effect of various carrier gases such as ammonia, nitrogen, argon on the growth of CNTs [2-7]. Through the developed theoretical model mentioned in Sec. II, we study the carrier gas effect on the MWCNT growth. The model consists of the dynamics of all the species present in the reactive plasma, and the CNT growth equation to examine the effect of carrier gases have on the MWCNT growth. The results are discussed in Sec. III and conclusion in Sec. IV.

II. MODEL

In the investigative model, following the detailed analysis done by Sharma *et al.* [8], we have considered a plasma composed of acetylene (C_2H_2) as the hydrocarbon source gas; ammonia (NH_3), nitrogen (N_2), and argon (Ar) are different carrier gases, respectively. The reactive plasma

consists of electrons, neutrals, and ions of acetylene denoted as type A, and carrier gas (namely NH_3, N_2, and Ar) denoted as type B in the present study. All the terms have their usual notations [8].

$$\partial_\tau(n_e) = \sum_j \beta_j n_j - \sum_j \alpha_j n_e n_{ij} - \gamma_e \frac{n_{CNT}}{\lambda_s}\left(\overset{tip}{\underset{e}{\xi}} + \overset{cyl}{\underset{e}{\xi}}\right) - \mathrm{K}^{e}_{wall} n_e, \quad (1)$$

The first term of Eq. (1) represents the rise in electron density via ionization of neutrals, the second term denotes decrease due to recombination of electron and ions, the third term represents decrease owing to their accumulation at the MWCNT surface, and the discharge loss to the walls of the chamber is denoted lastly.

$$\partial_\tau(n_{iA}) = \beta_A n_A - \alpha_A n_e n_{iA} - \frac{n_{CNT}}{\lambda_s}\left(\overset{tip}{\underset{iA}{\xi}} + \overset{cyl}{\underset{iA}{\xi}}\right) + \sum_{\ell AB} k_\ell n_A n_{iB} - \sum_{qBA} k_q n_B n_{iA} - \mathrm{K}^{iA}_{wall} n_{iA} + \left(\frac{P}{E^{diss}_{CH} V}\right), \quad (2)$$

$$\partial_\tau(n_{iB}) = \beta_B n_B - \alpha_B n_e n_{iB} - \frac{n_{CNT}}{\lambda_s}\left(\overset{tip}{\underset{iB}{\xi}} + \overset{cyl}{\underset{iB}{\xi}}\right) + \sum_{qBA} k_q n_B n_{iA} - \sum_{\ell AB} k_\ell n_A n_{iB} - \mathrm{K}^{iB}_{wall} n_{iB} + \left(\frac{P}{E^{diss}_{H} V}\right), \quad (3)$$

In Eqs. (2) and (3), first term denotes the elevates the ion density because of dissociation of neutrals, next term denotes the dissipation due to recombination of ions with electrons, third term represents the loss due to their accumulation at the MWCNT surface; the ion-neutral reactions are incorporated in the fourth and fifth terms, sixth terms denotes the discharge loss to the walls and the last term incorporates the dissociation by plasma power.

$$\begin{aligned}\partial_\tau(n_A) = {} & \alpha_A n_e n_{iA} - \beta_A n_A + \frac{n_{CNT}}{\lambda_s}\left(1-\gamma_{iA}\right)\left(\overset{tip}{\underset{iA}{\xi}} + \overset{cyl}{\underset{iA}{\xi}}\right) - \frac{n_{CNT}}{\lambda_s}\gamma_A\left(\overset{tip}{\underset{A}{\xi}} + \overset{cyl}{\underset{A}{\xi}}\right) \\ & + \mathrm{Q}^{IF}_A - \mathrm{Q}^{OF}_A - \sum_{lBA} k_l n_A n_{iB} + \sum_{qAB} k_q n_B n_{iA} - J^{adsp}_A + J^{desp}_A - \mathrm{K}^{A}_{wall} n_A, \end{aligned} \quad (4)$$

$$\begin{aligned}\partial_\tau(n_B) = {} & \alpha_B n_e n_{iB} - \beta_B n_B + \frac{n_{CNT}}{\lambda_s}\left(1-\gamma_{iB}\right)\left(\overset{tip}{\underset{iB}{\xi}} + \overset{cyl}{\underset{iB}{\xi}}\right) - \frac{n_{CNT}}{\lambda_s}\gamma_B\left(\overset{tip}{\underset{iB}{\xi}} + \overset{cyl}{\underset{iB}{\xi}}\right) \\ & + \mathrm{Q}^{IF}_B - \mathrm{Q}^{OF}_B - \sum_{qBA} k_q n_B n_{iA} + \sum_{\ell AB} k_\ell n_A n_{iB} - J^{adsp}_B + J^{desp}_B + J_{thd} - \mathrm{K}^{B}_{wall} n_B, \end{aligned} \quad (5)$$

In Eqs. (4) and (5), first term descibes the rise in neutral density due to the ion-electron recombination, next term denotes the dissipation due to dissociation of neutrals the rise in neutral atom density due to ion neutralizationis descibed in third term, the fourth term characterizes the dissipation due to neutral atom accumulation at the MWCNT surface, the neutral plasma species

influx and outflux is denoted by the fifth and sixth terms, the neutral-ion reactions are incorporated in the seventh and eighth terms, ninth and tenth term represents the adsorption and desorption of neutrals onto/from the catalyst. The eleventh term of Eq. (5) indicates an increase in hydrogen species density due to the thermal de-hydrogenation process. The ending terms both equations signify the discharge loss.

$$\partial_\tau \left(N_g \right) = \left[D_s + D_b + 2\pi\kappa v r^2 \right] \times n_C S - n_{cg} S, \qquad (6)$$

The first term of Eq. (6) signifying the rate of cylindrical graphitic shell formation [9] denotes surface diffusion over catalyst surface, the second term denotes bulk diffusion through the catalyst, the third term incorporates the C-atom precipitation around the catalyst rear end forming cylindrical shells where n_C is the number density of carbon species on the catalyst nanoparticle [8] and, the last term denotes etching effects of the carrier gas species where *S* is the surface area of catalyst nanoparticle.

III. RESULTS AND DISCUSSION

The Eqs. (1-6) have been solved simultaneously for a typical glow discharge plasma parameter with the initial boundary conditions at $\tau = 0$ as mentioned in the work of Sharma *et al.* [8]. The variation of the carbon number density and number of graphitic shells for the MWCNT with time for different carrier gases is illustrated in Figs. 1 and 2, respectively. The observations drawn from the results are discussed alongside.

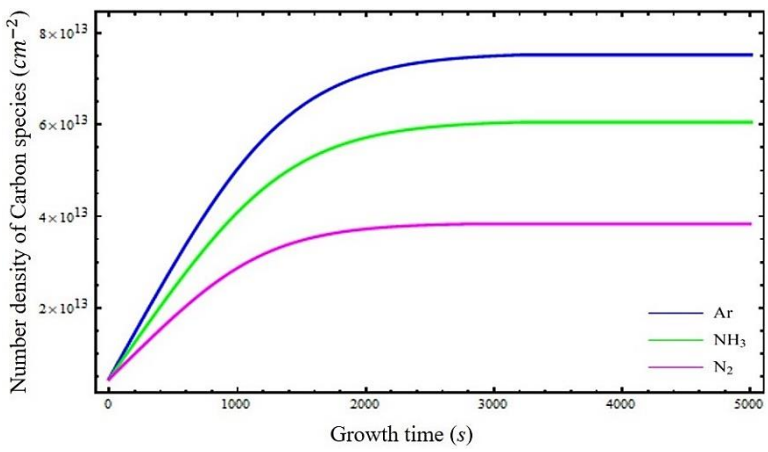


Figure 1. Time progression of carbon number density on catalyst particle.

The dissociation of hydrocarbon species, increases the number density of carbon species on the catalyst nanoparticle. This Carbon number density is higher for Ar compared to NH_3 and N_2, hence it would offer better MWCNT growth compared to others, as shown in Fig 1.

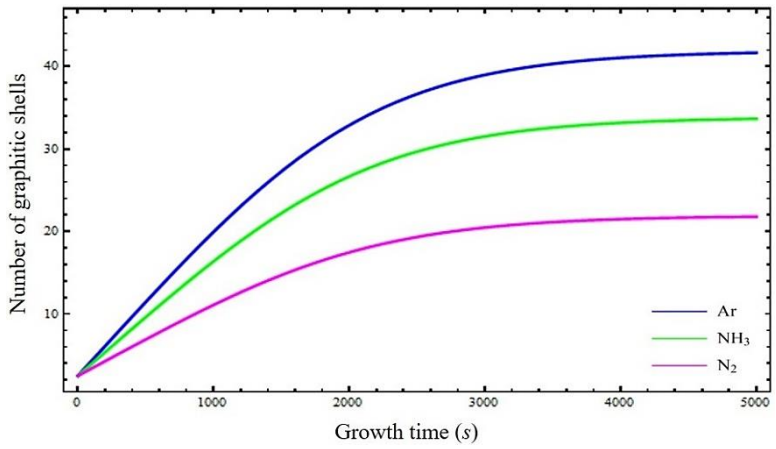


Figure 2. Time progression of graphitic shells of MWCNT.

As in case of Ar, the carbon number density is higher, so the number of cylindrical graphitic shells formed in this case are more compared to NH_3 and N_2, as shown in Fig 2. This increase in number of shells increases the radius of MWCNT, consequently, Ar offers a better MWCNT growth than NH_3 and N_2.

IV. CONCLUSION

The dynamics of the plasma species and MWCNT growth has been studied using the developed model incorporating the effects of various gases in plasma-assisted CNT growth process. The effect of gases; ammonia (NH_3), nitrogen (N_2), and argon (Ar) on the growth of the MWCNT are investigated. These results are in compliance with various experimental results [2-7]. The study can be proved valuable in the understanding of the use of CNT as field emitter devices.

ACKNOWLEDGEMENT

The authors are thankful to the Department of Science and Technology (DST), Government of India for providing the financial assistance under INSPIRE Fellowship.

REFERENCES

[1] S. Iijima, Nature **354,** 56 (1991).

[2] V. I. Merkulov, A. V. Melechko, M. A. Guillorn, D. H. Lowndes, and M. L. Simpson, Appl. Phys. Lett. **79,** 2970 (2001).

[3] W. Mi, J. Y. Lin, Q. Mao, Y. Li, and B. Zhang, J. Natural Gas Chem. **14,** 151 (2005).

[4] S. M. Toussi, A. Fakhru'l-Razi, A. L. Chuah, and A. R. Suraya, Sains Malays. **40,** 197 (2011).

[5] M. Jung, K. Y. Eun, J.-K. Lee, Y.-J. Baik, K.-R. Lee, and J. W. Park, Diamond Relat. Mater. **10,** 1235 (2001).

[6] W. Mi, J. Y. Lin, Q. Mao, Y. Li and B. Zhang, J. Nat. Gas Chem. **14,** 151 (2005).

[7] V. Kayastha, Y. K. Yap, S. Dimovski, and Y. Gogotsi, Appl. Phys. Lett. **85,** 3265 (2004).

[8] U. Sharma and S. C. Sharma, Physics of Plasmas **25,** 103509 (2018).

[9] R. Gupta, N. Gupta and S. C. Sharma, Physics of Plasmas **25,** 043504 (2018).

Lower hybrid waves excitation by a relativistic electron beam in a magnetized dusty plasma: kinetic theory

Anshu[1], S. C. Sharma[1], J.Sharma [2]

[1] *Department of Applied Physics, Delhi Technological University (DTU)*

Shahbad Daulatpur, Bawana Road, Delhi 110042, India.

[2]*Amity School of Applied Sciences, Amity University Haryana, Gurgaon-122413, India*

Email: anshudahia@gmail.com, suresh321sharma@gmail.com, plasmajyoti@gmail.com

ABSTRACT

In the present paper, we study the kinetic theory of instability of electrostatic lower hybrid waves[1] (LHWs) in dusty plasma containing positive potassium ions (K+). A relativistic electron beam (REB) propagating through the plasma medium containing dust grains, positive potassium ions, and electrons, drives the LHWs to instability via Cerenkov interaction. Analytical expressions and numerical calculations using Vlasov equations [2] have been carried out for evaluating the growth rate of LHWs for existing experimental parameters. It is found that the rate of growth of the instability increases as the relative density ratio of dust grains increases. Analytical results comply with some of the experimental observations.

INTRODUCTION

Lower Hybrid Waves (LHWs) have been the topic of several experimental, hypothetical and numerical investigations in the plasma because of their low-frequency range.[3-4] LHWs in contrast to other wave modes like magnetohydrodynamic waves, surface plasma waves, etc. can interact resonantly with ions along with electrons. These plasma waves oscillate longitudinally because of ions present in the magnetized plasma system. The propagation direction of these waves is at the right angle to the stationary magnetic field. The presence of dust grain changes the properties of the plasma system. The dust grain charge variations are the reason for the damping and growth of the LHWs. Merlino *et al.*[5] have examined experimentally and theoretically, the outcomes of electrostatic waves of low-frequency comprising dust grains which are negatively charged in a plasma.

In this paper, we study the LHWs excitation using a kinetic theory model in a dusty magnetized plasma by a relativistic electron beam. The Vlasov approach has been followed up to examine

the response of an electron beam. An electron beam interacts via Cerenkov interaction with the plasma medium and drives LHWs to instability.

MODEL

We consider a plasma filled system which comprises of electrons and ions in the existence of a static magnetic field B_0 in the axial direction. A low-frequency LHWs propagates in the x-y plane which must be perpendicular to the magnetic field direction, and a relativistic electron beam also propagates in the y-direction. The Perturbed electrostatic potential is taken into account as

$\phi = \phi_l \exp[-i(\omega_l t - k_l x)]$, where ω_l *and* k_l are the frequency and wave vector, respectively of the electrostatic wave.

Using the Vlasov equation, the density perturbed term can be evaluated using the expression

$$\frac{\partial F}{\partial t} + \sum \frac{\partial}{\partial r_j} \times \dot{r}_j F = 0,$$

where F, r and $\dot{r}_j$ are the distribution function, phase space and phase time derivative, respectively.

The density perturbed term can be simplified by using this equation

$n = \frac{m_\alpha^2}{\omega_c^\alpha} \iiint f_{md} d^3\vec{v}$, where m_α is the mass, $\alpha =$ b, i ,e and d denotes beam, ion, electron, and dust term resp., $\omega_c^\alpha \left(= \frac{eB_0}{m_\alpha c} \right)$ is the cyclotron frequency, f_{md} is the Maxwellian distribution function and v denotes the frequency.

On substituting the perturbed density in the poison's equation, the growth rate expression is obtained as

$$\Gamma = \omega_p^\alpha \left[\left\{ I_n^2 e^{-2a_\alpha} \left(X^2 + \pi Y^2 e^{-2\zeta^2} \right) + \left(1 - 2I_n e^{-a_\alpha} X \right) \right\}^{1/2} - \left(1 + I_n(a_\alpha) e^{-a_\alpha} X \right) \right]^{1/2}, \quad (1)$$

where $X = \frac{\omega_e + \omega_e^*}{\omega - kv_e} + \frac{\omega_i - \omega_i^*}{\omega - n\omega_c^i} + \frac{\omega_b + \omega_b^*}{\omega - kv_b}$

$$Y = \frac{\omega_e + \omega_e^*}{\omega_e} + \frac{\omega_i - \omega_i^*}{\omega_i} + \frac{\omega_b + \omega_b^*}{\omega_b} + 1$$

$\omega_p^{\alpha 2}\left(=\frac{4\pi n_\alpha^0 e^2}{m_\alpha}\right)$ is the plasma frequency and $\omega_\alpha^*\left(=\frac{kv_\alpha{}^2}{2\omega_c^\alpha}\right)$ is the diamagnetic drift. $a_\alpha\left(=\frac{k^2 v_\alpha{}^2}{2\omega_c^{\alpha 2}}\right)$ is the drift frequency, assuming $\left(\frac{\omega_l - n\omega_c^\alpha}{kv_\alpha}\right) = \xi,$ and I_n is a modified Bessel's function.

RESULTS AND DISCUSSION

In the present computation, we have taken ion density $n_i = 1\times 10^9 cm^{-3}$, electron density $n_e = 0.8\times 10^9 cm^{-3}$ and dust grain mass $m_d = 10^{12} m_p$, respectively. The relative density ratio (δ) has been varied from 1 to 10. Fig.1 depicts the variation of growth rate $\Gamma\left(\sec^{-1}\right)$ with density ratio $\delta\left(=\frac{n_i}{n_e}\right)$ taking dust grain size, i.e. $a = 10^{-4}$ cm .

The growth rate increases as the relative density ratio increases. The growth rate scales as one-sixth power to the dust grain size as seen from equation (1). The growth rate of the unstable mode in the presence of dust charge stabilizes more as the dust grain size decreases with the relative density ratio of the negatively charged dust grain $\delta\left(=\frac{n_i}{n_e}\right)$. Our results are in good agreement with the Barkan et al. experimental observations[6].

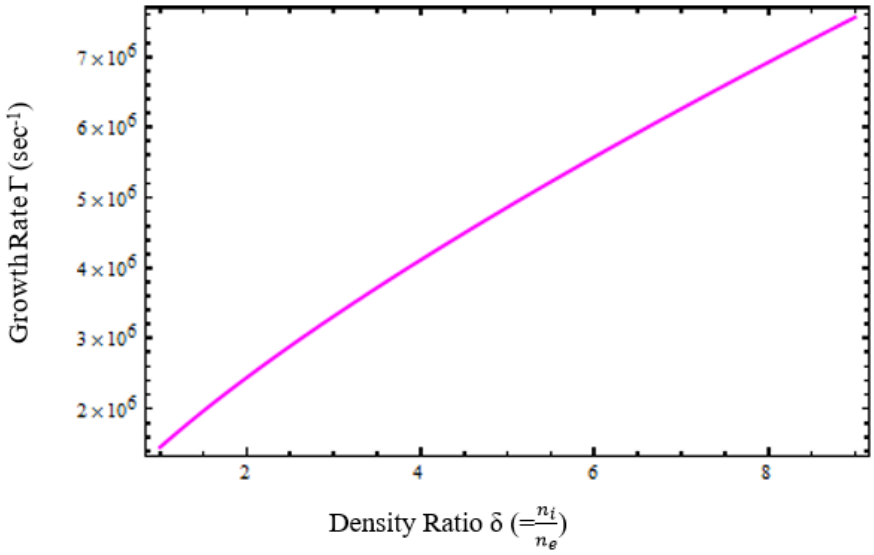


Fig.1 Growth rate $\Gamma\left(\sec^{-1}\right)$ variation with the density ratio of ions and electrons (δ) of lower hybrid waves.

CONCLUSION

The lower hybrid wave excitation has been studied in a magnetized dusty plasma via Cerenkov interaction by a relativistic beam of the electron. The growth rate of unstable mode augments with the relative density of negatively charged dust grains.

ACKNOWLEDGEMENT

The authors are thankful to SERB-DST New Delhi for financial assistance VID Grant No: EMR/2016/002699.

REFERENCES

1. V. Prakash, Vijayshri, S. C. Sharma, and R. Gupta, Physics of Plasmas **20**, 053701 (2013).

2. C. S. Liu and V. K. Tripathi, Physics Reports **130**, 143 (1986).

3. J. Sharma, S. C. Sharma, V. K. Jain, and A. Gahlot, Phys. Plasmas **20**, 033706 (2013).

4. K. Papadopoulos and P. Palmadesso, Physics of Fluids **19**, 605 (1976).

5. R. L. Merlino, A. Barkan, C. Thompson, and N. D'Angelo, Phys. Plasmas **5**, 1607 (1998).

6. A. Barkan, N. D'Angelo, and R. L. Merlino, Planet. Space Sci. **43,** 905 (1995).

Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Mechanical suitability of carbon fiber based composite for chopart prosthetic socket

To cite this article: F H Abdulsadah *et al* 2019 *J. Phys.: Conf. Ser.* **1240** 012154

View the article online for updates and enhancements.



IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 14.139.251.100 on 13/08/2019 at 11:31

IOP Conf. Series: Journal of Physics: Conf. Series **1240** (2019) 012154 doi:10.1088/1742-6596/1240/1/012154

Mechanical suitability of carbon fiber based composite for chopart prosthetic socket

F H Abdulsadah[1,3], Q Murtaza[2], F Hasan[1] and S Bhardwaj[1]

[1]Department of Mechanical Engineering, Aligarh Muslim University, Aligarh, India
[2]Department of Mechanical Engineering, Delhi Technological University, Delhi, India.
[3]Correpnding author: falahsci@gmail.com

Abstract. Material selection has important role in the design of prosthetics. Presently, polyethylene (PE) has been majorly used in the manufacture of chopart prosthetic socket due to it mouldability and low cost, but suffers from large weight to strength ratio. In the present study, carbon fiber composite (CF) is compared with PE for its usability in chopart prosthetic sockets. 3 perlon layers-2 carbon fiber layer-3 perlon layers (3P-2C-3P) lamination was developed and tested for the mechanical properties. Later, the tested mechanical properties of PE and CF were used in finite element modelling to analyze stresses and deformation in the Chopart prosthetic socket. The study showed that use of CF in chopart prosthetic sockets reduces stress by 70.6% as compared to PE, and could be suggested as mechanically superior material in design of chopart prosthetic socket compared to PE.

1. Introduction

Ability to perform ambulation measures the independency in living and over all wellbeing of the person [1]. The need of performing an amputation suffocates the independency in performing day to day activities [2]. According to the United States National Health Survey in 1996, it was found that 1.27 million people in America were living with lower limb loss, and this number is estimated to rise by 3.6 million by 2050 [3]. Partial foot amputation (PFA) is the most common kind of amputation suffered by patients. Lisfranc and chopart amputations are such PFA that were introduced in 19th century by French surgeons. Lisfranc is disarticulation of metatarsal joint, while chopart shortens the foot by removing mid and forefoot of the patient [4,5]. Besides comfort studies and ease of use of prosthetics design [6], finite element (FE) modelling has become an important tool for material selection and design analysis of such prosthetics. FE analysis has been used to study polypropylene (PP) based ankle foot orthosis (AFO) for drop foot patients [7]. Biomechanics together with FE modelling was earlier used to study impact of shoe soles in rehabilitation [8]. FE models have been developed to study tarsal bone degeneration in Hansen's disease [9]. A dynamic FE analysis of human foot was implemented with explicit dynamics with assumptions to be isotropic, linearly elastic and homogeneous [10].

In general, the polymeric material used in manufacturing prosthetics socket consist of polymers, Polyamide, Polyester, Copolyester, Polypropylene, Polyethylene, Polyphenylene sulfide, Polyvinylidene fluoride [11]. Polypropylene has been a much sough much sought material for the manufacture of prosthetic sockets owing to the low cost. However, studies have shown the loss of mechanical properties of polypropylene with time. Furthermore, manufacturing of polypropylene is associated with the risk for skin allergies, allergic bronchitis which may even lead to lung cancer in

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

manufacturing personals [12]. In recent time carbon fiber and glass fibers composites, owing to their better mechanical properties, have gained interest as a manufacturing material for prosthetic sockets [13].

The purpose of the present study is to evaluate carbon fiber composite as a potential material for the chopart socket prosthetic in terms of mechanical properties. Carbon fiber composite based socket is compared with the much popular polyethylene based socket using FE modelling.

2. Materials and Method

2.1. Estimation of Mechanical Properties

Test specimens for polyethylene (PE) and carbon fiber composite (CF) were fabricated for estimating the mechanical properties which was later used for finite element analysis.

2.1.1. *Carbon fiber composite*. The CF composite tested in the present study consisted of 2 layers of CF (Ottobock healthcare 616G15) sandwiched between 3 layers of perlon stockinet (Ottobock healthcare 623T3). 80:20 polyurethane and c-orthocryl lamination resin was used as an adhesive for the composite.

2.1.2. *Fabrication of specimens.* Vacuum molding with gypsum molds was used for the fabrication of CF composite lamination. Polyvinyl-alcohol (PVA, Ottobock healthcare 99B71) was used to cover the mold and lamination. Lamination resin and hardener was used in the standard ratio i.e. for each 100 part of acrylic resin mixed 2/3rd part of hardener. Constant vacuum of 35 mm of Hg was maintained until the composite materials come to room temperature. Three tensile and flexural test specimens were then machined according to ASTM D638 and ASTM D790 respectively [14]. Also, three PE specimens were machined from the 4 mm PE sheet for testing. Hence, the total 12 specimens were tested for tensile and flexural strength. Figure1 shows the tested specimens and the test machine.

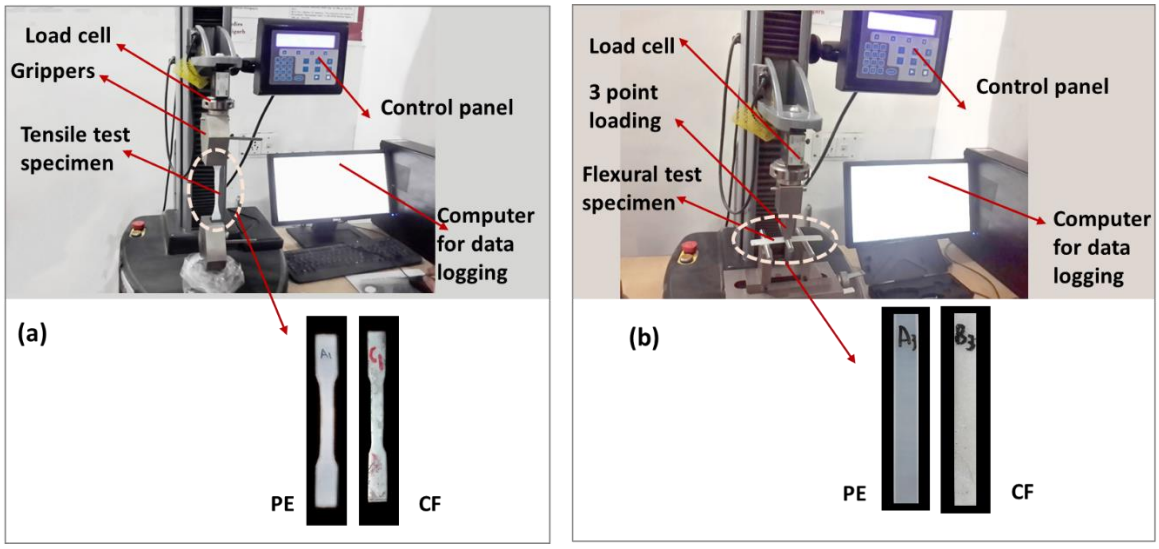


Figure 1. Testing setup and specimen for PE and CF for (a) Tensile test and (b) Flexural test

2.2. Finite Element (FE) Analysis

2.2.1. 3D Model Generation

Digital prototyping for the chopart socket was done using 3D scanning. First, cloud point of full scale human lower limb was obtained using 3D scanner (RangeVision smart). The point cloud was later imported in Autodesk Meshmixer for developing the solid model. The solid model file was then used to design the chopart socket using Autodesk Inventor as shown in figure 2.

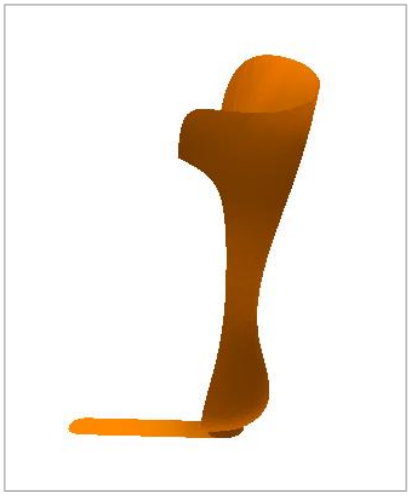


Figure 2. Chopart socket 3D model designed from full scale model of human leg.

2.2.2. FE model setup

The model was analyzed for stresses and deformation while 75 N force was applied at the chopart socket. The foot of the prosthetic was given the fixed constrain. The FE analysis was done in ANSYS (2016). The total number of mesh elements were 39744 with 61879 number of nodes. Figure 3 shows the ANSYS model. Von Misses theory of failure was assumed for the analysis. The analysis was done for the two materials considered in the study i.e. PE and CF.

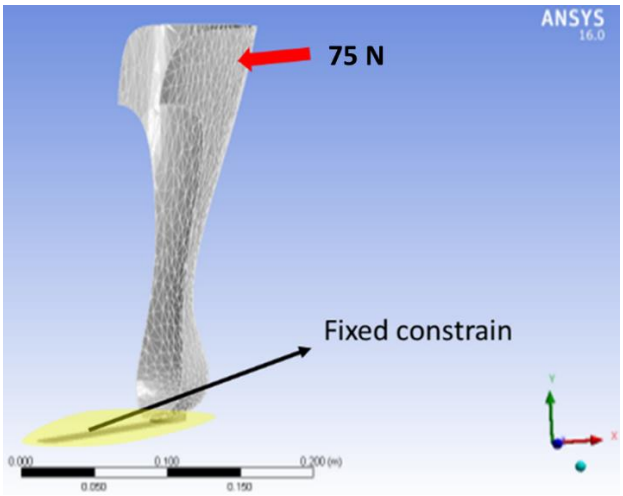


Figure 3. Generation mesh of prosthetics socket

3. Results and discussion

3.1. Experimental Results

The mechanical properties (σ_y, σ_{ult} and E) as obtained from the test specimens is shown in Table 1. These properties were later used as input data in the ANSYS to simulate the prosthetic socket and analyse the stresses with the tested load condition.

Table 1. The mechanical properties of the material tested.

Materials	Yield Stress σ_y (MPa)	Ultimate Stress σ_{ult} (MPa)	Young's Modulus E (GPa)
PE	18	22	1.9
CF	22.5	27	2.2

3.2. Simulation Results

The model was tested for the deformation and stresses for two materials considered in the study. Maximum deformation of 43 mm was found for PE compared to 11 mm for CF. The equivalent Von-Mises Stresses were evaluated. The highest stress value was 11.6 MPa in the socket made from PE, while the highest stress for CF based socket was 3.4 MPa. Heal was found to be the most stress concentrated region. Figure 4 shows the result from FE analysis.

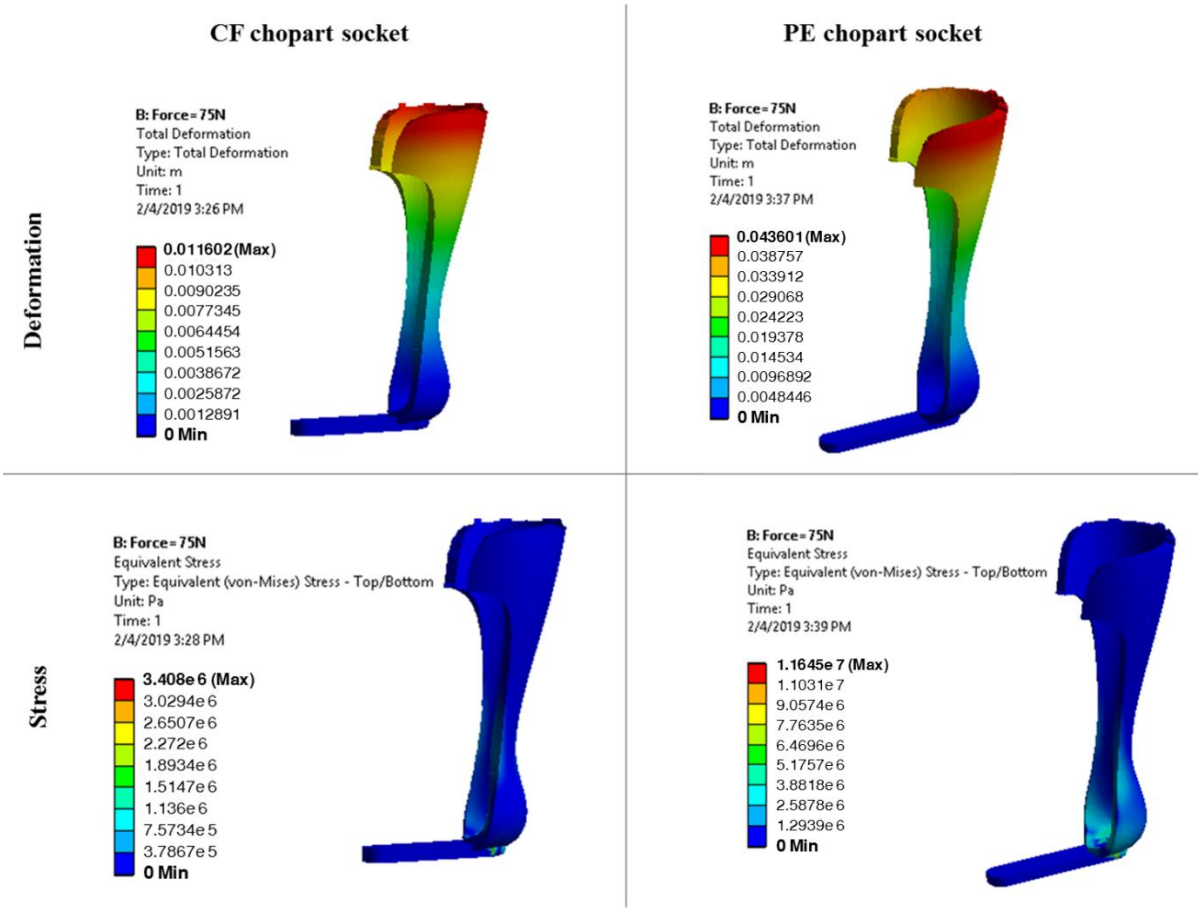


Figure 4. Total deformation and equivalent stresses for CF and PE chopart socket

While both of the CF and PE prosthetic socket gave a safe factor of safety, the CF socket has maximum equivalent stress of 3.4 MPa which was 70.6% lesser than the equivalent peak stress in PE socket. Hence, the CF socket can also be tested for lesser thickness at the heal as compared to PE socket. CF could be used a light weight alternative to reduce the physiological cost of the patient using the chopart socket. It is for this reason that CF is desired for the manufacture of prostheses, regardless of its high cost [15].

4. Conclusions

The goal of the current study was to access the suitability of carbon fibre composite in the design of chopart socket. From the mechanical tests, CF was found to have 20.4% and 18.5% high yield stress and ultimate respectively as compared to PE. Simulation study showed decrease in equivalent stress by 70.6% compared to PE. Also, CF provide significantly higher strength to weight ratio than PE. The simulated results indicated the suitability of CF based composite as a mechanically suitable choice for chopart prosthetics socket design.

References

[1] Atnip R G 2005 Toe and partial foot amputations *Oper. Tech. Gen. Surg.* **7** 67–73

[2] Owings M F and Kozak L J 1998 Ambulatory and inpatient procedures in the United States, 1996. *Vital Health Stat. 13.* 1–119
[3] Ziegler-Graham K, MacKenzie E J, Ephraim P L, Travison T G and Brookmeyer R 2008 Estimating the Prevalence of Limb Loss in the United States: 2005 to 2050 *Arch. Phys. Med. Rehabil.* **89** 422–9
[4] Bowker J H 2007 Partial Foot Amputations and Disarticulations: Surgical Aspects *Jpo J. Prosthetics Orthot.* **19** P62–76
[5] Dillon M P and Barker T M 2006 Can Partial Foot Prostheses Effectively Restore Foot Length? *Prosthet. Orthot. Int.* **30** 17–23
[6] Janisse D J and Janisse E J 2010 Shoes, Orthoses, and Prostheses for Partial Foot Amputation and Diabetic Foot Infection *Foot Ankle Clin.* **15** 509–23
[7] Hassan F, Murtaza Q and Bhardwaj S 2017 Manufacture, Design and Structural Analysis of a Custom Ankle Foot Orthoses on Foot Drop Patients Using 3D Scanning *Int. J. Sci. Eng. Res.* **8** 201–3
[8] Lemmon D, Shiang T Y, Hashmi A, Ulbrecht J S and Cavanagh P R 1997 The effect of insoles in therapeutic footwear--a finite element approach. *J. Biomech.* **30** 615–20
[9] Jacob S and Patil M K 1999 Three-dimensional foot modeling and analysis of stresses in normal and early stage Hansen's disease with muscle paralysis. *J. Rehabil. Res. Dev.* **36** 252–63
[10] Qian Z, Ren L, Ding Y, Hutchinson J R and Ren L 2013 A Dynamic Finite Element Analysis of Human Foot Complex in the Sagittal Plane during Level Walking ed J Langowski *PLoS One* **8** e79424
[11] Kumar S 2014 Selective Laser Sintering/Melting *Compr. Mater. Process.* 93–134
[12] OSHA 1977 *Criteria for a recommended standard: occupational exposure to fibrous glass.*
[13] Phillips S L and Craelius W 2005 Material Properties of Selected Prosthetic Laminates - Journal of Prosthetics and Orthotics, 2005 | *J. Prosthetics Orthot.* **17** 27–32
[14] Tanikella N G, Wittbrodt B and Pearce J M 2017 Tensile strength of commercial polymer materials for fused filament fabrication 3D printing *Addit. Manuf.* **15** 40–7
[15] Warder H H, Fairley J K, Coutts J, Glisson R R and Gall K 2018 Examining the viability of carbon fiber reinforced three-dimensionally printed prosthetic feet created by composite filament fabrication *Prosthet. Orthot. Int.* **42** 644–51

Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Microstructural and wear characterisation of as-cast AS21A magnesium alloy

To cite this article: Sumit Joshi *et al* 2019 *J. Phys.: Conf. Ser.* **1240** 012127

View the article online for updates and enhancements.



IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 14.139.251.100 on 13/08/2019 at 11:37

IOP Conf. Series: Journal of Physics: Conf. Series **1240** (2019) 012127 doi:10.1088/1742-6596/1240/1/012127

Microstructural and wear characterisation of as-cast AS21A magnesium alloy

Sumit Joshi[1*], Ramesh Chandra Singh[1], Rajiv Chaudhary[1]

[1]Department of Mechanical Engineering, Delhi Technological University, Bawana Road, New Delhi-110042, India

*Email: sumitrke1989@gmail.com

Abstract. The present study consists of wear investigations performed on AS21A magnesium alloy. AS21A alloy falls in the category of Mg-Al-Si alloy system which is recommended for high-temperature applications. The condition of the studied alloy was as-cast and annealed. Microstructural features of AS21 alloy comprised of a soft α-Mg matrix reinforced with hard and brittle coarse Mg_2Si intermetallic. Wear behaviour of the alloy experimented through pin-on-disk apparatus and at varying loads from 10-40 N. It was observed that the wear resistance and Coefficient of Friction (COF) decrease with an increase in applied load. Further, the presence of Mg_2Si intermetallic in the magnesium matrix had a significant effect on the wear properties of the AS21A alloy.

Keywords: AS21A; Mg-Al-Si; Mg_2Si; intermetallic; wear

1. Introduction

In the last few decades, the strict government norms on emission encouraged the demand for lightweight materials in the aerospace and automotive sectors. Magnesium alloys which are lightest among structural materials have become a suitable choice in commercial industries. Magnesium alloys have earned a reputation because of the attributes such as high specific strength, excellent castability, low density, high thermal and electrical conductivity and shock absorption capability [1-3].

Despite remarkable bulk properties, magnesium alloys are inferior in surface properties like wear and corrosion resistance. The literature concerning wear behaviour of magnesium alloys is quite a few and limited to Al-Zn (AZ) [4], Al-Mn (AM) [5] and Rare Earth [6] series. Chen and Alpas [4] reported the dry sliding wear behaviour of AZ91 alloy subjected to a load range of 1–350 N and sliding velocity range of 0.1–2.0 m/s. They observed a mild wear regime consisting of an oxidational and delamination regime while the severe regime comprising of severe plastic deformation and a melt regime. Similarly, Taltavull et al. [5] and Lopez et al. [6] reported a similar wear mechanism corresponding to different load and sliding velocity during the wear study of AM60B and ZE41A alloy respectively.

The Mg-Al-Si (AS series) is another potential alloy for structural applications because of in-situ synthesised hard and brittle thermally stable Mg_2Si phase [7]. Mg_2Si exhibit attributes such as high melting point (1085 ^{0}C), low density (1.99×10^3 kg/m^3), high elastic modulus (120 GPa), high hardness (460 HV) and low coefficient of thermal expansion (7.5×10^{-6} /K) [8,9]. The study on Mg-Al-Si alloys is limited to microstructure modification and synthesis. There is limited literature [10,11] available

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

regarding the wear characterisation of Mg-Al-Si based alloys. Akyuz et al. [10] studied the behaviour of AS series magnesium alloys and reported that the Mg_2Si presence in the alloy had a strong influence on the wear and machining properties. Ajith Kumar et al. [11] reported the reduction in wear rate of Mg–Si based alloys with an increase of Si reinforcement and with the load. The primary aim of the past studies was to study the effect of the varying composition of Aluminium [10] and silicon [11] on the wear characterisation of magnesium alloys. However, there is no literature available specifically dedicated to AS21 alloy. In the present experiment, the microstructural and wear characterisation were performed on as-cast AS21 alloy. The aim was to know the effect of intermetallic on the mechanical and wear properties of AS21A alloy.

2. Material and methods

The material selected for the present study was as-cast AS21A magnesium alloy fabricated through gravity die casting in the form of a plate having 15 mm thickness. The as-received casting plate was machined in conventional machines along with the plate softening for further investigations. The casting plate was softened to mitigate the effect of casting defects. It was achieved through heating followed by atmospheric cooling. Figure 1 depicts the as-cast and machined surface of the studied alloy. The spectroscopy analysis of AS21A carried out as per ASTM B94-2013 for chemical composition of the alloying elements is presented in table 1. The chemical analysis was performed using ICP-OES (Inductively Coupled Plasma-Optical Emission Spectroscopy) spectrometer, Model: Spectro Max, Made in Germany.

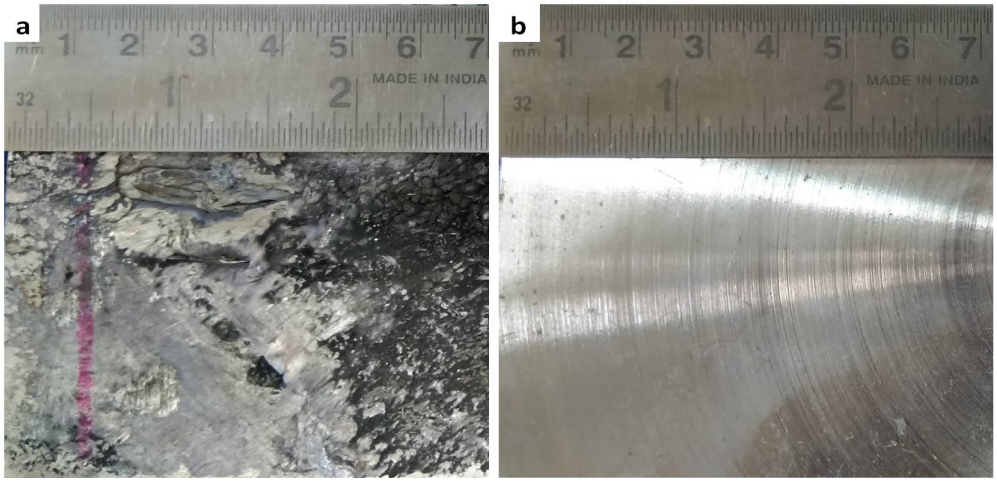


Figure 1. The surface of AS21A alloy casting plate (a) as received; (b) machined.

Table 1. The chemical composition of as-cast AS21A alloy

Alloy	Al	Si	Mn	Zn	Cu	Fe	Ni	Mg
Percentage	2.12	1.04	0.221	0.058	0.002	0.003	<0.001	Rest

The samples for characterisation were taken from AS21A casting plate through CNC wire cut. The AS21A microstructural features revealed through dry and wet polishing followed by etching. The etchant solution employed was acetic-picral comprised of ethanol (100 ml), picric acid (6 g), acetic acid (5 ml) and water (10 ml) for 5-10 seconds. The Optical Microscope (Olympus, model GX 41) was utilised to capture the Optical Microscopic (OM) images of the microstructure. Further, Scanning Electron Microscope (SEM) equipped with EDS facility was utilised for micrographs and elemental analysis. The microhardness analysis was carried out using a microhardness tester attached with

Vickers diamond pyramidal indenter (136º) at a load of 100 g and a dwell time of 10 seconds. The tensile test was performed on Universal Testing Machine (UTM) set up (Tinius Olsen H50KS) at a constant crosshead speed of 1 mm/min and room temperature. Figure 2 shows tensile specimens prepared as per ASTM B557M-06. SEM micrographs of the fractured samples were also taken for failure characteristics. Wear investigation was done on the samples shown in figure 2 as per ASTM G99 standard. Pin of 8 mm diameter was used as top specimen while EN-24 steel counter disc (hardened to 55-60 HRC) as a bottom specimen. The top and bottom specimen surface roughness (R_a) measured through was nearly 100 µm and 0.2 µm respectively. The surfaces of pin samples were polished with 1200 grit emery before performing wear test. The Pin-on-Disk apparatus (TR-20, Ducom) depicted in figure 3 was used for the study of wear behaviour. The parameters considered during wear test were: Load ranging from 10 N-40 N in steps of 10 N, the constant sliding velocity of 1 m/s and sliding distance of 2500 m. Wear rate (mg/m) was calculated as the ratio of weight loss (mg) and sliding distance (m) for all three samples. Further, weight loss was measured through precision weighing balance (0.001g) as the difference in the weight of the sample before and after wear. Wear analysis was done by utilising SEM micrographs to understand the wear mechanism involved.

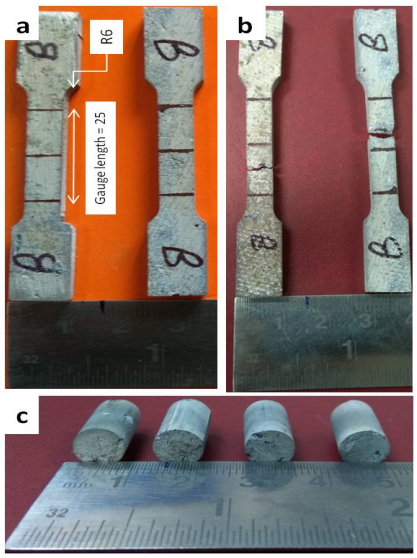


Figure 2. Characterisation specimens for: (a) tensile test before fracture; (b) tensile test after fracture; (b) wear test.

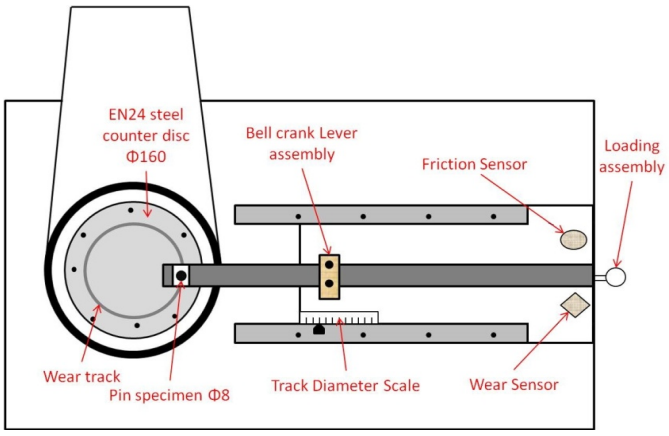


Figure 3. Schematic of Pin-on-Disk equipment.

3. Result and discussions

3.1. Microstructure characterisation

The as-cast AS21A alloy microstructure depicted in figure 4 typically comprised of soft α–Mg matrix surrounded by hard and brittle eutectic Mg_2Si intermetallic compound. The morphology of Mg_2Si particles was found to be coarse and Chinese script. Further, the FESEM of AS21A alloy shown in figure 5a also depicted the presence of intermetallics in the soft magnesium matrix. The Mg_2Si nature of intermetallic was confirmed by the EDS spectrum as shown in figure 5b. The Mg_2Si behaved as the in-situ reinforcement in the soft α-magnesium matrix thus enhancing the overall mechanical and wear properties of as-cast AS21A alloy. The in-situ reaction during solidification of casting was found responsible for the formation of the Mg_2Si phase. The microstructural findings of AS21A alloy were consistent with the available literature [12-14].

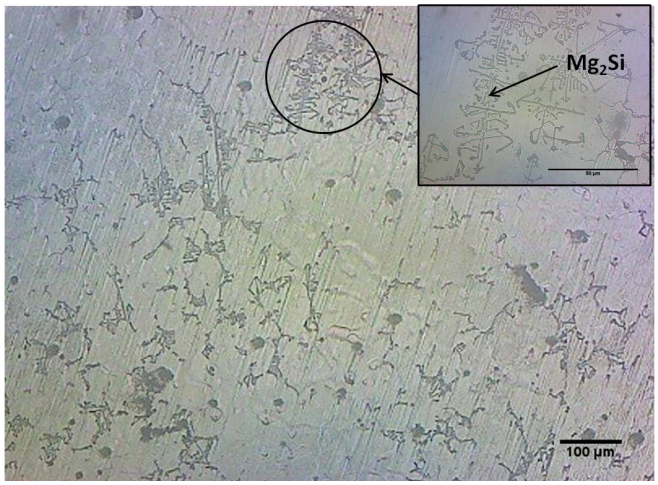


Figure 4. Optical Microscopy image of as-cast AS21A alloy

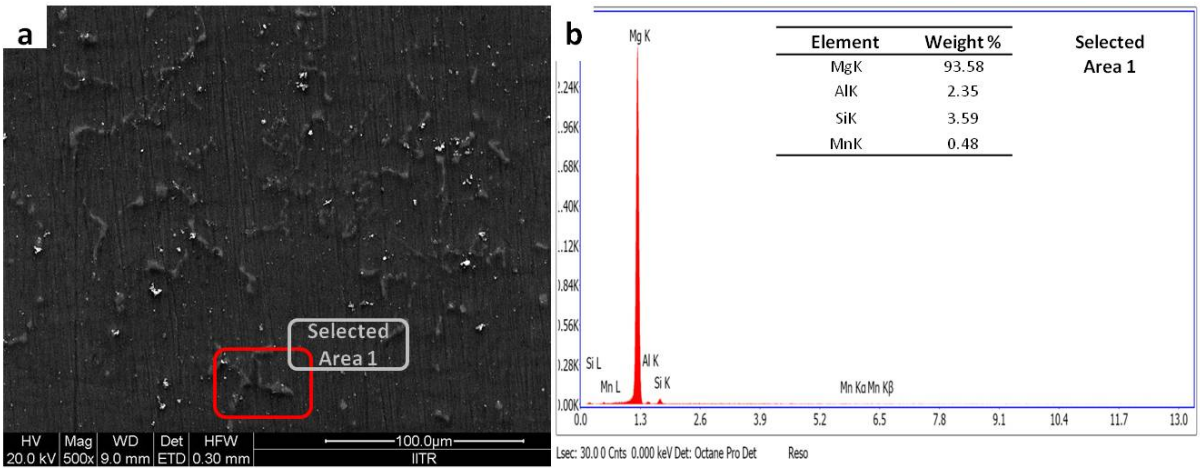


Figure 5. Microstructure characterisation of as-cast AS21A alloy: (a) SEM micrograph; (b) EDS spectrum.

3.2. Mechanical and wear characterisation

Table 2 represents the mechanical and wear characteristics obtained. The AS21A alloy exhibited enhanced ductility that can be attributed to the softening treatment. The tensile morphologies are

shown in figure 6 exhibited cracks, elongated dimples and cleavage facets on fractured surfaces which confirmed the blend of ductile-brittle nature.

Table 2. Mechanical and wear properties of the AS21A alloy

UTS (MPa)	YS (MPa)	Elongation (%)	Hardness (HV)	Load during wear test (N)	Wear rate ($\times 10^{-3}$ mg/m)	COF
117	74	15.5	55	10	11.64	0.32
				20	17.96	0.26
				30	26.40	0.22
				40	39.76	0.12

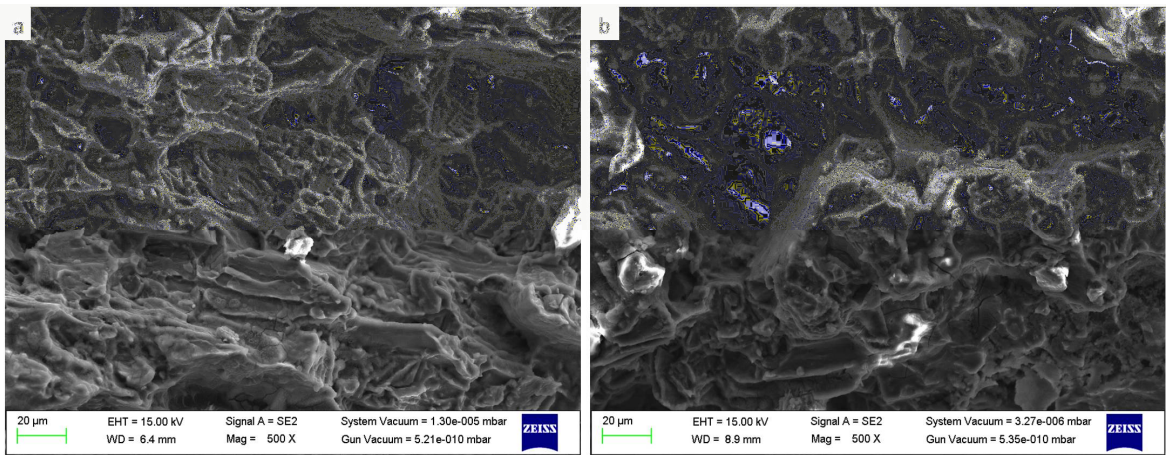


Figure 6. SEM micrographs of fractured tensile specimens.

Table 2 highlights the enhancement of wear rate with an increase in load. It was observed from the Coefficient of Friction (COF) Vs Sliding Distance (m) plot shown in figure 7 that an increase in normal load decreased the fluctuation in COF value. COF at a lower load of 10 N exhibited large fluctuations as compared to small fluctuations in case of a higher load of 40 N. The coarse Mg_2Si characteristics of AS21A alloy (shown in figure 4) was in a capacity to support the normal load acting during wear test thus decreasing the fluctuation at higher loads. A similar observation was evaluated by An et al. [15] during sliding wear behaviour of $Mg_{97}Zn_1Y_2$ alloy. The worn-out surfaces of wear samples were captured in the form of SEM micrographs, as shown in figure 8, to study the wear mechanism of the AS21A alloy. The major wear mechanisms observed during wear study were adhesion, oxidation, abrasion, and delamination. At 10 N loads, the presence of narrow grooves along with material sticking confirmed the combination of abrasion and adhesion wear mechanism. At 20 N loads, there were signs of grooves and oxide layer which confirmed the combination of abrasion and oxidation wear mechanism. At a higher load of 30 N and 40 N, the wear subsurface undergoes plastic deformation, and material removal takes place in the form of platelets. It confirmed the onset of delamination wear mechanism at higher loads. The wear mechanism observed was consistent with the literature [16, 17]. Selvan et al. [16] and Taltavull et al. [17] reported that the desired wear characteristics are a strong function of the applied load and sliding velocity.

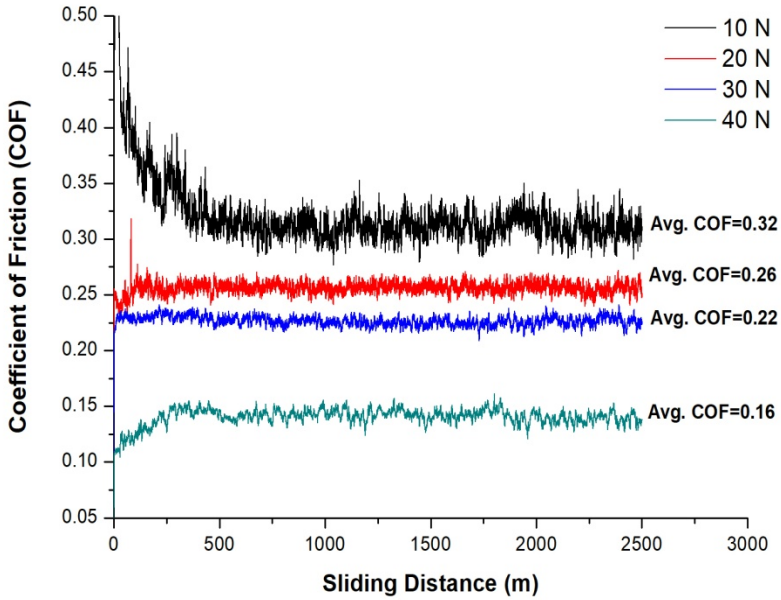


Figure 7. Graph showing COF variation with sliding distance

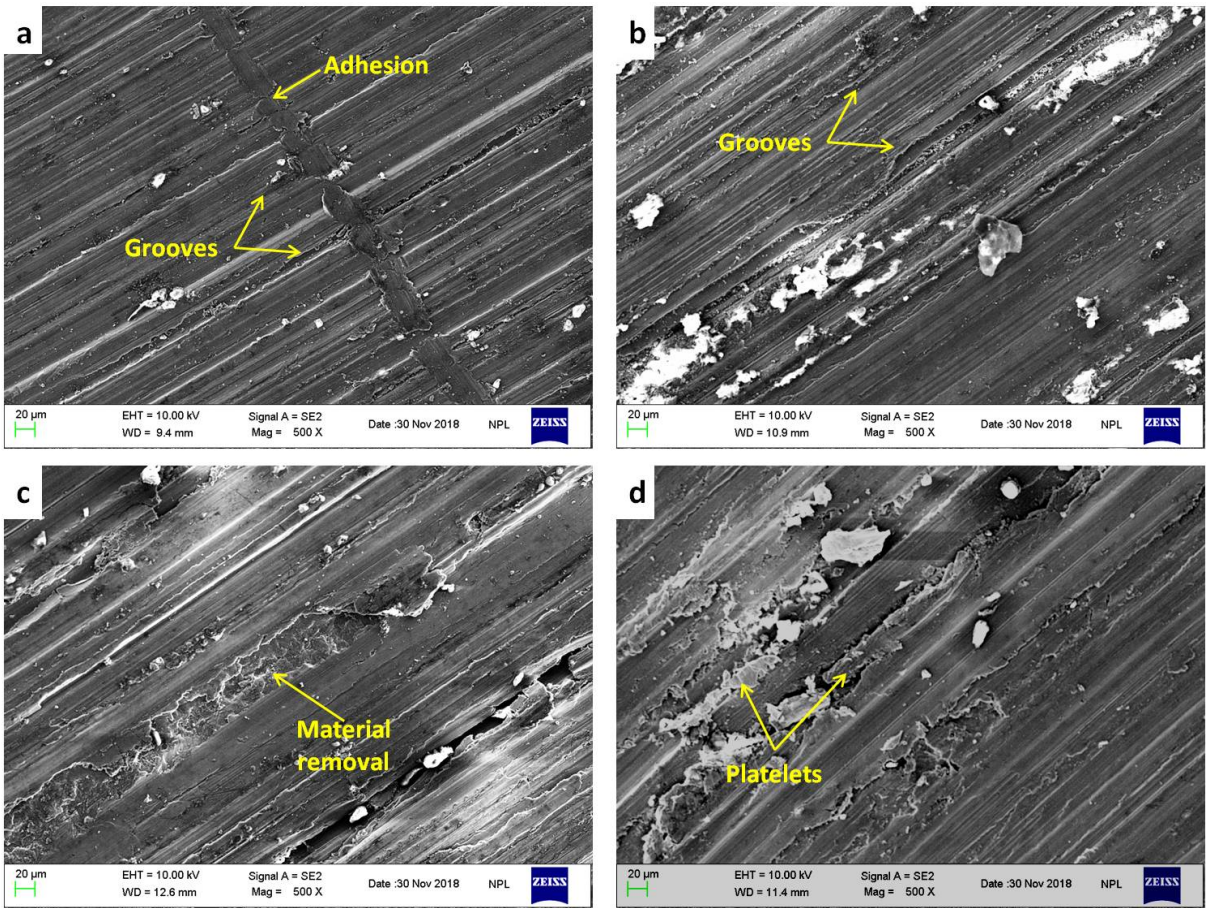


Figure 8. SEM micrographs of wear samples tested at an applied load of (a) 10 N; (b) 20 N; (c) 30 N and (d) 40 N.

4. Conclusions

The conclusions derived from the microstructural and wear study of as-cast annealed AS21A alloy are as follows:

1. The microstructural features of as-cast annealed AS21A alloy comprised of a primary α-Mg matrix reinforced with in-situ Mg_2Si intermetallic compound. The morphology of Mg_2Si found to be coarse and Chinese script.
2. The mechanical properties obtained were UTS of 117 MPa, YS of 74 MPa and % elongation of 15.5. Further, the high ductility of the studied alloy was confirmed by the fracture morphologies.
3. Wear resistance and coefficient of friction (COF) were found to be decreased with an increase in applied load. Furthermore, the Mg_2Si phase in the Mg matrix was found responsible for the decrease in COF fluctuations.
4. AS21A alloy exhibited a variety of wear mechanism at different loads: adhesion and abrasion at a low load of 10 N, abrasion and oxidation at an intermediate load of 20 N and finally delamination at a higher load of 40 N.

References

[1] Polmear I, St. John D, Nie JF and Qian Ma 2017 *Light alloys: metallurgy of light metals* 5th ed. (Oxford: Butterworth-Heinemann).
[2] Mordike BL and Ebert T 2001 *Mater Sci Eng A* **302** 37–45.
[3] Avedesian MM and Baker H 1999 *Magnesium and magnesium alloys: ASM specialty handbook* ISBN 0-87170-657-1.
[4] Chen H and Alpas AT 2006 *Wear* **246** 106–16.
[5] Taltavull C, Torres B, Rodrigo P Lopez AJ and Rams J 2013 *Wear* **301** 615–25.
[6] López AJ, Rodrigo P, Torres B and Rams J 2011 *Wear* **271** 2836–44.
[7] Luo A and Pekguleryuz MO 1994 *J Mater Sci* **29** 5259–71.
[8] Mabuchi M, Kubota K and Higashi K 1996 *Mater Sci Technol* **12** 35–9.
[9] Mabuchi M, Kubota K and Higashi K 1996 *J Mater Sci* **31** 1529–35.
[10] Akyuz B 2014 *Proc Inst Mech Eng B J Eng Manuf.* 1–9.
[11] AjithKumar KK, Pillai UTS, Pai BC and Chakraborty M 2013 *Wear* **303** 56–64.
[12] Li GH, Gill HS and Varin RA 1993 *Metall Mater Trans A* **24** 2383–91.
[13] Evangelista E, Gariboldi E, Lohne O and Spigarelli S 2004 *Mater Sci Eng A* **387–389** 41–5.
[14] Dargusch MS, Bowles AL, Pettersen K, Bakke P and Dunlop GL 2004 *Metall Mater Trans A* **35** 1905–09.
[15] An J, Li RG, Lu Y, Chen CM, Xu Y, Chen X and Wang LM 2008 *Wear* **265** 97–104.
[16] Anbu selvan S and Ramanathan S 2010 *Mater. Des.* **31** 1930–36.
[17] Taltavull C, Rodrigo P, Torres B, Lopez AJ and Rams J 2014 *Mater. Des.* **56** 549–56.

Acknowledgements

The authors would like to thank National Testing Laboratory, New Delhi for spectroscopy and Institute Instrumentation Centre, IIT Roorkee for the characterisation facility.

UNDER REVIEW SINCE 04TH JANUARY 2019

Microwave derived *monoclinic* $Ba_{1-x}Sn_xNb_2O_6$ materials as an alternative of ITO

Ashish Joshi[a], Vaibhav Shrivastava[b,†], and Anurag Pritam[b]

Microwave synthesis was optimized for preparing novel monoclinic Tin-doped Barium Niobate ceramics ($Ba_{1-x}Sn_xNb_2O_6$; x= 0.0, 0.01, 0.05, 0.1, 0.2, 0.3) BSN. The intensity of monoclinic phase formation was observed to decrease on increasing tin as dopant indicating decreased crystallinity. Strained crystalline phase was observed in undoped sample that became severe on doping tin. Monoclinic metal alloy Sn_2O_3 formation was confirmed on increasing tin doping beyond 5%. Electronic configuration of tin (II) oxide supports the local site-wise monoclinic disorder in crystal structure due to sterically active lone pair. Such a disorder arranges increased number of degenerate energy states and reduce effective energy gap between conduction band minimum and valence band maximum. All BSN compositions were investigated for monoclinic phase stabilization, ultra-violet absorption, dielectric response, raman modes and type of carrier concentration along with hall resistivity. All measurements possessed inflexion corresponding to 5 atomic % tin doping indicating successful site substitution and estimated Sn_2O_3 metal formation beyond this. The values of reducing optical energy band gap, transparency to visible spectrum and hall resistivity indicated utility of these materials as a substitute transparent conducting oxide (TCO) for well-known indium tin oxide (ITO).

Keywords: Microwave synthesis, TEM, FTIR, Raman modes, Optical energy band gap, Jonscher's law, Hall resistivity.

I. INTRODUCTION

Tungsten bronze (TB) structure based polar dielectrics form a very important class of non-toxic materials offering competent structural, electrical, optical, photocatalytic and electrocaloric response compared to corresponding toxic perovskite oxides like lead magnesium niobate (PMN), lead zirconium niobate (PZN), lead zirconium titanate (PZT), lanthanum-modified lead zirconate titanate (PLZT), lead scandium niobate (PSN) etc[1-3]. Many TB structures offer variety of applications emerging from NbO_6 octahedral distortions and phase transition between different polymorphs. Carruthers and Grasso *et al.*[4] discussed ternary room temperature phase relationship

†Author for correspondence: vaibhav.shrivastava@snu.edu.in

[a]Department of Applied Physics, Delhi Technological University, Delhi-110042, India

[b]Dielectric Laboratory, Department of Physics, Shiv Nadar University, G.B.Nagar-201314, India

between BaO-SrO-Nb_2O_5 oxides. The members of mixed matrix $Ba_{1-x}Sr_xNb_2O_6$, barium niobate BN ($BaNb_2O_6$) and Strontium Niobate ($SrNb_2O_6$) have not been discussed in detail for electro-optic coefficients, pyroelectric coefficients, and ferroelectric properties[5-7]. Current work was planned to utilize high transparency of barium-based dielectrics and instill high electric conductivity simultaneously to offer competitive alternative of Indium Tin Oxide, popular transparent conducting electrode material[8-10]. Indium tin oxide is an exhaustively used transparent conducting oxide (TCO) due to its high optical transparency (~85%) and electrical resistivity (~10^{-4} Ω-cm.) and optical energy band gaps (~3.8eV). The thin film geometries of ITO are highly used in solar cells, touch screens, flat panel displays etc. Further, ITO being an n-type degenerate semiconductor was also proposed as photocathode material for dye-sensitized solar cell structure[10].

Taking versatility of applications offered by ITO, our choice was to start with polymorph barium niobate BN ($BaNb_2O_6$) whose phase transition studies makes it further interesting optimize for selective phase for diverse applications. However, the studies on its structural, optical and dielectric properties are sparse and need to be reported. Mixed tungsten-bronze SBN system was reported mostly in orthorhombic form instead tetragonal[6, 7]. In these structures, cationic site A1 (tetragonal site) is proposed to be occupied by Sr^{2+} whereas A2-sites (pentagonal) can be occupied by both Sr^{2+} and Ba^{2+} in mixed matrix[5]. In BN, since there is no strontium thus A2-sites are expected to be occupied by barium ions. In the current work, microwave synthesis is optimized to prepare BN compound in monoclinic phase and further distort the unit cell by doping litharge structure type SnO for generating local disorder, hence, reduced crystallinity. The target is to introduce tin (Sn^{2+}) onto barium (Ba^{2+})-sites to prepare $Ba_{1-x}Sn_xNb_2O_6$ (BSN) composition series (x = 0.0, 0.01, 0.05, 0.1, 0.2, 0.3). This proposed substitution is expected to impart lateral covalent π-π overlapping to increase electron density between two ionic positions of Ba-O bond. High electron density was expected to be achieved by inducing degeneracy in undoped BN structure on

introducing tin. Thorough mechanical milling followed by homogenous microwave heating was expected to generate special *nano* dimension heat sinks on tin-sites in metal alloy form. Conducting metals possessing short skin depth re-radiate absorbed microwaves to the surrounding and generate local high temperature regions. On the other hand, atomic sites exhibiting prominent dielectric character, utilize absorbed microwave to resonate and generate controlled temperature regions by friction. The prepared compositions were further characterized for expected changes in crystalline phase formation, particle size variation, optical energy band gap, raman modes of vibration, *ac* conductance in correlation with rating of covalent character/ionic character in undoped BN and hall activity.

II. EXPERIMENTAL PROCEDURE

Each stoichiometric composition $Ba_{1-x}Sn_xNb_2O_6$ (BSN) was prepared by taking high purity (≥99.9%) $BaCO_3$, Nb_2O_5 and SnO powders (all from Sigma Aldrich Inc. USA). The mixed powders were grounded for three hours in covered electrical agate pulverizer in intermediate stages of spatula mixing. The mixtures were calcined at 1050^{o}C for 3 hours using a microwave furnace operating at 2.45 GHz with deliverable power of 2.2 kW in automatic magnetron changing mode. The SiC susceptor cavities were used for concentrating microwave power in a region of 60X60X60mm. The heating (3^{o}C per minute) and cooling (2^{o}C per minute) profiles were monitored through Eurotherm 2416 controlled AST infrared pyro sensor. The calcined powders were uniaxially pressed into disc-shaped pellets to minimize density gradient by applying a pressure of 700MPa. The densification of pellets was carried via microwave sintering at 1050^{o}C for 1 hour for removal of PVA added as an organic binder. X-ray diffractograms were recorded using Bruker D8 Advance model in coupled (2θ-θ) mode using Cu$K\alpha$ radiation of wavelength 1.54406Å in the range of angle 2θ from 10^o to 80^o. Transmission Electron Microscopy (TEM) was engaged to

analyze the grain growth and to compare particle size with crystallite size as estimated using Scherrer's expression. These images were captured using LIBRA 200FE high resolution TEM with information limit of 0.13 nm equipped with EDS, which operated at 200kV. Optical energy band gap for each modified BSN composition was calculated using *Tauc* plots. These plots were derived using UV-Vis diffuse absorption spectra measured in an integral sphere mode using spectrophotometer SHIMADZU UV-Vis 3700. For reflecting incident radiations $BaSO_4$ was used as a medium co-mixed with each modified BSN composition. The vibrational characteristics of all present chemical bonds in BSN compositions were studied using Fourier transform infrared (FTIR) spectroscopy. These FTIR spectra were recorded on FTIR Spectrophotometer of Thermo Fisher Scientific make and model Nicolet iS5 in the range of 500 and 4500 cm^{-1}. For recording this spectrum, each BSN composition was mixed with *KBr* powder and infrared light was shone on the prepared pellet. The pellets of prepared BSN composition were coated using silver and gold for the dielectric measurements. Dielectric data was collected using high frequency LCR meter NF2376 from NF Corporation, Japan in the frequency range 20 to 2MHz at temperatures from 25 to 575 ^oC.

III. RESULTS AND DISCUSSION

1. Crystal Structure and Morphology Analysis

Perfect monoclinic phase formation is confirmed in all microwave prepared BSN samples using JCPDS 01-077-541, Fig.1. All the x-ray peaks are well indexed for monoclinic phase of polymorph $BaNb_2O_6$, also this phase remains preserved overall on tin doping. Most reports have a mention about orthorhombic phase of strontium doped $BaNb_2O_6$ generally admixed with tetragonal phase[9,10] yielding crystalline planes diffused within each other along oblique axes of orthorhombic and tetragonal symmetries. A few unindexed new peaks (110)/(101) prominently emerge between 25-26.9^o and 33.1-34.3^o on increasing tin doping after 5%. These peaks belong to tin in oxide

forms, Sn_2O_3 (JCPDS 00-025-1259), SnO (JCPDS 00-006-0395) and SnO_2 (JCPDS 00-088-0287) in descending order of their peak matching percentage respectively. These fix Bragg angle peaks support metal alloy formation embedded in dielectric BSN unit cell structure in preferred orientation. Minor increase in x-ray intensities of these peaks is crucial in concluding the pivotal role of microwave heating followed after thorough milling of mixtures. The metal clusters of tin are expected to receive maximum microwave power in atomic network and melt *before* barium carbonate to flow randomly. This random flow in eutectic form presumably fills the vacant sites of barium to saturate coordination number. Current monoclinic structure is known to be natural tungsten bronze type and typically useful in photovoltaic applications[11]. Nearly undetectable intensity change in x-ray diffractograms for samples up to x=0.05 is an indicative of tin successfully substituting barium on respective sites.

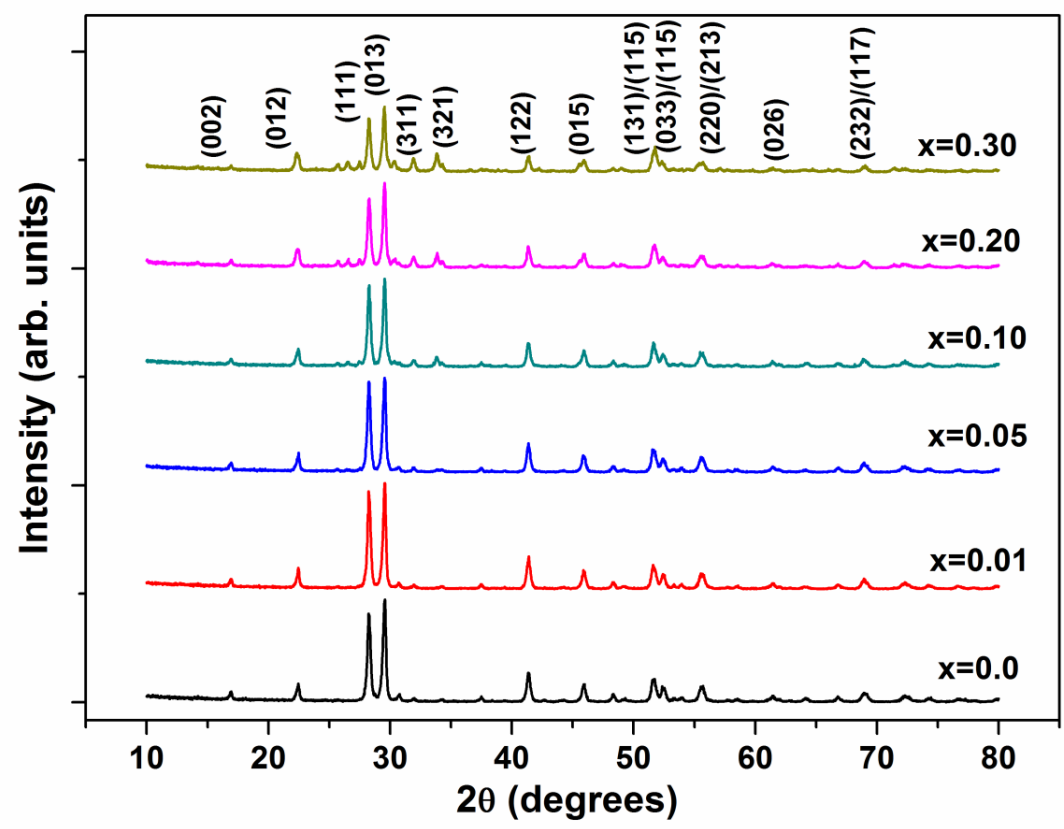


Fig. 1. X-Ray diffractograms of $Ba_{1-x}Sn_xNb_2O_6$ composition pellets

The introduction of tin beyond 5 % causes monoclinic metal alloy Sn_2O_3 formation in $BaNb_2O_6$ in addition to dominant monoclinic phase. The generated distortions are further investigated through calculations of distortion parameters like tetragonal strain and orthorhombic distortion. The lattice parameters are calculated using indexed peaks (012), (111) and (013) and used for computing distortion parameters. The values of lattice parameters (*a:*3.957Å, *b:*6.063Å, *c:*10.453Å) calculated are in accordance with those recommended for monoclinic BN. The variation in the values of lattice parameters *a*, *b* and *c* exhibits inflexion around 5% tin doping indicating exact site substitution without alloy formation till 5% doping. Tin (Sn: $Kr,4d^{10},5s^2,5p^2$) as dopant for barium (Ba: $Xe,6s^2$) promotes the formation of covalent bonds site wise compared to ionic interaction between barium and oxygen. Many reports have a mention about measure of increased covalence in the crystal structure in terms of tetragonality (c/a), the same is exercised in current work[12, 13]. An increase in tetragonal strain values (c/a) for initial doping up to 5% indicates about expected increase in covalence nature of chemical bonds due to c-axis oriented lateral π-π overlap. This finally results into lattice relaxation in *a*-*b* plane thereby decreasing orthorhombic distortion. Crystallite size (*t*) was estimated using the Scherrer's formula for the high intensity peaks including the maximum one: $t = \frac{k\lambda}{\beta cos\theta}$, where, λ is the wavelength of x-ray used, θ is the position of the maximum intensity peak, β is full width at half maxima (FWHM) of maximum intensity peak and k=0.9 is the Scherrer's constant. The crystallite size ranges from 27.1 nm to 29.5 nm and follows an inverse relationship with the tetragonal strain leaving sample x=0.01.

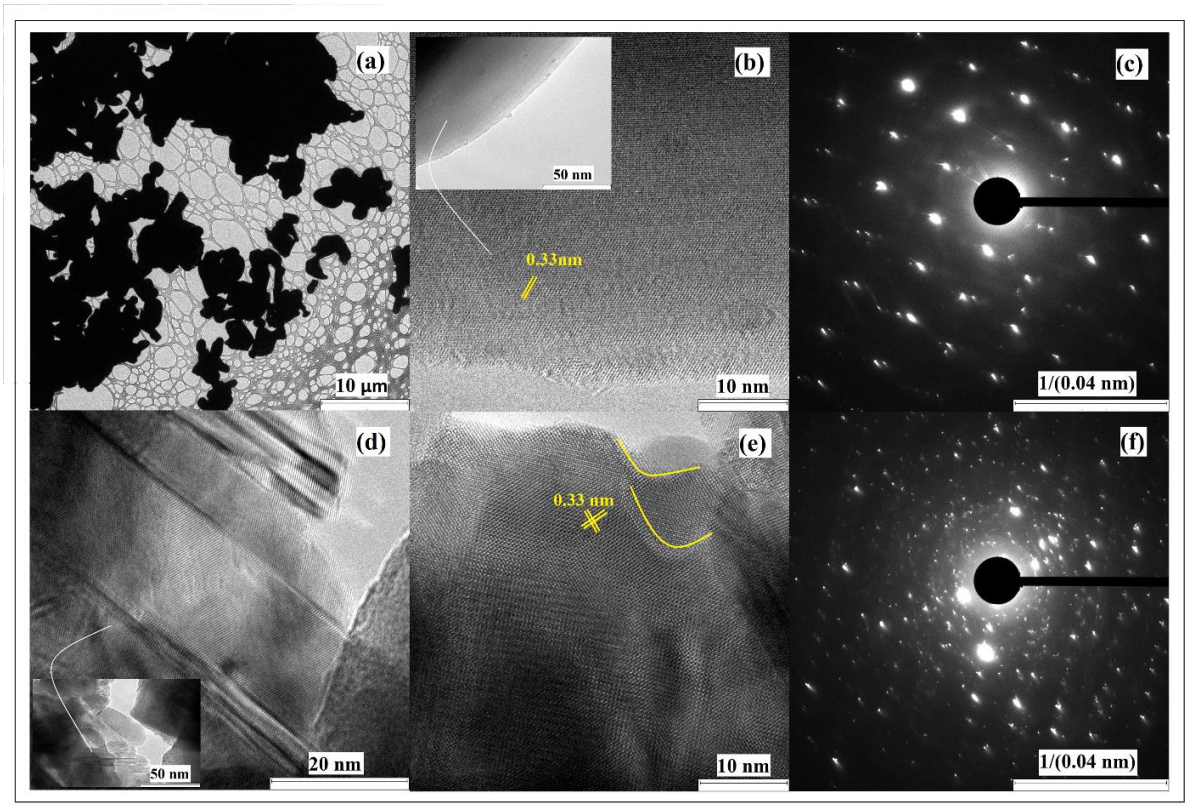


Fig. 2. TEM images of $Ba_{1-x}Sn_xNb_2O_6$ grains [(a)-(c) for x=0.00 and (d)-(f) for x=0.30]

In current work, premeditated interaction of microwaves is executed with tin metal alloys regions in doped BSN compositions. Beyond 5% doping, tin is observed to follow alloy formation, which is behaving finally as metal in dielectric network to attract maximum fed microwave power[14]. Two typical samples (undoped BSN and 30% doped BSN) were taken from the crushed pellet powder to investigate the microstructure by transmission electron microscopy (TEM), showing the formation of nanocrystals. The microwave prepared BSN compositions contained bigger particles of the order of 1μm along with a few regions of *nano* particles [see Fig.2(a)]. The micron-sized crystals are showing long range order as the one shown in Fig.2(b), confirming low lattice strain. The diffraction spots observed in selected area electron diffraction (SAED) regime [Fig.2(c)] indicate the formation of strained crystalline network in undoped $BaNb_2O_6$ sample. This strain

becomes severe in maximum doped BSN sample; showing clear interweaved zones indicated by arrows [see Fig.2(d)]. Further inspection of Figs.2(e) & (f) indicate the development of polycrystalline phase, while the grain boundaries are clearly visible in Fig.2(e)[15]. The local tin alloy regimes melted before barium carbonate and flow in eutectic form to channelize the quick preparation of monoclinic $BaNb_2O_6$ structure [see curved surfaces of grains in Fig.2(e)]. The distribution of energy dispersive spectrometry (EDS) peaks confirm presence of very close atomic stoichiometry as planned with no other elements present. EDS results indicate an increase in mass density of $Ba_{0.7}Sn_{0.3}Nb_2O_6$ matrix from 5.71 (for $BaNb_2O_6$) to 6.07 g/cm^3 on increasing tin by 30 atomic % (x=0.30). Tin possessing less atomic weight than barium should have caused decrease in mass density. This provides an indication of tin occupying sites other barium on high doping. The other available sites of 6-coordinated niobium favor tin occupancy; this was investigated further using spectroscopic methods.

2. Spectroscopic investigations

Fig.3 shows the optical absorption spectrum for investigated materials. The BSN compositions show major absorption in *uv* region starting from 250*nm* and concludes at 410*nm.* Such a broad absorption region is solely due to intrinsic transition of semiconductors and less likely due to defect energy states. The absorption peak broadens on increasing tin up to 10% and decreases thereafter. The broadening is investigated using full width at half maximum (FWHM), as an ancillary indication of grown semiconducting behavior of dielectric BSN compositions, Table-1. The absorption edge in all BSN compositions is found in the UV-Vis region from 380- 415*nm* indicating about energy gap between valence band maximum and conduction band minimum typically around 2.98-3.26eV, better than ITO[8,16,17]. The samples up to 5% doping show absorption edges shifting towards blue wavelength whereas the ones with higher doping exhibited red shift. The absorption values were below 0.1 for all samples, which is an essential requirement for TCO.

The samples with tin doping up to 5% possessed absorption lower than undoped BN that increased on tin doping beyond 5 atomic% justifying optimal tin substitution. Almost no change in the wavelength value for maximum absorption intensity supported typical strength of monoclinic phase of BN towards photo absorption after doping. The variation of absorption intensity is directly

Table-1 Optical energy band gap values and FWHM for UV absorption in BSN samples

Composition (x)	FWHM (*nm*)	E_g (*eV*)
0.00	112.433	3.51
0.01	113.133	3.45
0.05	141.308	3.47
0.10	182.838	3.12
0.20	170.207	3.01
0.30	150.655	3.22

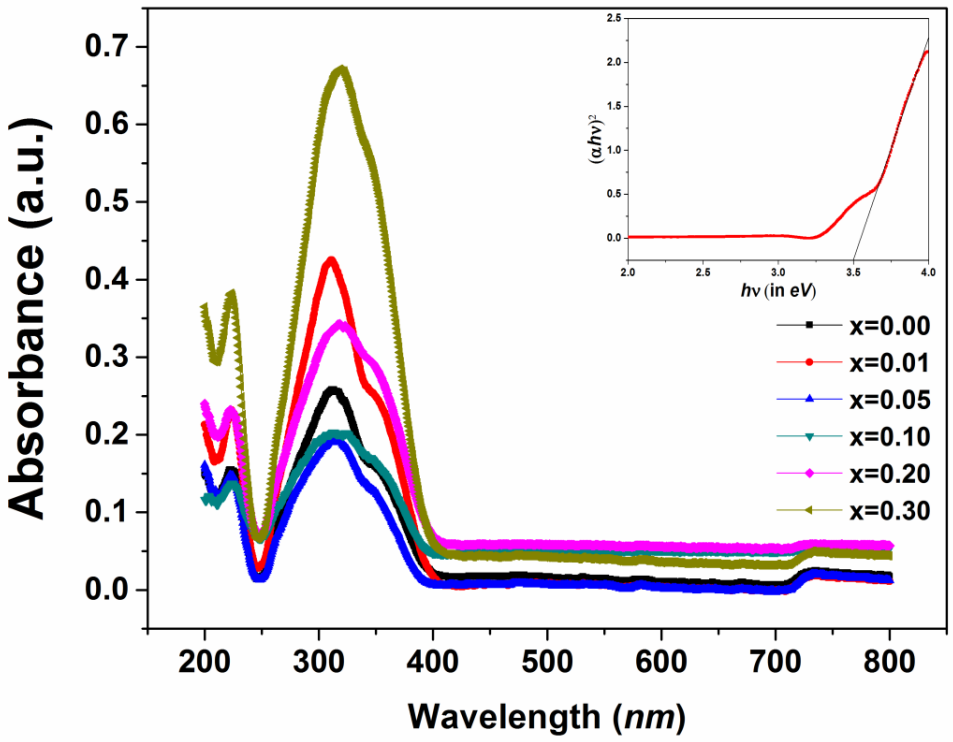


Fig. 3. UV-Visible absorption spectra of $Ba_{1-x}Sn_xNb_2O_6$ compositions

related to the crystallite size of the BSN samples. Samples with big crystallites absorb more UV radiation than those with small crystallites. Therefore, the response of absorption intensities possessed by samples x=0.05 and x=0.30 can be understood self. All BSN samples show a direct band gap transition and the calculated band gap energy values are listed in Table-1. A decrease in optical energy band gap on tin doping is due to significant increase in number of energy states compared to undoped BN. Additional transmittance measurements for a few samples; confirm the transparent response of BSN compositions to visible spectrum starting from 415*nm* because of transparent nature of constituent reduced oxides BaO, SnO and Nb_2O_5 [18-20].

Tin induced structural distortions in BSN compositions are further investigated using micro-raman spectroscopy, Fig.4. It is an effective method for rating the structural perfectness of an atomic framework. The distorted framework possess additional low energy coupled vibrations usually detectable as scattering losses. Therefore, raman scattering is the best tool for detecting these coupled vibrations emerging due to inelastic scattering of laser photons after interaction with electrons in atoms at different atomic sites. In addition, active raman modes in any structure correspond to number of atoms per unit cell in that structure. The obtained raman spectra of BSN samples are in excellent agreement with the work of Repelin *et al* [21], who concluded about tilting of corner shared NbO_6 octahedrons generating typical orthorhombic unit cell structure that can further distort to generate monoclinic system. Fifteen active raman modes are present in undoped BN, confirming the 15-folded A2 site occupancy by barium in BN, Table-2. There is no shift in all peak positions except tin-alloy induced growth and splitting of peaks 2 and 13. Thus, parent monoclinic phase of BN remains unaltered except diffusion of additional monoclinic phase of Sn_2O_3 at 10% doping and above[22]. Tin oxide (SnO) occupies less polar and more *c*-axis strained *tetragonal* litharge structure[23]. Hence, tin doping infuses a decrease in *phonon driven intrinsic*

polarizability per unit cell, therefore, all doped BSN compositions show decrease in intensity of all raman modes. Again, an increase in tetragonality (*c/a*) and no splitting of raman peaks can be correlated for tin doping up to 5% as estimated in previous sections. Beyond 5% doping, tin atoms not only occupy barium sites corresponding to unaltered raman peak positions rather generate local asymmetry to produce coupled vibrational modes due to monoclinic Sn_2O_3 formation as local distortion. These coupled modes appear as splitting of band-13 in present work. This is also confirmed by investigating full width at half maxima (FWHM) for vibration bands 1, 3, 5, 11 and 13. As tin is introduced in the BSN system (sample x=0.01), FWHM increases hence, raman bands become wider to adjust peak intensity to lower value. This further introduces a few more new peaks and shoulder bands starting from BSN composition x=0.1 in low frequency as well as in

Table 2. Raman bands in $Ba_{1-x}Sn_xNb_2O_6$ compositions

Band	Frequency (cm^{-1})	Band	Frequency (cm^{-1})
1	120 vs	9	376 w
2	146 w (for BN) + s (for Sn_2O_3)	10	490 w
3	181 s	11	556 vs
4	199 w	12	627 w
5	227 s	13	707 vs (for BN) + s (for Sn_2O_3)
6	278 m	14	771 w
7	303 m	15	841 w
8	360 w		

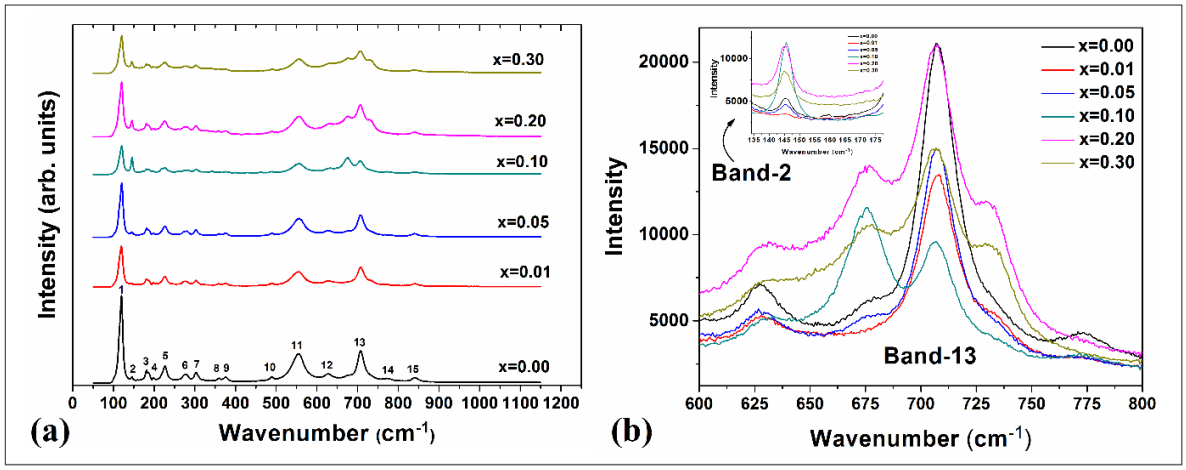


Fig. 4. Raman Spectra of BSN compositions

high frequency region. In x=0.2 and x=0.3 samples, the low frequency shoulder peaks are diminished while the shoulder peaks at high frequency bands become more pronounced.

To study the intrinsic contributions to real permittivity of BSN compositions, which is normally observable at high frequencies (~10^{12} Hz and beyond), FTIR spectroscopy is employed. Various chemical bonds present in a typical stochiometrically prepared composition vibrate after absorbing infrared radiation corresponding to typical expression given by,

$$\bar{\nu} = \frac{1}{2\pi c}\sqrt{\frac{k}{\mu}}$$

where $\bar{\nu}$ is the wavenumber (cm^{-1}), c is the velocity of light (cm./sec.), k-is the force constant (dynes/cm.), and μ is the reduced mass (gms.). The results of FTIR spectra of the BSN samples are in exact correlation with raman modes appeared from 500 to 1000 cm^{-1}, Fig.5. Observed IR absorption bands are compared with the data already reported[24-26] for Ba-O, Sn-O and Nb-O bonds. The absorption bands between 500 cm^{-1} and 600 cm^{-1} are assigned to antisymmetric stretching

vibrations of Nb-O bonds in NbO_6 octahedra. The absorption bands at 600 cm^{-1} and 630 cm^{-1} are corresponding to Ba-O and Sn-O bond antisymmetric stretching vibrations respectively. It can be

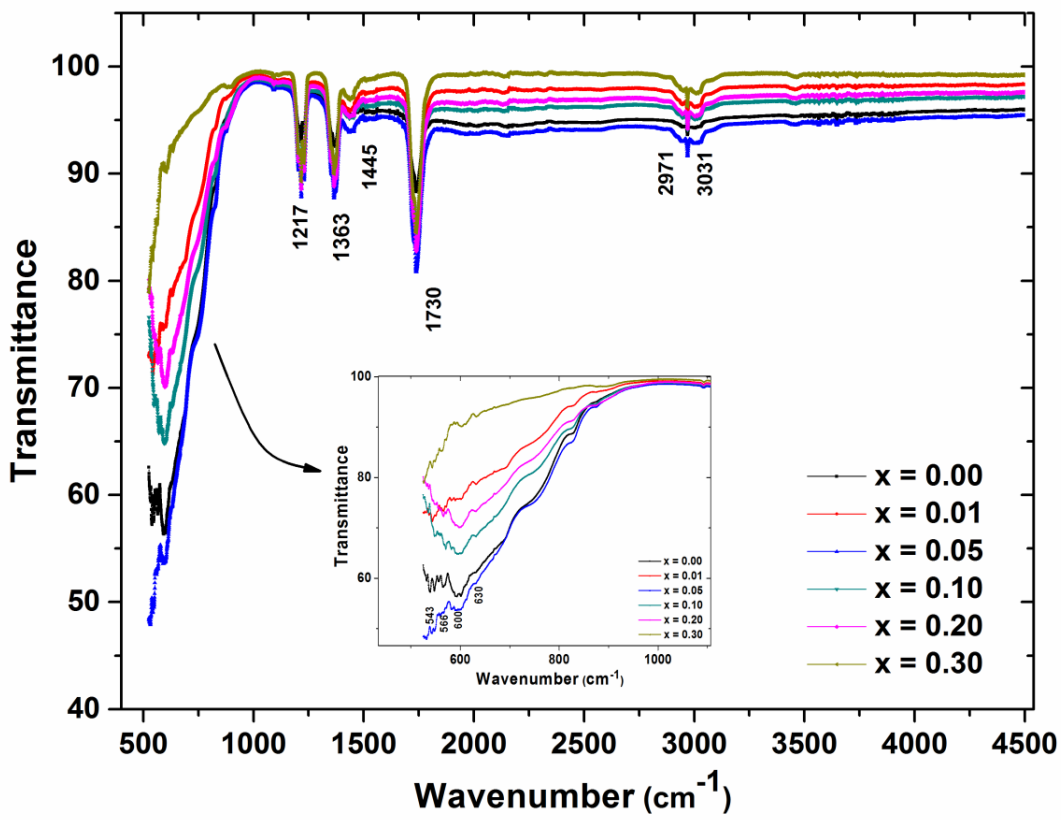


Fig. 5. FTIR Spectra of $Ba_{1-x}Sn_xNb_2O_6$ compositions

seen that on increase tin doping the peak intensity corresponding to Ba-O bond decreases and the same for Sn-O bond increases. This intensity change confirms the successful replacement of barium ions by tin ions onto 'A2' sites. The metal-oxygen bonds are observed to resonate at energies corresponding to wavenumbers lower than 1000 cm^{-1} range. While the vibration bands observed beyond 1000cm^{-1} are corresponding to the response of carbon assisted functional groups to absorbed radiation. The absorption band at 1217 cm^{-1} is attributed to C-O stretching vibrations, which could be either be due to residual carbonates in composition or because of adsorption of

CO/CO_2 from the atmosphere. The presence of the absorption bands at 1363 cm^{-1} and 1445 cm^{-1} verify the presence of $BaCO_3$ minor traces mostly present on open powder surfaces due to CO/CO_2 adsorption during FTIR measurement. All expected absorption bands are observed in the investigated BSN samples. Absorption bands at 1730cm^{-1} (bending), 2971 cm^{-1} (stretching) and 3031 cm^{-1} (stretching) are attributed to the O-H bond vibrations. The source of these O-H vibrations could be the moisture absorbed from atmosphere by the samples.

3. AC Conductivity

The degree of covalence determining augmented dielectric nature of BSN compositions is verified using *ac* conductivity (σ_{ac}) investigations as proposed by Jonscher's universal power law[27]: $\sigma_{ac} = \sigma_{dc} + A\omega^n$, where, σ_{dc} is the frequency independent *dc* conductivity contribution, A is a temperature dependent constant, ω is the angular frequency ($\omega = 2\pi f$) and *n* is the frequency exponent. The value of frequency exponent *n* determines the dominant conduction mechanism: $0<n<1$ associated with the hopping mechanism of ions and $n>1$ associated with the quantum mechanical tunneling of charge carriers over localized sites[28]. The conductivity values are calculated from the dielectric data using the relation $\sigma_{ac} = 2\pi f\varepsilon'\varepsilon_o\tan\delta$, where the symbols have their usual meaning. The variation of *ac* conductivity with frequency and temperature is investigated to conclude the nature of electronic transfer in the BSN samples, Fig.6. Note that *ac* conductivity values increase with frequency and tin concentration in BSN compositions. Typical values are 5.44µS/m (at 20^{o}C and 50Hz), 67.6µS/m (at 20^{o}C and 1 MHz) and 1028.5µS/m (at 300^{o}C and 1 MHz). The values of *n* for all samples is computed using the slope on linear part of data curves in Fig.6, Table-3. The value of *n* increase with the increase in tin concentration as estimated earlier between the limits of hopping mechanism[28] confirming additionally *an increase in covalence* on doping tin. The increase in conductivity values with frequency could be due to lowering of the capacitive energy barrier around domain boundary. The *ac* activation is also probed by utilizing temperature

dependence of dielectric data as shown in the inset of Fig.6 ($\ln\sigma_{ac}$ versus 1000/T). A typical insulating negative temperature coefficient of resistance (NTCR) behavior is observed for all BSN compositions. Near the phase transition temperatures, the rise in σ_{ac} values of all the BSN samples

Table-3 Covalence parameter and *ac* activation energy for $Ba_{1-x}Sn_xNb_2O_6$ compositions

x	*n*	E_a (eV)
0.00	0.24	0.258
0.01	0.26	0.261
0.05	0.30	0.265
0.10	0.37	0.375
0.20	0.42	0.239
0.30	0.49	0.198

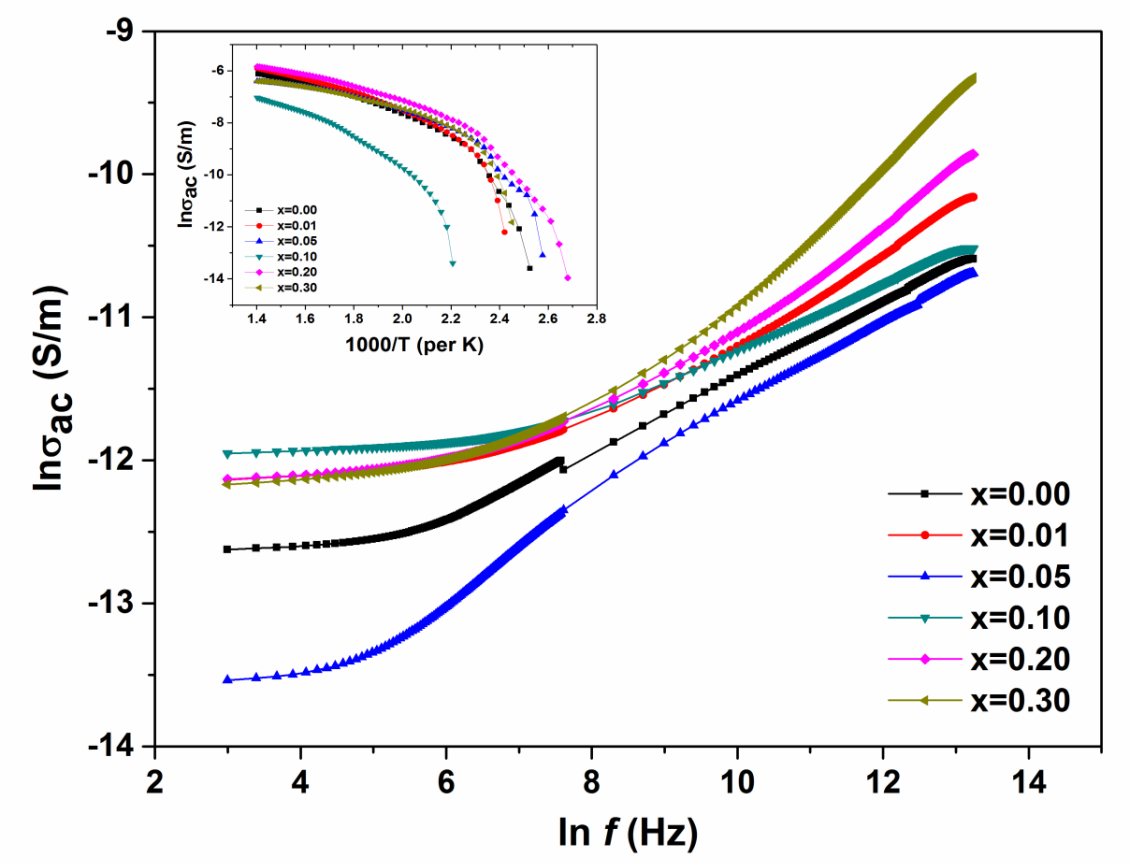


Fig. 6. Variation of σ_{ac} with frequency and temperature (inset)

is steep and is attributed to crystalline phase change from monoclinic to orthorhombic. The *ac* activation energy (E_a) is calculated using the slope of curves as plotted on Inσ_{ac} vs. 1000/T graph[29] using the equation: $\sigma_{ac} = \sigma_0 \exp(-E_a/k_BT)$, where, σ_0 is a constant, k_B is Boltzmann's constant and T is absolute temperature. The activation energies of all samples are listed in Table-3.

4. Hall Measurements

The measurements of absorption spectroscopy and microstruture study using TEM-SAED/HRTEM images in conjugation with dielectric investigations have indicated towards the following: a) tin doping beyond 5% generates *nano* regions of amorphous network as a result of metal alloy Sn_2O_3 introducing severe disorder in crystal structure of BN, b) overall low values of

dielectric constant are useful in facilitating exciton seperation and c) transmittance to major visible spectrum. These indications prompted us to investigate BSN compositons for type of carrier concentration and hall resistivity using van der Pauw geometry[30]. The measurments are recorded on randomly selected BSN composition thin pellets (~0.5mm) at fixed magnetic field of 500 Gauss using four conducting electrodes on arbitrary thin pellet surface. The Hall voltage values were recorded across two diagonal contacts by decreasing current across other two diagonal contacts from 10 *m*A to 0.5*m*A. The data of Hall voltage (Fig.S5) was normalized according to expression, $V_H = (V_{12} + V_{21} + V_{34} + V_{43})/2$. The nature of carrier concentration in undoped BN was found to be *n*-type that was gradually converted into *p*-type on increasing tin doping up to 5%. An increase in tin doping beyond 10% follows a reverse trend. The sample with 10% tin doping possess *n*-type conduction presumably due to emerging metal clusters and contributing electrons as majority carriers. All BSN compositions were found to exhibit carrier concentration higher than 10^{20} per cm^3 with minimum Hall resistivity value ~ 3.4×10^{-3} Ω-cm for 30% tin-doped BSN composition.

IV. Conclusions

Single *monoclinic* phase barium tin niobate $Ba_{1-x}Sn_xNb_2O_6$ BSN (x=0.00, 0.01, 0.05, 0.10, 0.20 and 0.30) ceramics have been successfully prepared using microwave synthesis. Tin occupied 15-coordinated A2-barium sites in all BSN compositions in addition to forming minor *locally distortive* monoclinic phase Sn_2O_3 metal alloy on high doping (≥10%). Tin doping introduced severe disorder in crystalline network promoting semiconducting nature of less strained dielectric BN compound. The reason for this disorder is the asymmetric electron distribution around tin sites due to well-known Sn $5s^2$ sterically active lone pair. The asymmetric electron distribution is due

to coupling between unfilled Sn (5p) with antibonding combination generated from interaction between Sn (5s) and O (2p) orbitals. These degenerate energy states affect intrinsic polarizability of atomic sites that is investigated through *raman* spectroscopy. The results of *uv-visible* absorption and *infrared* spectroscopy indicated towards significant changes in absorption edges and vibrational character of expected and new bonds. The values of absorbance, optical energy band gap, low dielectric constant and hall resistivity are in very close agreement with those reported for popular indium tin oxide (ITO). The investigated microwave synthesized BSN compositions offer uniqueness on many grounds like non-toxic, low cost, earth crust abundance and energy efficient method of preparation. This work proposed BSN compositions as a competent substitute for ITO.

References

1. Zhu X, Fu M, Stennett M C, Vilarinho P M, Levin I, Randall C A, Gardner J, Morrison F D, Reaney I M. *Chem. Mater*., *27* (9): 3250–3261 (2015)

2. Buixaderas E, Kadlec C, Kempa M, Bovtun V, Savinov M, Bednyakov P, Hlinka J, Dec J. Scientific Reports, 7:18034 (2017)

3. Lee K M, Jang H M., *Journal of the American Ceramic Society*, 81(10): 2586-2596 (1998)

4. Carruthers J R, Grasso M., *J. Electrochem. Soc.: Solid State Science*, Vol.117, No.11:1426-1430 (1970)

5. Kim D W, Hong H B, Hong K S, Kim C K, Kim D J., *Japanese Journal of Applied Physics*, 41(10R), 6045 (2002)

6. Kim D W, Hong K S, Kim C H and Char K, *Applied Physics A*, 79(3), 677-680 (2004)

7. Gaikwad S P, Samuel V, Pasricha R, Ravi V., *Bulletin of Materials Science*, 28(2): 121-123 (2005)

8. Xia N, Gerhardt R A., *Materials Research Express*, 3:116408 (2016)

9. Cong Y, Han D, Dong J, Zhang S, Zhang X, Yi Wang Yi, *Scientific Reports*, 7: 1497 (2017)

10. Yu Ze, Perera I R, Daeneke T, Makuta S, Tachibana Y, Jasieniak J J, Mishra A, Bauerle P, Spiccia L, Bach U., *NPG Asia Materials*, 8, e305 (2016)

11. Tomaszewski P E., *Phase Transitions*, 38, Iss.3: 127 (1992)

12. Cohen R E., *Nature*, 358: 13 (1992)

13. Fu D, Itoh M, Koshihara S Y., *J.Phys.:Cond.. Matter*, 22: 052204 (2010)

14. Singh S, Gupta D, Jain V, Sharma A K., *Materials and Manufacturing Processes*, 30: 1–29 (2015)

15. Jamieson P B, Abrahams S C, Bernstein J L., *J. Chem. Phys.*, 48, 5048 (1968)

16. Chen D, Ye J., *Chem. Mater.*, 21: 2327 (2009)

17. Cho I S, Bae S T, Kim D H, Hong K S., *International Journal of Hydrogen Energy*, Vol.35, Iss.23: 12954-12960 (2010)

18. Madelung O, Rössler U, Schulz M., *Semimagnetic Compounds. Landolt-Börnstein - Group III Condensed Matter* (Numerical Data and Functional Relationships in Science and Technology) Springer, Berlin, Heidelberg, Vol. 17B-22A- 41B: 1-7 (1999)

19. Vogel E M., *J. Am. Ceram. Soc.*, 72: 719 (1989)

20. Reddaway S F, Wright D A., *British Journal of Applied Physics*, Vol.16, No.2: 195-198 (1965)

21. Repelin Y, Brussat E H H., *Spectrochimica Acta*, Vol.35A: 937 (1979)

22. Maki-Jaskari M A, Rantala T T., *Modelling Simul. Mater. Sci. Eng.*, Vol.12: 33–41 (2004)

23. Walsh A, Watson G W., *Physical Review B,* Vol.70: 235114 (2004)

24. Mevada K C, Patel V D, Patel K R., *Arch Phy Res, 3*: 258-263 (2012)

25. Xun S, Zhang Z, Wang T, Jiang D, Li H., *Journal of Alloys and Compounds, 685*: 647-655 (2016)

26. Augsburger M S, Pedregosa J C, Sosa G M., *Journal of the Mexican Chemical Society*, Vol. 44, No. 2: 151-154 (2000)

27. Jonscher A K., *Journal of Physics D: Applied Physics*, *32*(14): R57 (1999)

28. El-Mallah H M., *Acta Physica Polonica-Series A General Physics*, *122*(1): 174 (2012)

29. Sharma S, Basu T, Shahee A, Singh K, Lalla N P, Sampathkumaran, E V., *Journal of Alloys and Compounds*, *663*: 289-294 (2016)

30. Ngo N, Niu H, Bharadwaj P, Bhatti H, Adhikari S., *Technical Report*, DOI: 10.13140/RG.2.2.23468.67208 (2017)

Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-019-00852-3

ORIGINAL ARTICLE

Modelling and simulation of repairable mechanical systems reliability and availability

Girish Kumar[1] · **Vipul Jain**[2] · **Umang Soni**[3]

Received: 2 May 2019 / Revised: 29 July 2019
© The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2019

Abstract Markov approach is applicable for reliability and availability modelling when time to failure and repair follow an exponential distribution. Since failure time of mechanical components follows Weibull distribution, Markov approach cannot be employed to these systems. In present work, Semi-Markov model, which is appropriate for repairable mechanical systems, is considered. In structural dependency, once a unit of a repairable system is failed due to one or more of its constituent components, the entire unit is taken for repair. Therefore, the repairs are considered at the unit level. This feature of structural dependency in the proposed approach addresses the problem of larger state space. The states at the unit level are derived from the component states to develop the system model. The solutions for reliability and availability are obtained using the Monte Carlo simulation. The suggested approach is realistic than the existing software's such as Rapid Algorithmic Prototyping Tool for Ordered Reasoning, Blocksim, etc., as failures and repairs are considered at different hierarchical levels. The recommended approach is illustrated for a centrifugal pumping system.

Keywords Weibull · Reliability · Semi Markov · Simulation · Mechanical systems

1 Introduction

Higher system availability and reliability are desirable in many industries where downtime adversely affects the production cost. In the past, the reliability and the availability models with various features such as imperfect repairs, common cause failure, human error, etc., are mostly analysed using the Markov approach (Hajeeh 2012). However, the Markov process can be used if the time to failure and repair follows an exponential distribution. System reliability is overestimated when structural dependency is not considered (Dao and Zuo 2017). Therefore, such dependency features should be incorporated in modelling for a realistic assessment of reliability. In general, for mechanical components, failure rates are not constant. Hence, it is appropriate to use a general probability distribution that better represents the stay time behaviour in a state-space model (Kumar et al. 2013a). The model so obtained follows a Semi-Markov process (Semi-MP) and its analytical solution is difficult (Welte 2009). The failure process of mechanical systems is better modelled with Weibull distribution (Fricks and Trivedi 1997). The Semi-MP that allows non-exponential distributions offers a more appropriate model (Perman et al. 1997). There is no closed form transient solution method for the Semi-MP model. However, some researchers attempted the

✉ Girish Kumar
girishkumar@dce.ac.in

Vipul Jain
vipul.jain@vuw.ac.nz

Umang Soni
umang.soni@nsut.ac.in

[1] Department of Mechanical Engineering, Delhi Technological University, Bawana Road, New Delhi 110042, India

[2] School of Management, Victoria Business School, Victoria University of Wellington, Wellington 6140, New Zealand

[3] Department of Manufacturing Process and Automation Engineering, Netaji Subhas University of Technology, Sector 3, Dwarka, New Delhi 110078, India

Published online: 14 August 2019

solution by approximations of non-exponential distributions to exponential. Welte (2009) derived the solution of the Semi-MP with the gamma distribution by approximating to a Markov process. Weibull distribution was approximated to exponential to obtain the Semi-MP model solution (Xie et al. 2004; Bowles and Dobbins 2004). Transient solutions of the Semi-MP reliability model are still in the evolutionary phase (Liminios 1997; Kumar et al. 2013a). The reliability analysis for a power-plant was carried using Semi-MP model with Weibull sojourn time (Perman et al. 1997) and the recursive procedures were employed to approximate transition probabilities and average availability. Relationships for reliability function, point availability and interval availability for Semi-MP using the integral approach were suggested (Csenki 1995) but these are difficult to apply for a complex system.

The Semi-MP has also been employed for reliability analysis in the areas of software (Koutras and Platis 2010), wind energy system (D'Amico et al. 2015), turbine blade (Kharoufeh et al. 2010), port transportation systems (Kolowrocki and Soszynska 2006), etc. Also, Semi-MP was applied for availability assessment of manufacturing system (Loganathan et al. 2016), performance modelling of flexible manufacturing system (Ahmed et al. 1989), condition-based maintenance policy optimization (Chen and Trivedi 2005; Kumar et al. 2018), availability analysis with opportunistic maintenance (Kumar et al. 2014), etc. Hybrid approaches such as Semi-MP and Bayesian belief networks (Maura and Droguett 2008), and Markov and Semi-MP (Vulpe and Carausu 2004) were also proposed for availability assessment.

These attempts for reliability and availability solutions with approximations and assumptions lack exactness. Although the stochastic simulation approaches are time-consuming, yet these are easier to implement. Recent advances in high-speed computations have made these approaches easy and solutions are possible within a shorter time and with the desired accuracy. Moreover, no implicit assumption is required for the system behaviour when Monte Carlo simulation (MCS) technique is employed to analyse a system. As dependencies among components and use of non-exponential distributions for failure/repair time lead to mathematical complexity in obtaining a closed-form solution, MCS approach provide a better alternative (Zio et al. 2007).The MCS is relatively an easier way to assess system mission reliability (Lewis et al. 1989). The MCS based model has been used to predict the reliability of microwave systems (Christou 2008), availability evaluation of multi-state systems with dependency features during operation (Zio et al. 2007) and to decide optimal interval for medical devices to maximize the equipment's achieved availability when preventive maintenance is practiced (Cruz and Rodriguez Duenas 2017).

In this paper, the system failure and repair behaviour are modelled by Semi-MP while the reliability and availability solutions are obtained by MCS. Due to structural dependencies, in typical industrial scenarios, repairs are carried out at the unit level which has failed due to one or more of its constituent components. The time to repair of each unit includes disassembly, repair/replacement of the components and assembly of the unit undergoing maintenance. It is, therefore; appropriate to consider the time to repair distribution at the unit level rather than at the component level. Depending upon the number of repair crews, the reliability and availability measures are to be evaluated at the different repair rates. It is assumed that repair will bring the unit to as good as new state. Commercially available software such as RAPTOR, Blocksim, etc., require failure as well as repair distribution parameters at the component level to analyse the system reliability (Kumar et al. 2013a, b). However, suggested approach considers the failures at the component level and repairs at the unit level.

Additionally, there are some demerits of RAPTOR and Blocksim software. The use of a single model for a complex system is error-prone and risky, and modelling in case of special cases such as standby, phasing, k-out-of-n redundancy, etc. is difficult when these softwares are employed. Moreover, the convergence of results is highly dependent on the consistency of the block failure rates (Brall et al. 2007). These limitations of the RAPTOR and Blocksim softwares are addressed in the proposed methodology. Hierarchical modelling addresses the problem of complex systems and separate algorithms are suggested for special cases in the proposed methodology.

The main contribution of the paper is modelling structural dependency that leads to reduced state space. Also, the expressions are derived for system downtime (SDT) for series, parallel and k-out-of-n redundancy configurations in terms of time to repairs (TR) of constituent units of the system to evaluate system availability. The suggested approach is demonstrated with an illustrative example for a centrifugal pump system.

The remaining paper is organized in 6 sections. In Sect. 2, an overview of Semi-MP and MCS method is presented. Semi Markov modelling is proposed in Sect. 3. In Sect. 4, the solution and analysis of the model are described and methodology is also proposed. In Sect. 5, reliability and availability analysis is carried out for a descriptive example. Usefulness and limitations of the technique are discussed in Sect. 6. Finally, Sect. 7 provides the conclusion.

2 Semi-Markov process and Monte Carlo simulation-overview

An overview of the Semi-MP process and MCS is briefly described below.

2.1 Semi-Markov process

A stochastic process is defined as Semi-MP when it possess the memory less property at each conversion instance and not at all intervals (Kulkrani 2009). On the other hand, continuous time Markov Chain is memoryless at all times. In other words, Semi-MP only lose their memory in a state transition unlike the Markov processes (MP), which are regenerative in every single instance. In Semi-MP, the sequential holding of states is decided by the conversion likelihoods of a Markov process and it is called embedded Markov process. However, holding time for every state is defined as "a random variable that depends on the present state as well as on the state to which the next transition will occur". The future behaviour of Semi-MP is governed by the present state and its sojourn time. It can trail any distribution, not only exponential (Fink and Zio 2013).

A Semi-MP is defined as a two random variable sequence, $\{(Y_n, T_n): n \in 1, 2,\ldots\}$, that satisfies following conditions (Kulkarni 2009):

- Y_n, is a discrete-time Markov chain, it belongs to set of feasible states of the system that represents its state after the nth transition.
- The sojourn-time, $T_n - T_{n-1}$, is a random variable that represents time between two transitions and its dispersal depends on the current state only viz, Y_n and also on the state afterwards the subsequent transition:

$$\begin{aligned} &\text{Prob}\{Y_{n+1} = j, T_{n+1} \quad - \quad T_n, \leq t | Y_0, \ldots, Y_n = i; T_O, \ldots, T_n, \} \\ &= \text{Prob}\{Y_{n+1} = j, T_{n+1} \quad - \quad T_n, \leq t | Y_n = i\} = R_{i,j}(t) \end{aligned}$$

Here, the Semi-MP kernel, R_{ij} (t) is expressed as:

$$R_{ij}(t) = q_{ij} \cdot F_i(t)$$

here, q_{ij} is transition probability of the Semi-MP and $F_i(t)$ is the cumulative distribution function for the holding time.

States of the system are generated according to the above Semi-MP procedure.

Failure/degradation of mechanical components/subsystems fit into Semi-MP and hence, its suitability for the mechanical systems. The analytical transient solution of the Semi-MP model is difficult than the simple Markov model because of the complexity in solving the convolution products. The simulation methods although time-consuming, yet are simpler than the analytical methods available in the literature (Zio et al. 2007). The simulation method used in this paper is based on MCS, which is introduced in the next sub section.

2.2 Monte Carlo simulation

In Monte Carlo simulation method probability of certain outcomes is approximated by running multiple trial runs (Billinton and Allan 2007). This technique was first used in World War II and named for Monte Carlo, a city in Monaco, famous for Casino (a game of chance). MCS is used to solve problems of various disciplines and is more useful when it is difficult to use other mathematical techniques. Initially, input variable distributions are selected. These distributions are used to generate a set of random variables in the form of random numbers. The system behaviour is simulated with the set of output obtained from a set of input random numbers. The different sets of possible output values are obtained by repeating the process with different set of input random numbers. National Aeronautics and Space Administration (NASA) developed a simulation program for system reliability assessment by generating block failure and repair times in terms of exponential and Weibull distributions.

The MCS is employed when individual probabilities at component level are known but the system model is complex to obtain closed form analytical solution. In other word, the MCS approach does not provide an equation for probability of success rather evaluates the probability of success at system level from the individual probabilities at component level. In MCS, large amount of repetitive trials are essential to converge results and it is quickly performed by the high computation machines. Many ideologies of probability and probability transformation are applied in MCS implementation. One such ideology is the law of large numbers, "It states that the sample mean is a good estimate of the population mean" if sample size is large (Billinton and Allan 2007). In MCS approach the system behaviour over time is analysed using statistical distributions for the time to failure and repair of individual components/subsystems. In this paper, failure time at unit level is based on failure time of components within the unit, with Weibull distribution. However, due to structural dependency, the time to repair is directly generated for the unit from its exponential distribution. Further, a sequence is generated using the failure and repair time at the unit level to find system states. MCS based algorithms are suggested in this work for various system configurations in Sect. 4.

Subsequent section deals with development of a system model using Semi-Markov approach.

3 Reliability and availability: Semi-Markov models

Let the system consists of several units and each unit consists of several components. Refer Fig. 1, for the system hierarchy. In the system structure, 'S' denotes the system, 'U_i' represents the ith unit, with i = 1 to n and C_j^i denotes the jth component of ith unit, with j = 1 to N. The total number of components in a unit, N may be different for different units. The units can be subassembly or subsystem which undergoes repair if one or more of its component fail. Hence, the unit's failure is dependent on the component(s) failure. The series structure is assumed for components in the unit. The state of a component is assumed as either 'operating' or 'failed'. However, at the unit level, the state of unit may be 'operating', when all its constituent components are in 'operating' state; 'failed'/ 'under repair' state, when one or more of its component has failed and the unit is under repair or 'in standby'; with the active unit operating.

In structural dependency, once the unit is failed due to failure of one or more of its component the entire unit is taken out for repair and not just the failed components. For such a repairable system, the model is developed considering the failures at the component level, while the repair at the unit level. The system states are derived based on the states of the units. A combinatorial theory is used to obtain the total number of states in a system. If binary states are assumed at unit level, then the total number of states of the system with 'n' units is 2^n. The possible states in the system structures like series, parallel system with active redundancy and k-out-of-n redundancy are given in Table 1.

The operating and failed state of the system is represented by 'O' and 'F' respectively. The symbol ^nCr is the combinatorial expression represents the number of selections out of 'n' picking r at a time i.e. $^nC_r = n!/r!(n-r)!$ and r is the number of failed units, which varies from '0' to 'n'. In the case of the series combination of n units; out of the total number of system states 2^n, only one state is operational, when all the units are functioning. In the case of parallel (active redundancy) of n units, out of the total number of system states 2^n, only one state is failed, when all the units are failed. In the case of parallel (k-out-of-n redundancy) combinations, the number of operating states are; $^nC_0 + {}^nC_1 + {}^nC_3 + \cdots + {}^nC_{n-k}$, where 'k' is the minimum number of operating units required for the operation. These states are shown in Table 1.

In standby redundancy structure, if 3 conditions are assumed at unit level, i.e. operating, failed and standby, then the total number of possible situations for the system with 'n' units are 3^n minus the combinations which are not possible. In the case of standby states, few combinations are not possible, e.g. all the units cannot be in standby states at a time and also all the units cannot be in operating state simultaneously. In case of standby systems with perfect switching for two-unit system, there are five possible states, as listed in Table 2. This is arrived by subtracting 4 infeasible combinations from 3^2, i.e. 9. The infeasible combinations are: when both units are considered in standby, both are operational, when first is in standby and second is in under repair and vice versa.

Once the states of the system in terms of states of the units are identified, a model showing the transition rate of failure and repair among the various states is developed. In the transition rate diagram, the transition rate for the failure of a unit is a function of Weibull parameters of components, i.e. $\lambda_i = f(\beta_1, \theta_1, \ldots, \beta_N, \theta_N)$. While the repair rate, μ_i, is considered from the repair time distribution of the entire unit, which includes time of disassembly for the unit, repair action on a failed component and assembly for the unit. This not only gives the reduced state space in the model but also ease in the solution. Since the transition rates are non-exponential, this model is called Semi-Markov model. The solution procedure suggested in this work is based on stochastic simulation, which is detailed in the next section.

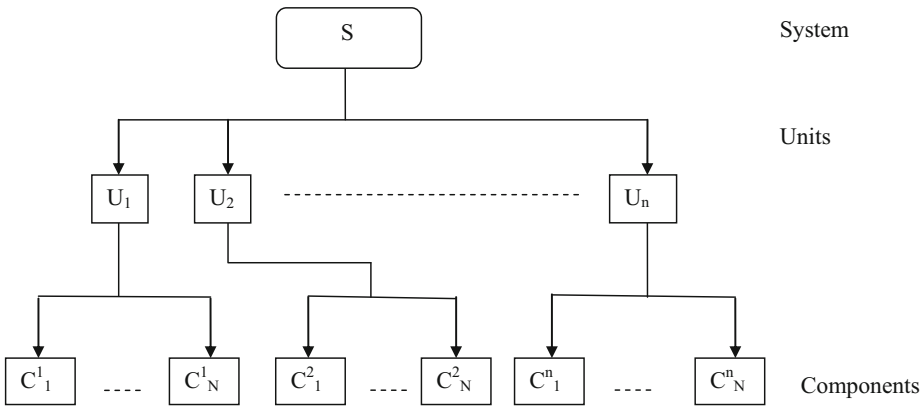


Fig. 1 System hierarchical structure

Table 1 States of n-unit system with series, parallel (active redundancy) and parallel (k-out-of-n redundancy) arrangements

Number of the failed unit in the system 'i'	No. of states	System state		
		Series arrangement	Parallel (active redundancy)	Parallel (k-out-of-n redundancy
0	${}^{n}C_0 = 1$	O	O	O
1	${}^{n}C_1 = n$	F	O	O
2	${}^{n}C_2 = n(n-1)/2$	F	O	O
–	–	F	O	O
–	–	F	O	O
n−k	${}^{n}C_{n-k} = n(n-1)(n-2)\ldots..(n-k+1)/k!$	F	O	O
n−k + 1	${}^{n}C_{n-k+1} = n(n-1)(n-2)\ldots..(n-k+2)/(k\text{-}1)!$	F	O	F
–	–	F	O	F
–	–	F	O	F
n − 2	${}^{n}C_{n-k} = n(n-1)/2$	F	O	F
n − 1	${}^{n}C_{n-k} = n$	F	O	F
n	${}^{n}C_n = 1$	F	F	F
Total states	2^n	2^n	2^n	2^n

O Operating, *F* failed

Table 2 States of two-unit system with standby redundancy

State	State of the unit '1'	State of the unit '2'	State of the system
1	O	S	O
2	F	O	O
3	S	O	O
4	O	F	O
5	F	F	F

O Operational, *F* failed, *S* standby

4 Solution and analysis: Monte Carlo simulation (MCS)

The state's probability transient solution in case of Semi-MP model is too cumbersome by analytical methods due to the complicated convolution of integrals involved in it. Therefore, the solution is obtained by the MCS, which is easier to implement than the analytical approaches. In the first instance, failure time for the components and repair time for the units are simulated by the converting random numbers (RNs) using suitable transformation technique. The converted random variable (RV) is denoted as 'T', failure/repair time. The procedure is detailed in the next paragraph.

Initially, the random numbers are generated in the interval (0, 1) based on uniform distribution. As per the classification behaviour, the uniform RNs are transformed into RVs of the appropriate non-uniform distributions. Inverse transform method, acceptance rejection method and composition methods are the most preferred conversion methods in the literature (Billinton and Allan 2007). The Inverse transform method being efficient is employed for this study. However, it is applicable if the distribution is invertible analytically. In this method, cumulative distribution function (cdf), $F_T(t)$ of some distribution that we use for sampling is considered initially. The function takes't' as input and gives the probability of obtaining $T \leq t$. Therefore,

$$F_T(t) = \Pr(T \leq t) = p$$

The inverse of such function, F_T^{-1} would take 'p' as input and return't' as output, i.e.

$$t = F_T^{-1}(t).$$

It is assumed that probability values 'p' is uniformly distributed. This could be used for any distribution provided it is invertible. Since Weibull and exponential distributions are considered for failure and repair time in this work, the sampling for these distributions is explained here.

The probability density function for Weibull distribution is:

$$f_T(t) = \left(\frac{\beta}{\theta}\right)\left(\frac{t}{\theta}\right)^{\beta-1} e^{\left(\frac{t}{\theta}\right)^{\beta}}$$

where 'θ' and 'β' are Weibull parameters.

Following the inverse transform technique:

$$U = F_T(t) = 1 - e^{-\left(\frac{t}{\theta}\right)^{\beta}} \text{ and } t = \theta(-\ln(1-U))^{\left(\frac{1}{\beta}\right)}$$

where F_T is the cumulative distribution function (CDF) and U is the random variable which is uniformly distributed over the interval (0,1).

Since the random variable (1-U) follows distribution similar to U, then:

$$t = \theta(-\ln(U))^{\left(\frac{1}{\beta}\right)}.$$

On the similar lines, the random variable't' (time to failure) that follows exponential distribution using this method is derived as: t =−(1/λ) ln (U).

Once the component failure time and the unit repair time are simulated according to the appropriate distribution, a system failure and repair time sequence is created for the mission time using these time values. Simulations are repeated to converge the result for the desired accuracy. Reliability for the given mission time is assessed as the fraction of number of successful missions to total number of simulations. In each simulation, the mission is successful when the system never enters into the failed state for a given mission time, else mission is a failure. Availability measure is obtained by dividing total up time with cumulative mission time of all the runs.

In the following Sects. 4.1–4.4, system down time (SDT) expressions are derived for commonly used structures.

4.1 Series arrangement

In case of series arrangement, the system remains in operating state when all the units are working. For reliability evaluation, the mission is a failure if any of the unit fails before mission time, otherwise it is success. For availability, the SDT is evaluated as:

$$\begin{aligned} SDT = &\sum_{i=1}^{n} TR_i - \sum_{1\le i<j}^{n} TR_i \cap TR_j + \\ &\sum_{1\le i<j<k}^{n} TR_i \cap TR_j \cap TR_k - \cdots \\ &+(-1)^n TR_i \cap TR_j \cap TR_k \cdots \cap TR_n \end{aligned} \tag{1}$$

where SDT: System down time for a given mission time, TR_i: Total repair time for unit 'i' for a given mission time, where i = 1 to n

In this expression, the possibility of simultaneous repairs due to failures of more than one unit at a time is also considered. The Eq (1) is a generic expression for system down time with 'n' units configured in series. In deriving this expression set theory concept is utilized. The system down town will be obtained by adding the repair time of all individual units, subtracting the overlapping repair period taken two at time, adding the overlapping repair period taken three at time and so on till the term containing overlapping repair period of all 'n' unit is attained. For example, in case of two unit serial systems, the SDT will be obtained by adding the repair time of two units and subtracting the overlapping period in their repair.

4.2 Parallel (active redundancy) arrangement

In case of a parallel (active redundancy) arrangement, the system enters into failed state when all the units fail. For reliability evaluation, the mission is considered as failure if all the units are under repair simultaneously; otherwise, it is a success. For availability, the SDT is evaluated as:

$$SDT = TR_i \cap TR_j \cap TR_k \cdots \cap TR_n \tag{2}$$

The Eq (2) is a generic expression for system down time with 'n 'units configured in parallel and shows the overlapping repair period for 'n' units. For example, in case of two unit parallel systems, the system down town (SDT) will be overlapping period in their repair.

4.3 Redundant arrangement with k-out-of-n redundancy

In case of parallel (k-out-of-n redundancy) arrangement, the system remains in working states when 'k' or more than 'k' units are working simultaneously. For reliability evaluation, the mission is considered as failure if more than 'n − k' units fail at a time during the mission, otherwise it is a success. For availability, the SDT is evaluated as:

$$\begin{aligned} SDT = &\left\{ \sum_{1\le i<j<k\cdots<i=n-K+1}^{n} TR_i \cap TR_j \cap TR_k \cdots \cap TR_{i=n-K+1} - (n-K+2) \sum_{1\le i<j<k\cdots<i=n-K+2}^{n} TR_i \cap TR_j \cap TR_k \cdots \cap TR_{i=n-K+2} \right\} \\ &+ \left\{ \sum_{1\le i<j<k\cdots<i=n-K+2}^{n} TR_i \cap TR_j \cap TR_k \cdots \cap TR_{i=n-K+2} - (n-K+3) \sum_{1\le i<j<k\cdots<i=n-K+3}^{n} TR_i \cap TR_j \cap TR_k \cdots \cap TR_{i=n-K+3} \right\} \\ &+ \cdots + \left\{ \sum_{1\le i<j<k\cdots<i=n-1}^{n} TR_i \cap TR_j \cap TR_k \cdots \cap TR_{i=n-1} - (n) TR_i \cap TR_j \cap TR_k \cdots \cap TR_n \right\} + \left\{ TR_i \cap TR_j \cap TR_k \cdots \cap TR_n \right\} \end{aligned} \tag{3}$$

The Eq (3) is a generic expression for system down time with 'k-out-of-n redundancy' configuration. First term in parenthesis is the overlapping repair period of 'n − k + 1' units. This is obtained by summing the overlapping repair periods of 'n − k + 1' units and subtracting overlapping repair period of 'n − k + 2' units each with 'n − k + 2' such combinations. Similarly, the second term in parenthesis is the overlapping repair period of 'n − k + 2' units. This is obtained by summing the overlapping repair periods of 'n − k + 2' units and subtracting overlapping repair period of 'n − k + 3' units each with 'n − k + 3' such combinations. This procedure is repeated till the last term with overlapping repair period of all 'n' units is obtained.

For example, in case of '3-out-of-5 redundancy' systems, the system down town will be obtained by summing the down time of 3, 4 and 5 units at a time. The down time with 3 units is obtained by summing the overlapping repair periods of 3 units and subtracting overlapping repair period of 4 units each with 4 such combinations. The down time with 4 units is obtained by summing the overlapping repair periods of 4 units and subtracting 5 times overlapping repair period of 5 units. Finally, the down time with 5 units is the overlapping period of 5 units repair time.

The Eq. (3) can be simplified as

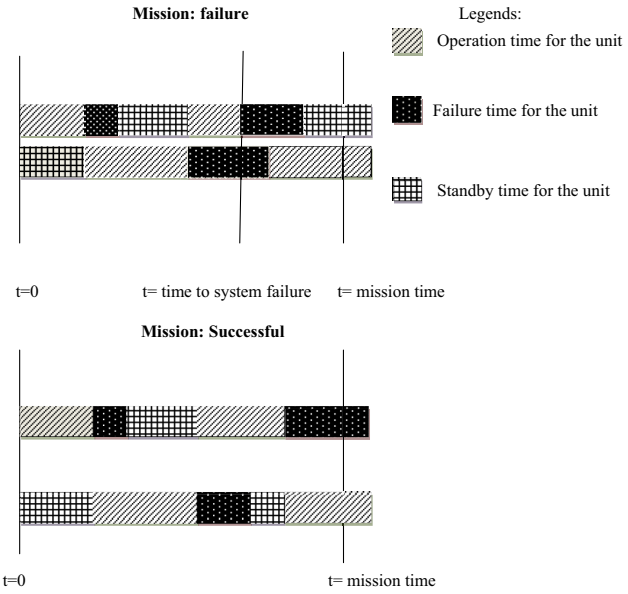


Fig. 2 Single cycle/sequence of the system run for the mission time

before the mission is completed, then mission is a failure, otherwise it is considered as success. Refer Fig. 2: for mission failure and mission success in a single cycle/sequence. Similar, cycles are generated for large number of simulations till desired accuracy is achieved.

$$SDT = \left\{ \begin{array}{l} \sum\limits_{1 \le i<j<k\cdots<i=n-K+1}^{n} TR_i \cap TR_j \cap TR_k \cdots \cap TR_{i=n-K+1} - (n-K+1) \sum\limits_{1 \le i<j<k\cdots<i=n-K+2}^{n} TR_i \cap TR_j \cap TR_k \cdots \cap TR_{i=n-K+2} \\ -(n-K+2) \sum\limits_{1 \le i<j<k\cdots<i=n-K+3}^{n} TR_i \cap TR_j \cap TR_k \cdots \cap TR_{i=n-K+3} - \cdots \\ -(n-1) TR_i \cap TR_j \cap TR_k \cdots \cap TR_n \end{array} \right\} \quad (4)$$

In Eq. (4), system down time is obtained by considering the repair overlapping time with terms containing n − k + 1 units overlap, n − k + 2 units overlap,…, and n units overlap.

4.4 Standby redundancy

In case of standby repairable system from reliability point of view; if both units are under repair simultaneously

The reliability after the desired number of simulations is calculated; as the ratio of cumulative count of successes to the total number of simulations performed.

For availability measure, a sequence of failure and repair is generated for the duration at least equal to mission time. The system down time is evaluated in each cycle. These are repeated till the desired accuracy is achieved. After the desired number of simulations, the availability

measure is evaluated by dividing the total system up time with total time in all runs.

Solution procedure described in this section is illustrated with the help of an algorithm detailed in the following subsection.

4.5 Algorithm

In the literature, algorithms employed for reliability and availability assessment are restricted to one hierarchical level of the system, i.e. state of the system depends on the combination of the states of its components. Also, studies are limited to use of exponential distribution in modelling. The algorithms suggested in this work for reliability and availability are extended to two hierarchical levels, i.e. units and components of the system. Further, Weibull distribution is assumed for the failure time of component(s), while exponential distribution is considered for unit(s) repair time. The repair is considered at the unit level only. The step by step details of the algorithm are given in the following paragraph.

4.5.1 Reliability algorithm

R_1 For 'N' number of serially connected components in the unit 'U_1' use uniform distribution to generate random numbers

R_2 Convert these numbers into time values based on inverse transform technique considering Weibull failure distribution for each component

R_3 Select the minimum out of these failure time values of the components in unit 'U_1' to get the unit time to failure in case of series combination of components in the unit

R_4 Repeat steps R_1–R_3 for the remaining units in the system, i.e. U_2, U_3…,U_n

R_5 For the first unit 'U_1', generate a random number based on uniform distribution

R_6 Transform this number into a repair time value based on inverse transform method considering exponential repair time distribution of unit 'U_1'

R_7 Repeat steps R_5–R_6 for remaining units in the system, i.e. U_2, U_3…,U_n

R_8 Create a sequence with failure and repair time for all units in the system

R_9 Repeat the steps R_1–R_8 till cumulative sequence time is greater than or equal to mission time

R_{10} Success or failure of the mission depends upon the configuration of the units in the system and for various system structures it is as:

(i) Series arrangement: The mission is considered failure, if the time to failure of any unit is less than the mission time during first sequence, otherwise it is a success.

(ii) Parallel (active redundancy) arrangement: The mission is considered failure, if the time to failure of all the units is less than mission time during the first sequence, otherwise it is a success.

(iii) Redundant arrangement with k-out-of-n redundancy: a mission is considered failure when the time to failure of more than 'n − k' units is less than the mission time during the first sequence, otherwise it is a success.

(iv) Standby redundant configuration: The system is considered as failed during the mission time when all the units are under repair due to their failure at a time and the mission is considered as failure, otherwise mission is a success.

R_{11} Reliability is calculated as division of cumulative number of successful missions by the total number of runs for the mission time

R_{12} R_1–R_{11} steps are repeated until the value of reliability converges

4.5.2 Availability algorithm

A_1 For 'N' number of serially connected components in the unit 'U_1' use uniform distribution to generate random numbers

A_2 Convert these numbers into time values based on inverse transform technique considering Weibull failure distribution for each component

A_3 Select the minimum out of these failure time values of the components in unit 'U_1' to get the unit failure time in case of series combination of components in the unit

A_4 Repeat steps A_1–A_3 for remaining units in the system, i.e. U_2, U_3…,U_n

A_5 For the first unit 'U_1', generate a random number based on uniform distribution

A_6 Transform this number into a repair time value based on inverse transform method considering exponential repair time distribution of unit 'U_1'

A_7 Repeat steps A_5–A_6 for remaining units in the system, i.e. U_2, U_3,…,U_n

A_8 Create a sequence with failure and repair time for all units in the system

A_9 Repeat the steps A_1–A_8 till the cumulative sequence time is greater than or equal the mission time

A_{10} The time for which the system remains operational depends upon the configuration of the units in the system and for various system structures it is as:

(i) Series arrangement: For a mission time, the system remains in upstate when all the units are operational during the mission and the corresponding time is the uptime. This is determined by subtracting SDT from the total mission time using Eq. (1)

(ii) Parallel (active redundancy) arrangement: For a given mission time, the system remains in downstate when all the units are failed. The duration of downstate is considers as system down time. The SDT is determined as the difference between mission time and uptime using Eq. (2)

(iii) Redundant arrangement with k-out-of-n redundancy: A mission is considered as success if there are 'k' or more than 'k' units in operational state at a time. The SDT is the time when system remains in failed state and it is calculated using Eq. (3). Further, the uptime is calculated by subtracting the SDT from the mission time

(iv) Standby configuration: The system is considered as failed during the mission time when all the units are under repair due to their failure at a time. The SDT is the time when the system is in failed state. The uptime is calculated by subtracting the SDT from the mission time.

A_{11} Availability measure is determined dividing cumulative uptime by the total mission times of completed missions

A_{12} Steps A_1–A_{11} are repeated till the availability value converges

A MATLAB code is developed using these algorithms for the configurations described in Sect. 4.1–4.4. However, the results are obtained and shown only for the standby redundancy case in graphical and tabular form.

4.6 Methodology

The step-by-step methodology which is based the discussion in Sects. 3 and 4 is given as below:

Step 1 Consider the system. Identify its units and constituent components and their arrangement. Refer Sect. 3

Step 2 Define Weibull parameters for the random variable, time to failure; 'θ_j^i' and 'β_j^i' for the jth component within the ith unit

Step 3 Define distribution parameters for time to repair of each unit

Step 4 Identify all possible states of the system in terms of states of the constituent units and Refer Sect. 3

Step 5 Develop a Semi-Markov model showing the transitions among the states as illustrated in Sect. 3

Step 6 Solve the model using MCS using the solution procedure/algorithms suggested in Sects. 4.1 to 4.5

5 Illustrative examples

A two pump standby system is chosen to illustrate the suggested approach for the analysis of availability and reliability. The schematic diagram of this system is given in Fig. 3. This system is a part of a centralized AC plant that operates 24 h a day and round the year. The suggested approach is used to assess the availability and reliability for different mission time. The Weibull failure data for different pump components is collected as per the data source (Barringer 2018). For failure data of pump components, refer Table 3. Exponential distribution is assumed for repairs. The time to repair for the pump varies from 12 to 150 h depending upon the severity of faults of pump components and amount of repair resource employed

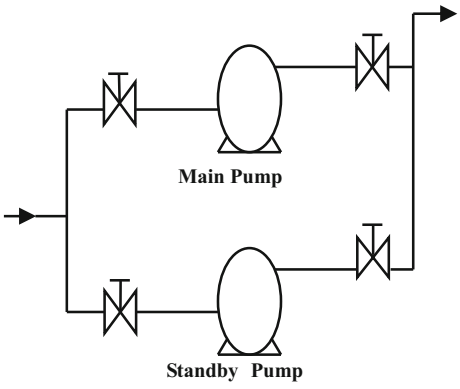


Fig. 3 Schematic diagram-pumping system

Table 3 Failure data for condenser pump components (Barringer 2018)

S.No.	Item	Beta values Weibull shape factor (β)	Eta values Weibull characteristic life-hours (θ)
1	Impeller	0.5	1,25,000
2	Shaft	0.8	50,000
3	Bearing	0.7	14,000
4	Motor	0.5	1,000
5	Bush	0.7	14,000
6	Housing	0.5	1,25,000

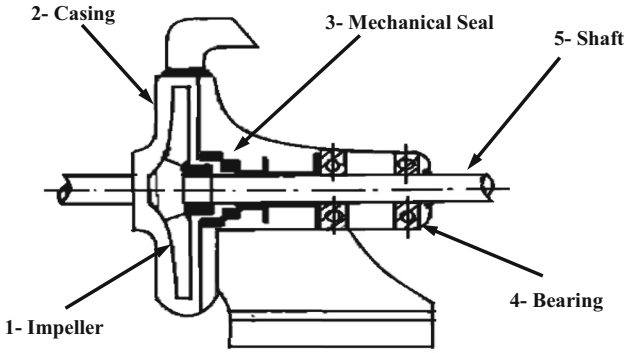


Fig. 4 Main components of the centrifugal pump

(OREDA 1998). The choice of values of Weibull parameters is critical for accuracy of results. In this work the failure and repair distribution parameter values are taken from the literature for illustration purpose. However, in practical applications these parameter values should be estimated from the real field data.

This example is illustrated step-by-step for availability and reliability assessment:

Step1 There are two units in the pumping system, one main pump and another standby unit. Both pumps are identical in its structure. The components of each pump are given in Fig. 4

Step2 Refer Table 3, for Weibull parameters 'θ' and 'β' for failure time of each component

Step3 The repair time distribution is assumed as exponential with parameter 'μ' for the pump is as 0.008, 0.0125 and 0.05 for some typical values of the time to repair for illustration purpose as repair time for the pump varies from 12 to 150 h depending upon the severity of faults and amount of repair resource employed (OREDA 1998)

Step4 There are five possible states in the system (Refer Table 2)

Step5 The transition rate diagram of two units system (one in standby) with perfect switching is analysed and is as shown in Fig. 5. λ_i^j and μ_i^j represent failure rate and repair rate respectively for ith unit at jth hierarchy level

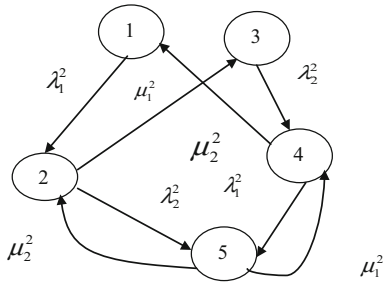


Fig. 5 State transition rate diagram of two units standby system with perfect switching

Table 4 Reliability/availability results (time = 1000 h)

Repair rate (repairs/h)	Reliability/availability values	
	Reliability	Availability
0.0080	0.6044	0.9265
0.0125	0.6426	0.9583
0.0500	0.7624	0.9935

Step6 The model developed in step 5 is solved using MCS. Refer Sect. 4.5. The results obtained are given in Table 4. The results are also graphically shown in Figs. 6 and 7

The analysis was done to investigate impact of employing additional maintenance resources (increased repair rate) to reduce the repair time for improving system reliability and availability. Refer Table 4, the reliability and availability values follow an increasing trend with increase in values of repair rates. Subsequent result analysis is presented in Table 5. As the repair rate is increased from 0.008 to 0.0125 (reducing repair time from 125 to 80 h), i.e., 1.56 times by employing extra repair resource, the % increase in reliability is 6.3 (i.e. from 0.6044 to 0.6426). The increase in reliability due to enhanced resource deployment is obvious; however, the model provides improvements in terms of actual figures. Further increasing repair rate from 0.0125 to 0.05 (reducing repair time from 80 to 20 h), i.e., 4 times, the % increase in

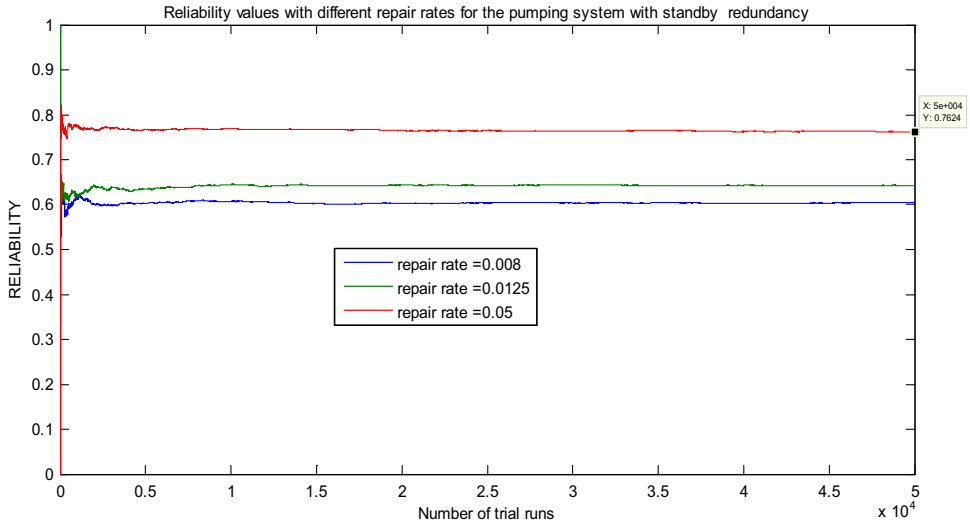


Fig. 6 Pumping system reliability values for mission time 1000 h with different repair rates

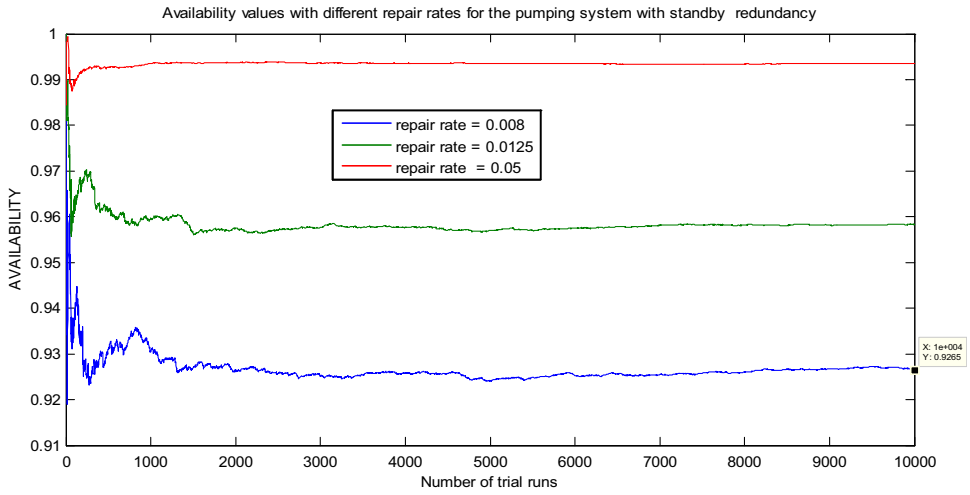


Fig. 7 Pumping system availability values for mission time 1000 h with different repair rates

Table 5 Results-analysis

Incremental increase in repair rate	% Increase in reliability	% Increase in availability	Rate of increase in reliability with respect to repair rate	Rate of increase in availability with respect to repair rate
1.56 times	6.3	3.4	8.5	7.1
4 times	18.6	3.6	3.2	0.94

reliability is 18.6 (i.e. from 0.6426 to 0.7624). On the other hand, for the same increase in repair rates, there is not much % increase in availability, i.e., 3.4% and 3.6%. These results are shown in Table 5. The results are also analysed by considering rate of increase in reliability and availability w. r. t. repair rate. The rate of increase in reliability and availability w. r. t. change in repair rate shows a decreasing trend for both. This is expected as the increase in repair rate by employing additional man power will not always result in substantial increase in reliability and availability. In view of this; beyond a certain limit, the benefit in reliability and availability becomes insignificant even with additional cost due to increased manpower. Therefore, replacement may be a better option than doing the subsequent repairs. The suggested approach is useful for designers and maintenance personnel in taking decisions. The cost has not been considered in the analysis. It will be meaningful to include it for the maintenance or repair actions to analyze the net gain due to enhance system availability.

6 Utility and limitation of the method

The suggested Semi-MP methodology is helpful for reliability and availability assessment "when the time to failure and the time to repair of components/subsystems within a system fit to a non-exponential distribution and the Markov approach becomes inapplicable". The Semi-MP method is suitable in case of mechanical components as their deprivation rate rises due to the aging process that is modelled by Weibull distribution. Although, an illustration with pumping system assuming Weibull and exponential distributions for failure and repair is provided in this work, yet the suggested method may be useful to any other system with non-exponential behaviour. The proposed approach furnishes realistic results for reliability and availability due to the selection of relevant and general distributions for failure and repair time. The MCS method does have some limitations. Generally, the simulation approach is, difficult to implement and takes more time for complex systems. On the contrary, analytical methods have fast and easy implementation. However, analytical methods also become complex when the assumptions in the model are relaxed. In case general distributions are considered in analytical approaches, the involved integrals and convolutions in such methods are hard to solve and the reliability and the availability evaluation cannot be done. In such situations, MCS approach is a workable alternative.

7 Conclusion and future directions

In this paper, Semi-MP based modelling and MCS based solution is suggested for reliability and availability evaluation of repairable systems. The paper has addressed the problem of state space explosion in state based models by merging the component states in unit states. Moreover, structural dependency is modelled successfully by consideration of repairs at the unit level rather that at the component level. The expressions for system down time (SDT) for series, parallel and k-out-of-n redundancy configurations are proposed in terms of time to repairs (TR) of constituent units of the system to evaluate system availability. These are unique contributions of this paper. The suggested approach is easy to implement as compared to the analytical solution. The modelling is helpful in selecting the different parameters to improve the various dependability attributes and gives accurate valuation of reliability and availability for mechanical systems. The proposed approach will be helpful to maintenance engineers in taking appropriate decisions. The approach can assist them in formulating suitable maintenance/replacement strategies. The designers can also apply the suggested approach in decision making for incorporating the type and the level of redundancy, provisioning of repair resource to achieve the system reliability and availability goals.

The suggested methodology is limited in its application to the systems with binary state of components (working or failed) and perfect repair of units (to restore it to as good as new). However, in practical situations, there is gradual deterioration of mechanical components and the failed units of the system can be restored to various level of its health. Therefore, in future work the modelling can be strengthened by considering multi states of the components with gradual deterioration and incorporating different types or levels of repairs of units such as; perfect, imperfect and minimal. Further, for repairs; lognormal distribution can be considered which is more realistic for mechanical systems.

References

Ahmed SB, Alam M, Gupta D (1989) Performance modelling and evaluation of flexible manufacturing systems using a Semi-Markov approach. Int J Comput Integr Manuf 2(5):275–280

Barringer (2018) Weibull database. http://www.barringer1.com/wdbase. Accessed 11 Dec 2018

Billinton R, Allan RN (2007) Reliability evaluation of engineering systems: concepts and techniques. Springer, New Delhi

Bowles JB, Dobbins JG (2004) Approximate reliability and availability models for high availability and fault tolerant systems with repair. Qual Reliab Eng Int 20(7):679–697

Brall A, Hagen W and Tran H (2007) Reliability block diagram modeling—comparisons of three software packages. In: 2007 annual reliability and maintainability symposium, Orlando, FL, 2007, pp 119–124

Chen D, Trivedi KS (2005) Optimization for condition-based maintenance with Semi-Markov decision process. Reliab Eng Syst Saf 90(1):25–29

Christou A (2008) Monte Carlo reliability model for microwave monolithic integrated circuits. Qual Reliab Eng Int 24(3):315–329

Cruz AM, Rodriguez Duenas WRR (2017) Estimation of the optimal maintenance frequency of medical devices: a Monte Carlo simulation approach. In: Torres I, Bustamante J, Sierra D (eds) VII Latin American Congress on biomedical engineering CLAIB 2016. Bucaramanga, Santander, Colombia, 26–28 October, 2016. IFMBE Proceedings, vol 60. Springer, Singapore

Csenki A (1995) Transient analysis of interval availability for repairable systems modelled by finite Semi-Markov processes. IMA J Math Appl Bus Ind 6(3):267–281

D'Amico G, Petroni F, Flavio P (2015) Reliability measures for indexed Semi-Markov chains applied to wind energy production. Reliab Eng Syst Saf 144(2015):170–177

Dao CD, Zuo MJ (2017) Selective maintenance of multi-state systems with structural dependence. Reliab Eng Syst Saf 159:184–195

Fink O, Zio E (2013) Semi-Markov processes with semi-regenerative states for the availability analysis of chemical process plants with storage units. Proc Inst Mech Eng Part O J Risk Reliab 227(3):279–289

Fricks RM, Trivedi KS (1997) Modelling failure dependencies in reliability analysis using stochastic Petri nets. In: Proceedings of

11th european simulation multiconference. Istanbul, Turkey, 1–4 June

Hajeeh MA (2012) Availability of deteriorated system with inspection subject to common-cause failure and human error. Int J Oper Res 12(2):207–222

Kharoufeh JP, Solo CJ, Ulukus MY (2010) Semi-Markov models for degradation-based reliability. Inst Ind Eng Trans 42:599–612

Kolowrocki K, Soszynska J (2006) Reliability and availability analysis of complex port transportation systems. Qual Reliab Eng Int 22(1):79–99

Koutras VP, Platis AN (2010) Semi-Markov performance modelling of a redundant system with partial, full and failed rejuvenation. Int J Crit Comput Based Syst 1(1/2/3):59–85

Kulkrani VG (2009) Modeling and analysis of stochastic systems. Chapman and Hall, London

Kumar G, Jain V, Gandhi OP (2013a) Availability analysis of repairable mechanical systems using analytical Semi-Markov approach. Qual Eng 25(2):97–107

Kumar G, Jain V, Gandhi OP (2013b) Feasibility of analytical solution for transient availability using Semi-Markov process. Int J Reliab Saf Eng 7(4):388–410

Kumar G, Jain V, Gandhi OP (2014) Steady-state availability analysis of repairable mechanical systems with opportunistic maintenance by using Semi-Markov process. Int J Syst Assur Manag 5(4):664–678

Kumar G, Jain V, Gandhi OP (2018) Availability analysis of mechanical systems with condition-based maintenance using Semi-Markov and evaluation of optimal condition monitoring interval. J Ind Eng Int 14(1):119–131

Lewis EE, Boehm F, Kircsch C, Kelkhoff BP (1989) Monte Carlo simulation of complex system mission reliability. In: Proceedings of the 1989 winter simulation conference, ACM Digital Library, New York, USA. 4–6 December, pp 497–504

Liminios N (1997) Dependability analysis of Semi-Markov systems. Reliab Eng Syst Saf 55(3):203–207

Loganathan MK, Kumar G, Gandhi OP (2016) Availability evaluation of manufacturing systems using Semi-Markov model. Int J Comput Integr Manuf 29(7):720–735

Maura MC, Droguett EL (2008) A continuous-time Semi-Markov bayesian belief network model for availability measure estimation of fault tolerant systems. Pesq Oper 28(2):355–375

OREDA (1998) Offshore reliability data handbook, 2nd edn. DNV, Hovik

Perman M, Senegacnik A, Tuma M (1997) Semi-Markov model with an application to power-plant reliability analysis. IEEE Trans Reliab 46(4):526–532

Vulpe A, Carausu A (2004) Stochastic evaluation of availability for subsystems by Markov and Semi-Markov models. In: Thirteenth world conference on earthquake engineering. Vancouver, B.C. Canada, 1–6 August, p 2483

Welte TM (2009) A rule-based approach for establishing states in a Markov process applied to maintenance modelling. Part O J Risk Reliab Proceed Instit Mech Eng 223(1):1–12

Xie M, Kong H, Goh TN (2004) Exponential approximation for maintained Weibull distributed component. J Qual Maint Eng 6(4):260–268

Zio E, Marelle M, Podofillini L (2007) A Monte Carlo simulation approach to the availability assessment of multi-state systems with operational dependencies. Reliab Eng Syst Saf 92(2007):871–882

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

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/326672903

Molecular characterization, modeling, and docking analysis of late phytic acid biosynthesis pathway gene, inositol polyphosphate 6-/3-/5-kinase, a potential candidate for developin...

Article · August 2018

DOI: 10.1007/s13205-018-1343-7

CITATION
1

READS
144

4 authors, including:



Mansi - Punjabi
Delhi Technological University
19 PUBLICATIONS **26** CITATIONS

SEE PROFILE



Archana Sachdev
Indian Agricultural Research Institute
100 PUBLICATIONS **244** CITATIONS

SEE PROFILE



Veda Krishnan
Indian Agricultural Research Institute
61 PUBLICATIONS **157** CITATIONS

SEE PROFILE

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

Developing low phytate soybean by manipulating pathway genes using CRISPR/Cas system View project

Pullulanase based identification of high resistant starch rice (Oryza sativa L.) View project

All content following this page was uploaded by Veda Krishnan on 14 November 2018.

The user has requested enhancement of the downloaded file.

3Biotech

Molecular characterization, modeling and docking analysis of late phytic acid biosynthesis pathway gene, inositol polyphosphate 6-/3-/5-kinase, a potential candidate for developing low phytate crops

--Manuscript Draft--

Manuscript Number:	BITC-D-18-00255R4
Full Title:	Molecular characterization, modeling and docking analysis of late phytic acid biosynthesis pathway gene, inositol polyphosphate 6-/3-/5-kinase, a potential candidate for developing low phytate crops
Article Type:	Original Article
Section/Category:	Plant Biotechnology
Funding Information:	National Funds for Basic, Strategic and Frontier Application Research in Agriculture (NFBSFARA/RNAi-2011/2011-12) — Dr. Archana Sachdev
Abstract:	The coding sequence of inositol polyphosphate 6-/3-/5-kinase (GmIPK2) gene was identified and cloned from popular Indian soybean cultivar Pusa-16. The clone was predicted to encode 279 amino acids long, 30.97 kDa protein. Multiple sequence alignment revealed an inositol-phosphate-binding motif, PxxxDxKxG throughout the IPK2 sequences along with other motifs unique to inositol phosphate kinase superfamily. Eight α-helices and eight β-strands in antiparallel β-sheets arrangement were predicted in the secondary structure of GmIPK2. The temporal analysis of GmIPK2 revealed maximum expression in the seed tissues during later stages of development while spatially the transcript levels were lowest in leaf and stem tissues. Endosperm-specific cis-regulatory motifs (GCN4 and Skn_1) which support high levels of expression, as observed in the developing seeds, were detected in its promoter region. The protein structure of GmIPK2 was modeled based on the crystal structure of inositol polyphosphate multikinase from Arabidopsis thaliana (PDB:4FRF) and subsequently docked with inositol phosphate ligands (PDB: 5GUG-I3P and PDB: 4A69-I0P). Molecular dynamics (MD) simulation established the structural stability of both, modeled enzyme and ligand-bound complexes. Docking in combination with trajectory analysis for 50 ns MD run confirmed the participation of Lys105, Lys126 and Arg153 residues in the formation of a network of hydrogen bonds to stabilize the ligand-receptor interaction. Results of the present study thus provide valuable information on structural and functional aspects of GmIPK2 which shall assist in strategizing our long-term goal of achieving phytic acid reduction in soybean by genetic modification of its biosynthetic pathway to develop a nutritionally enhanced crop in the future.
Corresponding Author:	Archana Sachdev Indian Agricultural Research Institute INDIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Indian Agricultural Research Institute
Corresponding Author's Secondary Institution:	
First Author:	Mansi Punjabi
First Author Secondary Information:	
Order of Authors:	Mansi Punjabi Navneeta Bharadvaja Archana Sachdev Veda Krishnan

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Order of Authors Secondary Information:	
Author Comments:	COVER LETTER To, The Editor-in-Chief 3 Biotech 1 July, 2018 Dear Sir, Thank you for the opportunity to revise our manuscript as per the thoughtful critiques provided by the esteemed reviewers. We have incorporated all of the suggestions and believe that the manuscript is substantially improved accordingly. The following pages include a detailed response to each reviewer's comments. All the modifications in the revised manuscript are highlighted in red. The final revisions were discussed and approved by all the co-authors. Thank you for your consideration. Yours faithfully, Archana Sachdev Principal Scientist Division of Biochemistry ICAR- Indian Agricultural Research Institute New Delhi-110012, India
Response to Reviewers:	Comments - If any reader wanted to retrieve the ligands from PDB database by typing I3P and I0P in the search box than this would show multiple entries. It is not acceptable. Moreover Authors should know that a PDB id hold the combination of FOUR number or/and alphabeted. I3P and I0P are the small molecules entry id and author should exclusively mentioned the source of that. eg PDB: 5GUG-I3P and PDB: 4A69-I0P. Response: The PDB ID for I3P has been corrected to 5GUG-I3P and I0P to 4A69-I0P.
Keywords:	Glycine max; phytic acid; inositol polyphosphate 6-/3-/5-kinase (IPK2); low phytate crops, spatiotemporal expression; Homology Modeling; Molecular Docking; molecular dynamics simulation

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Manuscript Click here to download Manuscript Manuscript.doc

Click here to view linked References

Molecular characterization, modeling and docking analysis of late phytic acid biosynthesis pathway gene, *inositol polyphosphate 6-/3-/5-kinase*, a potential candidate for developing low phytate crops

3 BIOTECH

MANSI PUNJABI[1,2], NAVNEETA BHARADVAJA[1], ARCHANA SACHDEV*[2], VEDA KRISHNAN[2]

[1] *Department of Biotechnology, Delhi Technological University (Formerly Delhi College of Engineering), New Delhi - 110042, India*

[2] *Division of Biochemistry, Indian Agricultural Research Institute, New Delhi - 110012, India*

**Corresponding author: Dr. Archana Sachdev, Tel: 91-11-25842038, Email: arcs_bio@yahoo.com*

Abstract

The coding sequence of *inositol polyphosphate 6-/3-/5-kinase* (*GmIPK2*) gene was identified and cloned from popular Indian soybean cultivar Pusa-16. The clone was predicted to encode 279 amino acids long, 30.97 kDa protein. Multiple sequence alignment revealed an inositol-phosphate-binding motif, PxxxDxKxG throughout the IPK2 sequences along with other motifs unique to inositol phosphate kinase superfamily. Eight α-helices and eight β-strands in antiparallel β-sheets arrangement were predicted in the secondary structure of GmIPK2. The temporal analysis of *GmIPK2* revealed maximum expression in the seed tissues during later stages of development while spatially the transcript levels were lowest in leaf and stem tissues. Endosperm-specific cis-regulatory motifs (GCN4 and Skn_1) which support high levels of expression, as observed in the developing seeds, were detected in its promoter region. The protein structure of GmIPK2 was modeled based on the crystal structure of inositol polyphosphate multikinase from *Arabidopsis thaliana* (PDB:4FRF) and subsequently docked with inositol phosphate ligands (PDB: 5GUG-I3P and PDB: 4A69-I0P).

Molecular dynamics (MD) simulation established the structural stability of both, modeled enzyme and ligand-bound complexes. Docking in combination with trajectory analysis for 50 ns MD run confirmed the participation of Lys105, Lys126 and Arg153 residues in the formation of a network of hydrogen bonds to stabilize the ligand-receptor interaction. Results of the present study thus provide valuable information on structural and functional aspects of *GmIPK2* which shall assist in strategizing our long-term goal of achieving phytic acid reduction in soybean by genetic modification of its biosynthetic pathway to develop a nutritionally enhanced crop in the future.

Keywords: Glycine max; phytic acid; inositol polyphosphate 6-/3-/5-kinase (IPK2); low phytate crops, spatiotemporal expression; homology modeling; molecular docking; molecular dynamics simulation

Molecular characterization, modeling and docking analysis of late phytic acid biosynthesis pathway gene, *inositol polyphosphate 6-/3-/5-kinase,* a potential candidate for developing low phytate crops

1. Introduction

Soybean is a phenomenal crop with a unique nutrient profile. It is the only plant source which provides a complete protein with quality equivalent to animal protein yet its human consumption is found to be very limited, even in the countries where its production is very high. Besides, majority of its uses in the food industry owe primarily to its functional properties rather than nutritional gain. The above stated can be attributed to the presence of a variety of bioactive components including phytic acid (PA), saponins, protease inhibitors, isoflavones, lectins etc. Amongst these, PA is probably the most well-known antinutrient that is found in abundance in cereal grains and legumes. It is a saturated cyclic acid whose metabolism has been substantiated to regulate phosphorous and myo-inositol homeostasis. Due to its highly negative chemical structure, it forms phytate-mineral-protein complexes which result in reduced bioavailability of phosphorus, other associated minerals and proteins and thus contribute to the debatable health effects associated with its consumption. Therefore, in a need to reduce the level of this compound several approaches were undertaken which blocked its constitutive synthesis but often resulted in crop with deplorable agronomic performance (Bilyeu et al., 2008; Cichy and Raboy, 2009; Feng and Yoshida, 2004; Kuwano et al., 2009, Nunes et al., 2006). Recently trending reverse genetic approaches that target tissue-specific knockdown provide a promising alternative to eliminate this negative impact (Ali et al., 2013; Ali et al., 2013). Research is thus underway to characterize PA biosynthetic pathway enzymes (Josefsen et al., 2007; Krishnan et al., 2015; Stiles et al., 2008; Sun et al., 2007; Sweetman et al., 2006; 2007) and study their expression patterns (Bhati et al., 2014; Fileppi et al., 2010; Suzuki et al., 2007) to generate adequate information for developing *lpa* soybean through metabolic engineering.

PA biosynthesis follows a succession of phosphorylation steps (Brearley and Hanke, 1996a; 1996b).

One of its late pathway enzyme inositol polyphosphate 6-/3-/5-kinase (IPK2, EC 2.7.1.151) is a multiple specificity enzyme which phosphorylates the D-6, D-3, and/or D-5 positions on a variety of inositol polyphosphate substrates (InsP3/InsP4/InsP5) (Frederick et. al, 2005) that makes it a key enzyme in regulating PA turnover and is thus a crucial target for perturbing PA dynamics (Stevenson-Paulik et al., 2005). The enzyme has previously been studied in *Arabidopsis* (Stevenson-Paulik et al., 2002), yeast (Saiardi et al., 2000; York et al., 1999), mouse (Frederick et al., 2005) and human (Majerus et al., 1992; 1996). A study has also been reported in soybean (Stiles, 2007), however, it discusses only the gene's expression and enzyme kinetics. Therefore, with the intention to extend the current knowledge on its transcript expression (both spatial and temporal) as well as to understand its structure, biochemical features and evolutionary history, we investigated this gene in soybean and compared it with those of another nineteen homologs in plants through various bio-computational tools. Furthermore, in order to elucidate complete molecular and biochemical mechanisms regulating PA synthesis as well as inositol metabolism and signalling in plants, an insight into IPK2 protein's structural features is required. Very less information regarding the three-dimensional structure of plant IPK2 is available. In order to facilitate the same, we determined the three-dimensional (3D) model of soybean GmIPK2 protein through homology modeling and performed its docking simulation with the substrate (1D-*myo*-inositol 1,4,5-trisphosphate (PDB: 5GUG-I3P) and 1D-*myo*-inositol-1,4,5,6-tetra*kis*phosphate (PBD: 4A69-I0P)) molecules.

Through this work, we made an effort to study *IPK2* gene to lay a groundwork for achieving the main goal of our work which is, to develop low phytic acid soybean, a nutritionally and agriculturally important crop.

2. Materials and Methods

2.1. Plant Material

Field grown soybeans (Glycine max [L.] Merr. cv. Pusa-16) were procured from Division of Genetics, IARI, New Delhi for use in this study. For spatial and temporal expression profiling, root,

stem, leaf and flower tissues were collected from 30 day old plants while developing seeds were collected regularly after flowering until maturation and sorted based on their sizes. After collection, the tissue samples were immediately frozen in liquid nitrogen and stored at -80°C until used.

2.2. PCR amplification, cloning and sequencing of partial GmIPK2 sequence

Total RNA was isolated from soybean seeds 8 mm in size using TRIzol reagent method (Invitrogen, USA). Approximately, 1 µg of isolated total RNA was used for single stranded cDNA synthesis by reverse transcription with RevertAid^{TM} H Minus First Strand cDNA synthesis kit (ThermoFisher Scientific, USA). Prior to cDNA synthesis, the RNA was treated with RNase-free DNase I (Thermo Scientific, USA) for removal of any residual genomic DNA. Polymerase chain reaction (PCR) amplification was performed using 0.2 µg of the initial RT reaction optimized at 94°C for 4 min; 35 cycles of 94°C for 30 s, 62.5°C for 30 s, 72°C for 30 s; 72°C for 10 min, using the *GmIPK2* specific primers: *IPK2*F 5′-ATGCTCAAGATCCCGGAG-3′ and *IPK2*R 5′-CAGTTAGTCTGCGACACTA ATTCAAGC-3′. On completion, 2 µl of amplified product was analyzed by electrophoresis on 1% agarose gel and purified using gel/PCR DNA Fragments Extraction Kit (DF100) by Geneaid. The purified amplicon was subsequently ligated in pGEM®-T Easy vector (Promega, USA) by TA cloning following protocol described in the manual and used directly for transformation of competent *E. coli* DH5α cells. The recombinant plasmids were identified by blue-white screening and confirmed by restriction digestion with *EcoRI* to release the amplicon fragment. The nucleotide sequences of *GmIPK2* thus isolated was determined by DNA sequence analysis of overlapping plasmid clones using universal primers (SP6 and T7) on an automated sequencer (ABI 3730xl DNA Analyzer, USA). The nucleotide sequence data was submitted to INSDC database GenBank.

2.3. Gene expression analysis by semi-quantitative reverse transcription PCR and quantitative real-time PCR

To analyze *GmIPK2* gene expression in different soybean tissues, we first performed semi-quantitative reverse transcription PCR (RT-PCR) to obtain an expression pattern and then further

estimated the transcript levels by quantitative real time PCR (qRT-PCR). For RT-PCR, first strand cDNA was synthesized from total RNA using RevertAid^{TM} H Minus First strand cDNA Synthesis Kit (Thermo Scientific, USA) and PCR was subsequently performed using the same pair of *GmIPK2* specific primers and thermal cycling conditions which were described for its cloning. We further monitored the quantitative amplification of *GmIPK2* by performing qRT-PCR analysis on a PikoReal 96 Real-Time PCR platform (ThermoFisher Scientific, USA). The reactions were set up using DyNAmo Flash SYBR Green qPCR Kit (Thermo Scientific, USA) with cDNA first strands as the template DNA. The expression of *GmIPK2* was normalized to an endogenous control, the housekeeping gene phosphoenolpyruvate carboxylase (*PEPCo*) (Sugimoto et al., 1992; Tuteja et al., 2004). The primers for the gene (q*IPK2*F 5′-CGCGGATCCGCGTTGCAGAAGCTCAAG-3′ and q*IPK2*R 5′-TCCCCGCGGGGAGCGACACTAATTCAAG-3′) and the internal control (q*PEPCo*F 5′-CATGCACCAAAGGGTGTTTT-3′ and q*PEPCo*R 5′-TTTTGCGGCAGCTATCTCTC-3′) were designed using PrimerQuest tool by IDT, USA. The reactions were setup following the standard protocol provided in DyNAmo ColorFlash SYBR Green qPCR Kit (Thermo Scientific). The thermal profile used for PCR amplification was: 95°C for 4 min; 40 cycles of 95°C for 15 s, 60°C for 30 s and fluorescence data collection. To minimize variation in the output, three technical replicates were carried out for each of the three biological replicates. The baseline data was collected for first 15 cycles to generate a baseline-subtracted plot of the logarithmic increase in fluorescence signal (ΔRn) versus cycle number. A standard fluorescence threshold (Rn) was set to 0.5 on the log fluorescence scale to determine the fractional cycle number (Ct value). The relative abundance of *GmIPK2* was calculated using the 2-ΔΔCT method (Livak and Schmittgen 2001). Dissociation curve analysis from 60-95°C was also performed at the end of the assay to check for any non-specific amplification and/or contamination.

2.4. Sequence analysis and phylogenetic tree construction

Homologous *IPK2* sequences from other plants were identified with National Center for Biotech-

nology Information (NCBI) Basic Local Alignment Search Tool (Protein BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi) using the above deduced *GmIPK2* sequence as a query. The primary sequence composition of IPK2 sequences were computed using PEPSTATS (Rice et al., 2000). The physio-chemical parameters of proteins were predicted by using the ProtParam tool of ExPASy web server (Gasteiger et al., 2005). TargetP 1.1 Server using the cutoffs of 95% specificity was implemented for subcellular location prediction of GmIPK2 protein (Emanuelsson et al., 2000) and the outcome was compared to predictions obtained from MemType-2L (Chou and Shen, 2007), WoLF PSORT (Horton et al., 2007), SubLoc v1.0 (Hua and Sun, 2001) and CELLO v2.5 (Yu et al., 2006). Its transmembrane topology was predicted with PSIpred (Buchan et al., 2013), TMpred (Hofmann K and Stoffel, 1993) and NPS@ web programs (Rost et al., 1996). A final consensus was drawn manually and the topology was generated and visualized using Protter version 1.0 (Omasits et al., 2014). Presence of potential secretory signal peptides or mitochondrial targeting peptides was analyzed with SignalP 4.1 web server (Petersen et al., 2011). M-Coffee multiple sequence alignment (MSA) of the selected amino acid sequences was carried out to produce quality alignments which served as the basis for phylogenetic analysis (Wallace et al., 2006) to detect its evolutionary placement and phylogenetic similarity with other similar genes. The evolutionary tree was constructed using Neighbor Joining (NJ) clustering method to compute distances and poisson model for amino acid substitution in MEGA Version 6.0 (Molecular Evolutionary Genetic Analysis, Tamura et al., 2013). Bootstrap replications were set at 1000 to assess the degree of confidence for each clade of the observed tree. The final image was rendered with the Interactive Tree of Life server (iTOL) (Letunic and Bork, 2016).

2.5. Promoter isolation and prediction of regulatory motifs

A motif search was carried out to define putative cis-elements in the promoter sequences involved in the regulation of *IPK2* expression using PlantCARE (Lescot et al., 2002). To identify these cis-regulatory elements, around 2 kb upstream sequence of the *IPK2* homologs were retrieved by

NCBI's nucleotide BLAST program and fed to PlantCARE web tool. The tool's database identified regulatory elements in the isolated upstream sequences and each of the elements observed were analyzed with previously reported properties of the particular element. Unlike CDS, the regulatory sequences indirectly influence their immediate phenotype.

2.6. Secondary structure analysis and domain prediction

M-Coffee multiple sequence alignment (MSA) of the selected amino acid sequences was carried out and the alignment file was imported to Jalview (Waterhouse et al., 2009) to identify and shade the conserved amino acid sequences. Ungapped motifs were also detected using MEME web tool available on MEME suite 4.11.1 (Bailey et al., 2009). The motifs present were further verified using My Hits motif scan tool (Pagni et al., 2007). The domain composition was analyzed using CDD tool (www.ncbi.nlm.nih.gov/cdd) on NCBI server. A secondary structure consensus from amino acid sequences was built based on the joint prediction with SOPMA (nearest-neighbor method) and PHD (neural networks method) correctly predicting 82.2% of the residues for 74% of co-predicted amino acids (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html) (Geourjon and Deléage, 1995; Rost and Sander, 1993). Cysteine species and disulfide connectivity of protein sequences were determined using web tool DiANNA (Ferrè F and Clote, 2005). A secondary structure topology map of the 3D model was generated using ProMotif (http://www.ebi.ac.uk/thornt onsrv/databases/pdbsum/ Generate.html) and rendered using TopDraw (Bond, 2003).

2.7. Homology modeling and quality assessment of predicted model

The 3D model of GmIPK2 was constructed by template-based homology modeling using automated comparative protein modeling servers, SWISS-MODEL (Biasini et al., 2014) and PHYRE2 (Kelly et al., 2015) as well as a standalone comparative modeling program MODELLER 9.16 (Webb and Sali, 2014). Comparative modeling consists of five main steps: search for related protein structures, selection of one or more appropriate templates, target-template alignment, model building and model evaluation. The final step of comparative modeling was additionally performed with RAMPAGE

(Lovell et al., 2003), VERIFY 3D (Eisenberg et al., 1997) and ProSA servers (Wiederstein and Sippl, 2007) to evaluate the stereochemical and energetic properties of the obtained models. In addition to this, the quality of models can also be assessed by structural comparison to the template using the MatchMaker tool in UCSF Chimera (Pettersen et al., 2004), a molecular visualization software package and calculating the Cα root mean square deviation scores (RMSDs) for each of the comparative models. The secondary structures of the final and template proteins were also compared by pairwise 3D alignment using MATRAS 1.2 (Kawabata, 2003). Agreement on the best model was made on the basis of the majority in best scores from these different quality analyses.

2.8. Refinement of the predicted homology model

Molecular dynamics (MD) simulation was performed to optimize the obtained GmIPK2 model. Simulation of the model was conducted in explicit solvent using the GROMACS (Groningen Machine for Chemical Simulations) 4.5.5 package (Pronk et al., 2013). The model was solvated with simple point charge (SPC216) water in a cubic box with edges that were 0.7 nm from the molecular boundary. Initially, energy minimization (maximum number of steps: 1000) was performed to remove steric conflicts between the protein and water molecules, using the steepest descent integrator. The system was then equilibrated by optimizing the solvent molecules surrounding the energy-minimized model with NVT (constant Number of particles, Volume, and Temperature) and NPT (constant Number of particles, Pressure, and Temperature) ensembles using Berendsen thermostat and Parrinello-Rahman barostat respectively. Finally, the system was simulated for 50 ns thrice maintaining the same temperature (300 K) and pressure (1 bar) using the Particle Mesh Ewald (PME) electrostatics method. After completion of production run, we collected the data using a post-processing tool, trjconv (strips out coordinates and correct for periodicity) and used this corrected trajectory as the input for studying the conformational stability of our simulated protein via *g_rms* and *g_rmsf* tools. We subsequently generated and analyzed the output plots using a simple plotting program, xmgrace (Turner, 2005).

2.9. Active site prediction, molecular docking and MD simulation of the docked complexes

The refined protein model thus generated was docked to characterize the 3D structure of the complex, to gain an insight into the crucial amino acid residues also referred to as active residues that are involved in the complex formation. Based on the available literature for IPK2, 1D-*myo*-inositol-1,4,5-trisphosphate (PDB: 5GUG-I3P) and 1D-*myo*-inositol-1,4,5,6-tetra*kis*phosphate (PDB: 4A69-I0P) ligands were selected for optimized GmIPK2 protein receptor. Docking of these ligands was performed using AutoDock Vina (version 1.1.2) (Trott and Olson, 2010) following the semi-flexible approach of docking. AutoDock Vina reads all the molecules in a simplified PDB file representation, termed PDBQT and thus the coordinates of both GmIPK2 protein and its ligands (downloaded from the RCSB protein data bank) were prepared by using MGL tools (version 1.5.4) AutoDock tools (Morris et al., 2009) prior to docking. All the water and solvent atoms of the protein were removed and the polar hydrogen atoms were added. The GmIPK2 molecule was kept rigid while the ligands were allowed to rotate and explore more flexible binding pockets. Vina used a customized rectangular 3D cartesian grid for specifying the binding site of the protein and for efficient geometric scoring. The dimensions of the grid were customized to make sure that the size of the search space is large enough for the ligand to rotate in. Once set, the docking run produced ligand poses each with a definite binding energy (kcal/mol) calculated based on the scoring function used in Vina. The conformations with the lowest binding affinity were chosen and the interaction diagrams were generated using Discovery Studio Visualizer 4.1 (Accelrys Software Inc., USA, 2013). The amino acid residues present at a distance of 2Å were considered as the binding partners of the ligands. The active amino acid residues were also predicted by combining results of three different interface prediction web servers, CASTp (Dundas et al., 2006), FTSite (Kozakov et al., 2015) and FunFOLD2 (Roche et al., 2013) into a consensus. The final complexes were equilibrated by MD simulation by following the procedure described for protein model in explicit water under the section 3.2.7 but by applying restrain on the ligands to prevent them from moving away from the binding site. Once

equilibrated, we simulated the complexes for 50 ns and analyzed the resulting trajectories for their stability using *g_rms, g_rmsf, g_hbond* and *g_mmpbsa* tools.

3. Results and Discussion

3.1. Cloning and sequencing of GmIPK2

We first performed a BLASTN analysis in the plant comparative genomics portal Phytozome v9.1 (Goodstein et al., 2012) with the soybean *IPK2* gene sequence available in NCBI (GenBank: NM_001250522) to retrieve a 1241 bp transcript sequence. Based on the sequence data thus derived, we designed primers specific to amplify complete *GmIPK2* coding sequence (CDS) as well as a fragment of its 3' untranslated region (UTR) with the aim to design silencing construct in the future specific to this highly conserved region because of its substantial role in gene regulation. To accomplish this, we converted total RNA isolated from 8-10 mm developing seed stage to cDNA and performed PCR amplification from the synthesized cDNA template following the protocol described under section 2.2. The amplicon was then cloned into pGEM-T easy vector system and introduced into the bacterial host *E. coli* (DH5α). The putative cDNA clones were verified by restriction analysis with EcoRI and sequence characterized to ~910 bp residues in length (Fig 1). It contains a single open reading frame ~840 bp long which potentially encode a single polypeptide of 279 amino acid residues and a ~70 bp 3′ UTR fragment. We submitted the obtained CDS data to NCBI (GenBank: KF297702) and used the information as query to conduct protein homology search using PSI-BLAST algorithm. Amongst the sequences producing significant alignment, we shortlisted 30 plant IPK2 sequences based on percentage sequence identity to carry out further *in silico* analysis.

3.2. Spatial and temporal expression profiling

As discussed above, tissue specific modulation of *IPK2* gene in order to generate a *lpa* mutant is essential to evade any possible pleiotropic effects. Before this can be achieved it is vital to investigate its spatial expression profile in different tissues as well as its temporal expression profile

in developing seeds, to provide an initial point for strategic achievement of a desired level of silencing.

To study the expression of *GmIPK2* gene in different tissues of a soybean plant and during seed development, semi-quantitative as well as real-time PCR expression analysis was performed with total RNA isolated from root, stem, leaf and flower tissues of 30-day-old *G. max* plants and developing seeds ranging from 0 to 16 mm in size distributed in eight different progressive stages. Semi-quantitative PCR analysis revealed a differential pattern of *GmIPK2* transcript expression across the set of experimental tissues analyzed, with the highest level of transcripts observed in seeds (Fig 2(a)). The same was confirmed by steady-state qRT-PCR analysis which also detected highest level of transcripts in seeds (Fig 2(b)). This suggests that *GmIPK2* play a key role in PA biosynthesis in this tissue for use as a primary source of energy during germination. In both the analyses, same level of amplification was observed for *PEPCo* housekeeping gene. Since PA is also required for several other vital functions throughout the plant system, a basal level of it is observed in all the tissues as well. Thus, amongst the other tissues analyzed, expression of *GmIPK2* was also recorded in roots and flowers but at a level lower than that observed in the cotyledons. The strong presence of *GmIPK2* transcripts in these tissues may be attributed to its role in regulating cytosolic calcium gradient which correlates with pollen germination, pollen tube growth (Franklin-Tong et al., 1996; Malho, 1998; Pierson et al., 1994; Xu et al., 2005), root growth and root hair development (Bibikova et al., 1997; Felle and Helper, 1997; Wymer et al., 1997). A low level of expression was also observed in stems and leaves as it is studied to be involved in regulating a vital function of axillary shoot branching by participating in auxin signaling (Zhang et al., 2007).

We then analyzed its temporal expression pattern during eight progressive seed development stages. In both semi-quantitative as well as real-time PCR expression analysis, we observed that its expression increased as the development progressed and reached a peak value at the later stages of seed development (Fig 3). This pattern of expression coincides evenly with the pattern of

accumulation of PA which is linear throughout most of seed development (Raboy and Dickinson, 1987). Similar results were obtained in a microarray transcriptome study conducted in the past in our laboratory (GEO: GSE69821). This observation can be explained by the increase in production of phosphorous compounds required to support growth and development during initial stages of seed development when the synthesis of phosphorous reserve, PA is minimal (Raboy and Dickinson, 1987).

The spatiotemporal analysis in summary identifies seed as the major tissue for its expression with maximum relative expression occurring during the later stages of its development. Thus, from the present study we can hypothesize that targeting the *GmIPK2* gene expression during late seed development stages may provide a potential strategy for generating *lpa* soybean with enhanced nutritional value.

3.3 Characterization of regulatory motifs in IPK2 gene promoter

In order to understand the mechanism regulating the spatiotemporal expression pattern of *GmIPK2* gene thus observed, we analyzed the promoter region of all its homologs and identified different cis-regulatory elements located therein. A 2 kb sequence upstream to the open reading frame was identified and subjected to PlantCARE analysis. Database search revealed the presence of many motifs related to seed-specific promoters, hormone-responsive cis-elements (HRE) and cis-elements responsive to stresses (DSRE) that together contribute to the differential regulation of our gene (Table 1). Light responsive elements (Arguello-Astorga GR and Herrera-Estrella, 1996; Feldbrugge et al., 1996; Lois et al., 1989; Lopez-Ochoa et al., 2007; Sessa et al., 1995) were observed most frequently which suggest a probable diurnal regulation of *IPK2* expression. The GCN4 and Skn_1 motifs highly conserved in the promoters of cereal seed storage protein genes were located within the *IPK2* promoter. These cis-acting elements play a central role in controlling endosperm-specific gene expression (Onodera et al., 2001; Takaiwa et al., 1996; Washida et al., 1999; Wu et al., 1998). Multiple HREs particularly those known to be involved in abscisic acid (ABA) and gibberellic acid

(GA3) sensing were also identified. A ABA-responsive element (ABRE) with the core sequence PyACGTG/TC and three GA-responsive elements (GARE) AAACAGA, TCTGTTG and TATCCAC/T (Niu et al., 2016; Mongkolsiriwatana et al., 2009; Yamaguchi-Shinozaki et al., 1989) were identified which previously were reported to correlate with PA accumulation during grain filling (Abe et al., 2003; Matsuno et al., 2014). Aggarwal et al., 2015 reported that *IPK2* is an ABA-induced gene which is antagonistically suppressed by GA3 underlining the crucial role played by these hormones in regulating PA pathway genes. Putative elements responsive to methyl-jasmonate, drought inducibility and anaerobic induction were also observed which help combat assorted types of abiotic factors that plants are exposed to under natural environment (Abe et al., 2003; Kim et al., 1993; Nguyen et al, 2003; Rouster et al., 1997; Shinozaki and Yamaguchi-Shinozaki, 2000).

3.4 Computation of physiochemical parameters and subcellular localization prediction

GmIPK2 sequence similarity search using PSI-BLAST revealed homology to IPK2 protein sequences from different plant sources showing maximum similarity with *Glycine soja* (98%) and *Vigna radiata* (74%) (Table 2). Primary sequence analysis of these homologs indicate that leucine is the most abundant amino acid which makes upto approximately 9 to 11 mole percent of its backbone residues whilst the percentage of tryptophan and methionine were found to be the least at approx. 1% across species. The isoelectric points (pI) of all IPK2 proteins were computed to be under 7 suggesting that they are most likely to precipitate in acidic buffers except *Glycine soja, Medicago truncatula, Zea mays and Sorghum bicolor* which has a pI of 7.1, 8.57, 10.48 and 8.26 respectively indicating their solubility in basic buffers. The calculated pI will be useful for empirical protein purification by isoelectric focusing and ion exchange chromatography. The extinction coefficient (EC) (Gill and Von Hippel, 1989) of IPK2 proteins measured at 280 nm in water was found ranging from 25440 to 34380 $M^{-1}cm^{-1}$ with respect to their concentration of aromatic amino acids (11-15%) and cystine (disulfide bonds). These EC values can be used to calculate protein concentration in a solution which in turn help in the quantitative study of biochemical interactions (protein-protein and

protein-ligand). The instability indices (Ii) computed for selected IPK2 proteins are used to determine their *in vivo* half-lives (Guruprasad et al., 1990). Rogers et al., 1986 reported that proteins having Ii values greater than 40 have an in vivo half-life of less than 5 hours while those proteins having Ii values less than 40 have a longer in vivo half-life of 16 hours. Our study showed that Ii values of all the homologs are less than 40 and hence are thermally stable with a long half-life except for *Lotus japonicas* and *Zea mays* kinases that have a Ii above 40 which indicate their possible thermal instability. Thermostability of proteins result from a combination of several factors acting synergistically. Here we assess thermal stability of our proteins in direct proportionality with their aliphatic index (Ai) which is a measure of the relative volume occupied by aliphatic side chains (Ikai, 1980). The Ai values determined for IPK2 kinases ranged from 68.79 to 96.70 with those from brassicaceae family showing lowest thermal stability. This in turn is indicative of their greater flexibility at a wide range of temperatures when compared to proteins of other families. Grand average of hydropathicity (GRAVY) number reflect the average hydropathy of a protein, the positively rated being hydrophobic and negatively rated being hydrophilic in nature Kyte and Doolittle, 1982). GRAVY index for IPK2 kinases was found ranging from -0.517 to -0.013 indicating that these proteins will interact favourably with water except for *Vigna radiata* with an index of 0.045 and hence is a potential hydrophobic protein.

Protein localization and target peptide predictions are significant studies as they aid in *in silico* protein function characterization as well as genome annotation. The acidic amino acid composition of IPK2 homologs determined by the physiochemical analysis conducted above suggest that they are cytoplasmic in nature as opposed to the basic amino acid composition of membrane proteins for their stability (Schwartz et al. 2001). Further sequence analysis based on TargetP scores (cTP: 0.163, mTP: 0.066, SP: 0.087, other: 0.906) also suggest that IPK2 kinases may be located anywhere in the cell besides chloroplast and mitochondria. The sequences were not predicted to have any signaling pre-sequence which was indicated by their low signal peptide (SP) score reinforced that they are

soluble in nature. Besides, a consensus of predictions obtained from WoLF PSORT, CELLO v2.5, SubLoc v1.0 and MemType-2L servers (Table S1) also established the cytoplasmic character of IPK2 protein. However, TMpred, PSIPRED, NPS@ and DAS servers also identified a single consensus C-terminal transmembrane region positioned at 259-274. This could have been mistaken due to the presence of large regions of hydrophobic residues in the C-terminal soluble region, as hydrophobicity alone is used as the criterion to predict membrane-spanning regions. The absence of any transmembrane helix is also well documented in the hydrophobicity plot generated using waveTM server (Fig S1) (Pashou et al., 2004) as majority of amino acids show negative for hydrophobicity.

3.5 Motif analysis and secondary structure characterization

Multiple sequence alignment (MSA) of the selected plant IPK2 homologs was performed using M-Coffee web server which is a meta-method for assembling MSA by combining the output of several individual methods into one to generate the best possible alignment. The consensus alignment thus generated revealed several significantly conserved motifs and sites unique to inositol phosphate kinases (Fig 4). A signature inositol-phosphate-binding motif, PxxxDxKxG was identified in all the aligned sequences (Odom et al., 2000; Saiardi et al., 1999) which confirms that they belong to inositol phosphate kinase (IPK) superfamily of IP kinases. Holmes and Jogl, 2006 state that the members of this superfamily share several strictly conserved signature motifs with each other and are predicted to assume the same overall fold, despite the low sequence conservation. The core catalytic tyrosine kinase motif, RxxxExxxY was also discovered in all the sequences which suggest that they are tyrosine-specific protein kinases (Cooper et al., 1984). IPK2 sequences from Solanaceae and Rosaceae families were found to contain a Glycine-rich consensus ATP-binding GxGxxG motif characteristic of protein kinase C (PKC) catalytic domain (Steinberg, 2008). The classical PKC and plant CDPKs recognized phosphorylation S/TxK/R motif (Nishikawa et al., 1997; Neumann et al., 1996; Roberts and Harmon, 1992) was also identified in some of the sequences speculating their role

in lipid-dependent PA biosynthetic pathway. Such promiscuous kinase activity suggests that both lipid-dependent and independent pathways regulate PA biosynthesis as well as basic nuclear and cellular processes in plants (Josefsen et al., 2007; Stevenson-Paulik et al., 2002). A protein recognition LxxLL motif common to all of the aligned sequences indicate their participation in protein-protein interactions, regulating cell signalling, cell adhesion and transcription (Plevin et al., 2005). Further analysis by MEME suite web server identified a total of eleven conserved ungapped motifs, with motif PxxxDxKxG being the most conserved amongst all IPK2 homologs as indicated by its lowest E-value of 2.9e-487 (Fig S2). The obtained motifs were subjected to BLASTP analysis for conformation of their annotations which established that they all belong to IPK superfamily domain (CDD Acc: cl12283) and thus substantiate our previous results. We further explored GmIPK2 protein sequence by Motif Scan which recognized diverse protein kinase phosphorylation sites (Table S2). Since phosphorylation acts as a molecular switch in modulating protein function, structural rearrangement and cellular localization it can be suggested that GmIPK2 play a critical role in many biological regulatory events of signalling, proliferation, differentiation and apoptosis. Its role in biological processes as diverse as mRNA export (York et al. 1999), DNA repair (Hanakahi et al. 2000), regulation of chromatin structure (Shen et al. 2003; Steger et al. 2003), maintenance of basal resistance to plant pathogens (Murphy et al. 2008) and apoptosis (Agarwal et al. 2009) have been studied in the past.

The secondary structure of a protein is more conserved than its nucleotide sequence and is therefore a prized source of information in understanding its classification, function, molecular evolution and interaction with macromolecules (Reehana et al. 2013). In addition, secondary structure provides the first framework for homology based prediction of a protein 3D-model. Thus, in the current study, we inferred the secondary structure composition of IPK2 kinases from a three-state prediction done by using NPS@ web server (Table 3). A high coil content was observed in most of the sequences including GmIPK2 while some showed them in nearly equal proportion with α-helix. This structural

state can be justified based on the rich content of highly flexible glycine and kink inducing proline amino acid residues. The percentage of extended strands (% Ee) in all the kinases were found ranging from 13-29% except for the Poaceae family kinases which showed a low % Ee conformation (below 10%). PDBSum tool PROMOTIF analysis of GmIPK2 polypeptide identified total eight α-helices and eight β-strands arranged to form three antiparallel β-sheets, interspersed throughout by regions of coil or turn conformations (Fig 5). We also recognized a varying number of bonded half-cystine pairs in all the IPK2 protein sequences using DIANNA server. It revealed the presence of 8 Cys residues in GmIPK2 and the most probable half-cystine pairs predicted by CYS-REC were 94-121, 123-276 and 158-186. These potential long-term interactions participate in stabilizing the native conformations of our proteins and may as well contribute to differences in their tertiary structures.

3.6 Evolutionary analysis

Conserved motif analysis of IPK2 protein sequences point at a distinct evolutionary association between these kinases. Phylogenetic analysis would provide a further basis to determine their relatedness as well as to understand their collective evolution from a common ancestor. Previous research has reported that IPK superfamily of kinases to which *IPK2* belongs evolved from a common ancestor (Irvine and Schell, 2001; Shears, 2004). In our study, based on the alignment obtained in section 3.5 of GmIPK2 protein and its homologs, we constructed a neighbor-joining phylogenetic tree using MEGA 6.0 software (Fig 6). The tree topology derived was supported by high bootstrap values. The IPK2s were clustered into six well delineated groups. The clusters consist of members of the Poaceae, Brassicaceae, Malvaceae, Rosaceae, Solanaceae and Fabaceae families. The Poaceae family of monocots (*Zea may, Sorghum bicolor and Aegilopus tauschii*) was found to be most distantly related to *G. max* in comparison to the fabaceae family of eudicots (*Glycine soja, Phaseolus vulgaris, Vigna radiata, Cajanus cajan, Trifolium pratense, Cicer arietinum, Medicago truncatula, Lotus japonicas, Arachis duranensis* and *Arachis hypogaea*) which are most closely related.

3.7 Three-Dimensional model construction

***3.7.1 Model building*:** To derive structural information about GmIPK2 protein we built its theoretical model by following homology modeling approach since no X-ray crystal or NMR structure of it is available. The homology modeling technique takes advantage of structural conservation found in similar proteins that have evolved from a common ancestor. Yeast IPK2 protein was the first member of inositol multikinase family whose crystal structure was determined (Holmes and Jogl, 2006) but it was found to show a very low sequence similarity with GmIPK2 protein. As we know, aligning two sequences can be a difficult process if the sequence similarity is low, we therefore used the above derived GmIPK2 protein sequence as a query in PSI-BLAST to find more similar sequences amongst the PDB database proteins with resolved 3D structures to use as a potential template. The only closest homologous sequence available in PDB was that of chain A of *Arabidopsis thaliana* inositol phosphate multikinase (PDB:4FRF) which showed 55% sequence identity with an e-value of 9e-99. The initial comparative models i.e. GmIPK2-S and GmIPK2-P were built using fully automated SWISS-MODEL and PHYRE2 servers respectively which also identified 4FRF_A as the most reliable template using sensitive hidden markov model searches and used the same as the structural input. A global quality estimation score (GMQE) of 0.62 was provided by SWISS-MODEL which indicates a reasonably reliable structure. Homology model was additionally built using MODELLER 9.16 program from the X-ray crystal structure coordinates of the previously identified template structure (4FRF_A). The software generated five different models by optimizing the objective function of spatial restrains in a cartesian space. Three different energy scores viz. molpdf, DOPE and GA341 were computed for each of these models and compared to one another to select the best 3D structure (Table S3). Model 4 (GmIPK2-M4), with the lowest molpdf and DOPE scores of 1734.48291 and -27846.58203 respectively, was chosen as the principal conformational structure.

***3.7.2 Structure validation*:** We then assessed the accuracy and reliability of all the three predicted models viz. GmIPK2-S, GmIPK2-P and GmIPK2-M4 using various online diagnostic tools (Table

4). RAMPAGE server which evaluates the 3D-structures based on Ramachandran plot calculations showed variable distribution of torsion angles in all the models. Fig S3(a) shows Ramachandran plot for GmIPK2-S model, showing 92.5% residues in favourable region and 4.9% in allowed region with just 2.7% residues in outlier region of the plot, which reflects its superior backbone geometry. We then utilized ProSA-web computational engine to analyze overall quality of the models based on their z-scores and local quality based on their residue energies. GmIPK2-S model showed a better z-score of -6.56 which is displayed in the energy distribution plot derived from a group of experimentally determined protein structures of similar size. The Z-score thus observed was very much within the range of scores typically found for native conformations of this group which indicates a good overall quality of the modelled protein (Fig S3(b)). Moreover, its residue energy was computed to be largely negative which further reflects that the local regions in the protein are modelled well (Fig S3(c)). The accuracy of 3D-models was further analyzed from energy profiles obtained by Verify3D program. Figure S3(d) shows 3D-1D profile for GmIPK2-S model with 97.82% of the model residues showing an average 3D-1D profile score >= 0.2 and hence validate that majority of its amino acid sequence reconcile to its environment in the 3D structure. Additionally, the pairwise 3D structural alignment of GmIPK2-S model with template protein, 4FRF_A using MATRAS 2.1 program revealed that both the structures shared 91.3% secondary structure identity (Fig S4(a)) and the average distance between the Cα backbone atoms of their 3D structures i.e. root mean square deviation (RMSD) measured through superimposition was 0.44 Å (between 196 atom pairs) (Fig S4(b)). Based on the majority of winning scores, GmIPK2-S (Fig 7) was chosen as the best comparative model for energy minimization and further analyses.

***3.7.3 Molecular dynamics simulation*:** We subsequently performed MD simulation on our predicted GmIPK2-S model using GROMOS96 53A6 force field to compute its stability and dynamics. Initial potential energy minimization of solvated model showed that the maximum force dropped below the defined value of 1000 $kJmol^{-1}$ nm^{-1}in nearly 500 steps. The protein structure was then subjected to

50 ns of equilibration run at a constant temperature and pressure to obtain its molecular trajectory. The trajectory thus obtained was used to determine RMSD of Cα backbone atoms of the model using its starting structure as the reference, to determine its convergence towards an equilibrium state. Fig 8(a) shows RMSD as a function of simulation time. We observed that the protein stabilized around 10 ns of production run and converged to ~0.55 nm at 50 ns. The initial increase of RMSD could be attributed to the restraints in the system applied in the equilibration phase and their release later at the beginning of production phase. Besides RMSD, we also calculated root mean square fluctuations (RMSFs) to study mobility of the protein structure to draw an idea of its flexibility regions. From the RMSF plot (Fig 8(b)), we identified Leu (75, 135, 272), Asp (76, 79, 127), Ala (77), Ser (78, 130-132), Gly (80, 136), His (84), Lys (126, 154), Glu (129, 230), Arg (128, 153) residues to be more flexible. This implies that the mentioned residues show greater movement from their native position i.e. are dynamic in nature and thus are functionally more relevant. Thus, overall, the simulation results highlight the stable and reliable nature of our protein model and find it fit to be used for further active site predictions.

3.8 Molecular docking with inositol phosphates

***3.8.1 Active site predictions*:** Once the protein model was refined, a comparative study was performed to detect possible binding pocket residues using CASTp, FTSite and FunFOLD2 binding site prediction servers (Fig S5). Lys105, Thr110, Lys122, Lys126, Ser130, Lys138, Ile139, Pro140, Arg153, Lys154, Gln157 and Ser219 were determined as possible active site residues in GmIPK2-S model. Similar predictions were also made in template protein (4FRF_A) which identified Arg104, Thr105, Pro108, Phe137, Lys149, Arg152, His216, Asn218, Ser219, Gln242 and Val246 as the probable binding site residues. From these studies, we could decipher that residues Lys, Arg, Pro, Thr, Gln and Ser are highly conserved in active sites of functionally identical model and template proteins.

***3.8.2 Docking and residue interaction analysis*:** Molecular recognitions are vital to many biological

processes. However, experimental determination of structures of molecular interactions is cost intensive, demand time and expertise. We therefore, chose computational molecular docking to model our protein-ligand binding and characterize the interactions between its binding pocket residues and known active ligands. Multiple substrate specificities have been described previously for IPK2 gene product, catalyzing primarily 5GUG-I3P and 4A69-I0P (Stevenson-Paulik et al., 2002). We thus docked these centroid ligands into the binding cavity of GmIPK2 protein using molecular docking program Autodock Vina based on a semi-flexible docking approach with the scaling factor defined within 0.1 nm to predict their bound geometry. VINA uses its iterated local search global optimizer algorithm to produce 9 different poses of which pose 1 corresponding to each ligand was identified as the best binding mode based on their lowest binding affinity score of -6.2 kcal/mol for 5GUG-I3P and -5.8 kcal/mol for 4A69-I0P computed by VINA's default statistical scoring function. We further inspected the molecular interactions between these protein substrate poses using Discovery Studio to predict functionally important amino acid residues and found that both the ligands were stabilized in their active site area by strong hydrogen bonding interactions (Fig 9(a) and 9(b)). Lys105 and Lys126 were identified as H-donors to the phosphate group oxygen of both 5GUG-I3P and 4A69-I0P while Arg153 a H-acceptor for hydrogen bonds formation suggesting that these binding pocket residues may play a pivotal role in enzymes function and protein structure stability. Based on the previous work conducted by Holmes and Jogl, we hypothesize that side chains of these amino acid residues may assist in inducing conformational changes on inositol phosphate binding, enabling the enzyme to interact with differently phosphorylated inositol polyphosphates in different orientations, thus endorsing its substrate versatility. Besides Ser217 was also found to form stabilizing hydrogen bond with I0P ligand. Gln157 and Pro140 were observed to form an unconventional carbon-oxygen hydrogen bond with 5GUG-I3P and 4A69-I0P respectively indicating their possible contribution to ligand binding affinity and ligand recognition (Klaholz and Moras, 2002). Lengths and angles of hydrogen bonds stabilizing the GmIPK2-I3P and GmIPK2-I0P

complexes are enlisted in Tables 5(a) and 5(b) respectively. Moreover, our analysis indicates that Thr110, Lys138 and Lys154 show non-bonding interactions with the ligands. 2D protein GmIPK2-I3P/I0P ligand interaction diagrams (Fig 9(c) and 9(d)) were also generated using Discovery Studio.

***3.8.3 Molecular dynamics simulation of GmIPK2-I3P and GmIPK2-I0P complexes*:** We subsequently subjected our docked protein-ligand complexes to MD simulation using GROMOS96 43A1 force field to understand their stability and dynamics. The trajectories obtained were utilized to construct their respective RMSDs, RMSFs and H-bond interactions. The GmIPK2-I3P and GmIPK2-I0P complexes exhibited a deviation between ~0.45 to 0.65 nm (Fig S6(a)) and ~0.45 to 0.58 nm (Fig S7(a)) respectively that converged to ~0.6 nm and ~0.55 nm respectively at 50 ns. This suggests that the structures were stabilized following simulation. From their RMSF plots, we observed fluctuations up to ~0.38 nm in the GmIPK2-I3P complex (Fig S6(b)) and ~0.29 nm in the GmIPK2-I0P complex (Fig S7(b)) which reveal the characteristic regional flexibilities of functional significance in each complex. We also analyzed the hydrogen bonds which participate in the maintenance of these complexes. Three main hydrogen bonds help to stabilize the ligands, 5GUG-I3P and 4A69-I0P, within the enzyme's active site, two of them acting as acceptors and only one as a donor. In particular, the hydrogen bonds formed with Lys105 and Lys126 and the oxygen atom of each ligand's phosphate group, were constantly held. Fig. 10 shows the time evolution of these main H-bonds. Other less significant hydrogen bonds were formed sporadically between the ligands and the protein molecule. Lastly, we calculated the binding free energies ($\Delta G_{binding}$) of both the complexes by using *g_mmpbsa* for their respective trajectories. $\Delta G_{binding}$ values of -12.4 kcal/mol and -11.9 kcal/mol were estimated for GmIPK2-I3P and GmIPK2-I0P complexes respectively. This indicates that the binding of ligand molecules is thermodynamically favourable and thus validates the reliability of our simulation.

4. Conclusion

The present work describes expression analysis and molecular characterization of *IPK2*

multifunctional kinase involved in the late phase of PA biosynthesis. To initiate the study, we identified and cloned the partial gene sequence of *IPK2* from *G. max* cv. Pusa-16. *IPK2* transcripts, when assessed during seed development showed a predominant expression in the later stages of its development which consequently suggests that perturbing *IPK2* gene at this stage can be a viable strategy for manipulating PA levels in soybean seeds to achieve a *lpa* trait. Computational analysis of the gene further highlighted its molecular features including cis-acting promoter elements potentially regulating its observed expression; primary and secondary structural features shedding light on its physicochemical features, conserved functional motifs and major structural elements; and its evolutionary relationship, which can be used to design several experimental studies. The 3D model of IPK2 protein developed can assist in the experimental determination of its 3D model as well as can be considered as a working model for generating hypothesis to make more accurate predictions of protein function and catalytic mechanism in the future. The docking studies performed subsequently can be used for structure-based designing of potent IPK2 inhibitors, useful in studying IPK2 active site and substrate selectivity and for studying PA biosynthesis pathway in detail. In conclusion, the obtained results provide very important preliminary data needed to manipulate PA content in soybean seeds as well as other crops for improving their nutritional quality by biotechnological intervention in the future.

5. Acknowledgements

Financial support for the work was provided by funding from the National Funds for Basic, Strategic and Frontier Application Research in Agriculture [Grant No. NFBSFARA/RNAi -2011/2011-12], ICAR, Government of India. The authors would also like to thankfully acknowledge the Supercomputing Facility for Bioinformatics and Computational Biology at IIT Delhi for the use of its facilities.

6. Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

7. References

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signalling. Plant Cell 15:63-78.

Accelrys Software Inc. Discovery Studio Modeling Environment, Release 4.0. San Diego, Calif, USA: Accelrys Software Inc. (2013).

Agarwal R, Mumtaz H, Ali N (2009) Role of inositol polyphosphates in programmed cell death. Mol Cell Biochem, 328:155-165.

Aggarwal S, Shukla V, Bhati KK, Kaur M, Sharma S, Singh A, Mantri S, Pandey AK (2015) Hormonal Regulation and Expression Profiles of Wheat Genes Involved during Phytic Acid Biosynthesis Pathway. Plants (Basel), 4:298-319.

Ali N, Paul S, Gayen D, Sarkar SN, Datta SK, Datta K (2013) Development of low phytate rice by RNAi mediated seed-specific silencing of inositol 1,3,4,5,6-pentakisphosphate 2-kinase gene (IPK1). PLoS ONE 8:e68161.

Ali N, Paul S, Gayen D, Sarkar SN, Datta SK, Datta K (2013) RNAi mediated down regulation of myo-inositol-3-phosphate synthase to generate low phytate rice. PLoS ONE 6:12.

Arguello-Astorga GR, Herrera-Estrella LR (1996) Ancestral multipartite units in lightresponsive plant promoters have structural features correlating with specific phototransduction pathways. Plant Physiol 112:1151-1166.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202-W208.

Bhati KK, Aggarwal S, Sharma S, Mantri S, Singh S, Bhalla S, Kaur J, Tiwari S, Roy J, Tuli R, Pandey AK (2014) Differential expression of structural genes for the late phase of phytic acid biosynthesis in developing seeds of wheat (Triticum aestivum L.), Plant Sci 224:74-85.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modeling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42:W252-W258.

Bibikova TN, Zhigiler A, Gilroy S (1997) Root hair growth in Arabidopsis thaliana is directed by calcium and an endogenous polarity. Planta 203:495-505.

Bilyeu Kristin D, Zeng P, Coello P, Zhang Zhanyuan J, Krishnan Hari B, Bailey A, Beuselinck Paul R, Polacco Joe C (2008) Quantitative conversion of phytate to inorganic phosphorus in soybean seeds expressing a bacterial phytase. Plant Physiol 146:468-77.

Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. Bioinformatics 19:311-2.

Brearley CA, Hanke DE (1996a) Metabolic evidence for the order of addition of individual phosphate esters to the myo-inositol moiety of inositol hexakisphosphate in the duckweed Spirodela polyrhiza L. Biochem J 314:227-233.

Brearley CA, Hanke DE (1996b) Inositol phosphates in barley (Hordeum vulgare L.) aleurone tissue are stereochemical similar to the products of breakdown of Ins P6 in vitro by wheat bran phytase. Biochem J 318:279-286.

Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench, Nucleic Acids Res 41:W340-W348.

Chou KC, Shen HB (2007) MemType-2L: A Web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. Biochem Biophys Res Commun 360:339-345.

Cichy K, Raboy V (2009) Evaluation and development of low-phytate crops. Ed. Hari Krishnan, Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America 177-200.

Cooper JA, Esch FS, Taylor SS, Hunter T (1984) Phosphorylation sites in enolase and lactate

dehydrogenase utilized by tyrosine protein kinases in vivo and in vitro. Biol Chem 259:7835-7841.

Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res 34:W116-W118.

Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396-404.

Emanuelsson O, Nielsen H, Brunak S, Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Mol Biol 300:1005-1016.

Feldbrugge M, Hahlbrock K, Weisshaar B (1996) The transcriptional regulator CPRF1: expression analysis and gene structure. Mol Gen Genet 251:619-627.

Felle HH, Hepler PK (1997) The cytosolic Ca^{2+} concentration gradient of Sinapis alba root hairs as revealed by Ca^{2+} selective microelectrode tests and fura-dextran ratio imaging. Plant Physiol 114:39-45.

Feng X, Yoshida KT (2004) Molecular approaches for producing low-phytic-acid grains in rice. Plant Biotechnol 21:183-189.

Ferrè F, Clote P (2005) DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Res 33:W230-2.

Fileppi M, Galasso I, Tagliabue G, Daminati M, Campion B, Doria E, Sparvoli F (2010) Characterisation of structural genes involved in phytic acid biosynthesis in common bean (Phaseolus vulgaris L.). Mol Breed 25:453-470.

Franklin-Tong VE, Drobak BK, Allan AC, Trewavas AJ (1996) Growth of pollen tubes of Papaver rhoeas is regulated by a slow moving calcium wave propagated by inositol 1,4,5-trisphosphate. Plant Cell 8:1305-1321.

Frederick JP, Mattiske D, Wofford JA, Megosh LC, Drake LY, Chiou ST, Hogan BLM, York JD (2005) An essential role for an inositol polyphosphate multikinase, Ipk2, in mouse embryogenesis

and second messenger production, Proc Natl Acad Sci 102:8454-8459.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server, The Proteomics Protocols Handbook pp. 571-607.

Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11:681-684.

Gill SC, Von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182:319-326.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178-D1186.

Guruprasad K, Reddy BVB, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 4:155-161.

Hanakahi LA, Bartlet-Jones M, Chappell C, PappinD, West SC (2000) Binding of inositol phosphate to DNA-PK and stimulation of double-strand break repair. Cell, 102:721-729.

Hofmann K, Stoffel W (1993) TMbase - A database of membrane spanning proteins segments. Biol Chem 374:166.

Holmes W, Jogl G (2006) Crystal structure of inositol phosphate multikinase 2 and implications for substrate specificity. Biol Chem 281:38109-38116.

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucl. Acids Res 35:W585-587.

Hua S, Sun Z (2001) Support vector machine approach for protein subcellular localization prediction. Bioinformatics 17:721-728.

Ikai AJ (1980) Thermostability and aliphatic index of globular proteins. J Biochem 88:1895-1898.

Irvine RF, Schell MJ (2001) Back in the water: the return of the inositol phosphates. Nat Rev Mol Cell Biol 2:327-338.

Josefsen L, Bohn L, Sorensen M, Rasmussen S (2007) Characterization of a multifunctional inositol phosphate kinase from rice and barley belonging to the ATP-grasp superfamily. Gene 397:114-125.

Kawabata T (2003) MATRAS: a program for protein 3D structure comparison. Nucleic Acids Res 31:3367-9.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10:845-858.

Kim SR, Kim Y, An G (1993) Identification of methyl jasmonate and salicylic acid response elements from the nopaline synthase (nos) promoter. Plant Physiol 103:97-103.

Klaholz B, Moras D (2002) C-H...O hydrogen bonds in the nuclear receptor RARgamma--a potential tool for drug selectivity. Structure 10:1197-1204.

Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, Xia B, Beglov D, Vajda S (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nature Protocols 10:733-755.

Krishnan V, Jain P, Tripathi V, Hada A, Manickavasagam M, Ganapathi A, Rai RD, Sachdev A (2015) Molecular modeling and in-silico characterization of Glycine max inositol (1,3,4) tris 5/6 kinase-1 (Gmitpk1) - a potential candidate gene for developing low phytate transgenics. Plant Omics 8:381-391.

Kuwano M, Mimura T, Takaiwa F, Yoshida KT (2009) Generation of stable 'low phytic acid' transgenic rice through antisense repression of the 1D-myo-inositol 3-phosphate synthase gene using the 18-kDa oleosin promoter. Plant Biotechnol 7:96-105.

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105-132.

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S (2002)

PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30:325-327.

Letunic I, Bork P (2016) Interactive Tree Of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucl Acids Res 44:W242-W245.

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using realtime quantitative PCR and the 2ΔΔC(T) method. Methods 25(4):402-408.

Lois R, Dietrich A, Hahlbrock K, Schulz W (1989) A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements. EMBO J 8:1641-1648.

Lopez-Ochoa L, Acevedo-Hernández G, Martínez-Hernández A, Argüello-Astorga G, Herrera-Estrella L (2007) Structural relationships between diverse cis-acting elements are critical for the functional properties of a rbcS minimal light regulatory unit. J Exp Bot 58:4397-4406.

Lovell SC, Davis IW, Arendall III WB, Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure Validation by C_α Geometry: Φ,Ψ and C_β Deviation. Proteins 50:437-450.

Majerus PW (1992) Inositol phosphate biochemistry. Annu Rev Biochem 61:225-250.

Majerus PW (1996) Inositols do it all. Genes Dev 10:1051-1053.

Malho R (1998) Role of 1,4,5-inositol trisphosphate-induced Ca^{2+} release in pollen tube orientation. Sex Plant Reprod 11:231-235.

Matsuno K, Fujimura T (2014) Induction of phytic acid synthesis by abscisic acid in suspension-cultured cells of rice. Plant Sci 217-218:152-157.

Mongkolsiriwatana C, Pongtongkam P, Peyachoknagul S (2009) In silico promoter analysis of photoperiod-responsive genes identified by DNA microarray in rice (Oryza sativa L.). Kasetsart J (Nat Sci) 43:164-177.

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009)

Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Computational Chem 16:2785-2791.

Murphy AM, Otto B, Brearley CA, Carr JP, Hanke DE (2008) A role for inositol hexakisphosphate in the maintenance of basal resistance to plant pathogens. Plant J., 56:638-652.

Neumann GM, Thomas I, Polya GM (1996) Identification of the site on potato carboxypeptidase inhibitor that is phosphorylated by plant calcium-dependent protein kinase. Plant Sci, 114:45-51.

Nguyen T, Sherratt PJ, Cecil B, Pickett CB (2003) Regulatory mechanisms controlling gene expression mediated by the antioxidant response element. Annu Rev Pharmacol Toxicol 43:233-260.

Nishikawa K, Toker A, Johannes FJ, Songyang Z, Cantley LC (1997) Determination of the specific substrate sequence motifs of protein kinase C isozymes. Biol Chem 272:952-960.

Niu J, Wang J, Hu H, Chen Y, An J, Cai J, Sun R, Sheng A, Liu X, Lin S (2016) Cross-talk between freezing response and signalling for regulatory transcriptions of MIR475b and its targets by miR475b promoter in Populous suaveolens. Sci Rep 6:20648.

Nunes ACS, Vianna GR, Cuneo F, Amaya-Farfan J, de Capdeville G, Rech EL, Aragao FJL (2006) RNAi mediated silencing of the myo-inositol-1-phosphate synthase gene (GmMIPS1) in transgenic soybean inhibited seed development and reduced phytate content. Planta 224:125-132.

Odom AR, Stahlberg A, Wente SR, York JD (2000) A role for nuclear inositol 1,4,5-trisphosphate kinase in transcriptional control. Science 287:2026-2029.

Omasits U, Ahrens CH, Müller S, Wollscheid B (2014) Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics 30:884-886.

Onodera Y, Suzuki A, Wu CY, Washida H, Takaiwa F (2001) A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. J Biol Chem 276:14139-14152.

Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Hau J, Martin O, Kuznetsov D, Falquet L (2007) MyHits: improvements to an interactive resource for analyzing protein sequences.

Nucleic Acids Res 35:W433-W437.

Pashou EE, Litou ZI, Liakopoulos TD, Hamodrakas SJ (2004) WaveTM: wavelet-based transmembrane segment prediction. In Silico Biol 4:127-31.

Petersen TN, Brunak S, Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8:785-786.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) "UCSF Chimera--a Visualization System for Exploratory Research and Analysis." Journal of Computational Chemistry 25 (13): 1605-1612.

Pierson ES, Miller DD, Callaham DA, Shipley MA, Rivers BA, Cresti M, Hepler PK (1994) Pollen tube growth is coupled to the extracellular calcium ion flux and the intracellular calcium gradient: effect of BAPTA-type buffers and hypertonic media. Plant Cell 6:1815-1828.

Plevin MJ, Mills MM, Ikura M (2005) The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. Trends Biochem Sci 30:66-69.

Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845-854.

Raboy V, Dickinson DB (1987) The timing and rate of phytic acid accumulation in developing soybean seeds. Plant Physiol 85:841-844.

Reehana N, Ahamed AP, Ali DM, Suresh A, Kumar RA, Thajuddin N (2013) Structure based computational analysis and molecular phylogeny of C-Phycocyanin gene from the selected cynobacteria. International J Biological, Veterinary, Agric. Food Eng. 7:47-51.

Rice P, Longden I, Bleasby A (2000) Emboss: The European molecular biology open software suite. Trends Genet 16:276-277.

Roberts DM, Harmon AC (1992) Calcium modulated proteins: targets of intracellular calcium signals in higher plants. Annu Rev Plant Physiol and Plant Mol Biol 43:375-414.

Roche DB, Buenavista MT, McGuffin LJ (2013) The FunFOLD2 server for the prediction of protein-ligand interactions. Nucleic Acids Res 41:W303-307.

Rogers S, Wells R, Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science 234:364-368.

Rost B, Casadio R, Fariselli P (1996) Refining neural network predictions for helical transmembrane proteins by dynamic programming. Proc Int Conf Intell Syst Mol Biol 4:192-200.

Rost B, Sander C (1993) Prediction of protein secondary structure at better than 70% accuracy. Mol Biol 232:584-599.

Rouster J, Leah R, Mundy J, Cameron-Mills V (1997) Identification of a methyl jasmonate-responsive region in the promoter of a lipoxygenase 1 gene expressed in barley grain. Plant J 11:513-523.

Saiardi A, Caffrey JJ, Snyder SH, Shears SB (2000) Inositol polyphosphate multikinase (ArgRIII) determines nuclear mRNA export in Saccharomyces cerevisiae. Biol Chem J 275:24686-24692.

Saiardi A, Erdjument-Bromage H, Snowman A, Tempst P, Snyder SH (1999) Synthesis of diphosphoinositol pentakisphosphate by a newly identified family of higher inositol polyphosphate kinases. Curr Biol 9:1323-1326.

Schwartz R, Ting CS, King J (2001) Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. Genome Res 11(5):703-9.

Sessa G, Meller Y, Fluhr R (1995) A GCC element and a G-box motif participate in ethylene-induced expression of the PRB-1b gene. Plant Mol Biol 28:145-153.

Shears SB (2004) How versatile are inositol phosphate kinases? Biochem J 377:265-280.

Shen X, Xiao H, Ranallo R, Wu WH, Wu C (2003) Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates. Science,299:112-114.

Shinozaki K, Yamaguchi-Shinozaki K (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Curr Opin Plant Biol

3:217-223.

Steger DJ, Haswell ES, Miller AL, Wente SR, O'Shea EK (2003). Regulation of chromatin remodeling by inositol polyphosphates. Science, 299:114-116.

Steinberg SF (2008) Structural basis of protein kinase C isoform function. Physiol Rev 88:1341-1378.

Stevenson-Paulik J, Bastidas GJ, Chiou ST, Frye RA, York JD (2005) Generation of phytate-free seeds in Arabidopsis through disruption of inositol polyphosphate kinases. Proc Natl Acad Sci 102:12612-12617.

Stevenson-Paulik J, Odom AR, York JD (2002) Molecular and biochemical characterization of two plant inositol polyphosphate 6-/3-/5-kinases. Biol Chem J, 277:42711-42718.

Stiles A, Qian X, Shears S, Grabau E (2008) Metabolic and signaling properties of an ITPK gene family in Glycine max. FEBS Letters 582:1853-1858.

Stiles AR (2007) Identification and characterization of late pathway enzymes in phytic acid biosynthesis in Glycine max, Dissertation, Virginia Polytechnic Institute and State University.

Sugimoto T, Kawasaki T, Kato T, Whittier RF, Shibata D, Kawamura Y (1992) cDNA sequence and expression of a phosphoenolpyruvate carboxylase gene from soybean. Plant Mol Biol 20: 743-747.

Sun Y, Thompson M, Lin G, Butler H, Gao Z, Thornburgh S, Yau K, Smith D, Shukla V (2007) Inositol 1,3,4,5,6-pentakisphosphate 2-kinase from maize: Molecular and biochemical characterization. Plant Physiol 144:1278-1291.

Suzuki M, Tanaka K, Kuwano M, Yoshida K (2007) Expression pattern of inositol phosphate-related enzymes in rice (Oryza sartiva L.): Implications for the phytic acid biosynthetic pathway. Gene 405:55-64.

Sweetman D, Johnson S, Caddick S, Hanke D, Brearley C (2006) Characterization of an Arabidopsis inositol 1,3,4,5,6-pentakisphosphate 2-kinase (AtIPK1). Biochem J 394:95-103.

Sweetman D, Stavridou I, Johnson S, Green P, Caddick S, Brearley C (2007) Arabidopsis thaliana

inositol 1,3,4-trisphosphate 5/6-kinase 4 (AtITPK4) is an outlier to a family of ATP-grasp fold proteins from Arabidopsis. FEBS Letters 581:4165-4171.

Takaiwa F, Yamanouchi U, Yoshihara T, Washida H, Tanabe F, Kato A, Yamada K (1996) Characterization of common cis-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. Plant Mol Biol 30:1207-1221.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725-2729.

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. Computational Chem 31:455-461.

Turner PJ. (2005) XMGRACE, Version 5.1.19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR.

Tuteja JH, Clough SJ, Chan WC, Vodkin LO (2004) Tissue-Specific Gene Silencing Mediated by a Naturally Occurring Chalcone Synthase Gene Cluster in Glycine max. The Plant Cell 16:819-835.

Wallace IM, Sullivan O, Higgins DG, Notredame C (2006) M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Res 34:1692-1699.

Washida H, Wu CY, Suzuki A, Yamanouchi U, Akihama T, Harada K, Takaiwa F (1999) Identification of cis-regulatory elements required for endosperm expression of the rice storage protein glutelin gene GluB-1. Plant Mol Biol 40:1-12.

Waterhouse AM, Procter JB, Martin DMA, Clamp M and Barton GJ (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench Bioinformatics 25(9):1189-1191.

Webb B, Sali A (2014) Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics 5.6.1-5.6.32.

Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407-W410.

Wu CY, Adach T, Hatano T, Washida H, Suzuki A, Takaiwa F (1998) Promoters of rice seed storage

protein genes direct endosperm-specific gene expression in transgenic rice. Plant Cell Physiol 39:885-889.

Wymer CL, Bibikova TN, Gilroy S (1997) Cytoplasmic free calcium distribution during the development of root hairs of Arabidopsis thaliana. Plant J 12:427-439.

Xu J, Brearley CA, Lin WH, Wang Y, Ye R, Mueller-Roeber B, Xu ZH, Xue HW (2005) A Role of Arabidopsis Inositol Polyphosphate Kinase, AtIPK2α, in Pollen Germination and Root Growth. Plant Physiol 137:94-103.

Yamaguchi-Shinozaki K, Mundy J, Chua NH (1989) Four tightly linked rab genes are differentially expressed in rice. Plant Mol Biol 14:29-39.

York JD, Odom AR, Murphy R, Ives EB, Wente SR (1999) A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. Science 285:96-100.

York JD, Odom AR, Murphy R, Ives EB, Wente SR (1999) A phospholipase c-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. Science,285:96-100.

Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. Proteins 64:643-651.

Zhang ZB, Yang G, Arana F, Chen Z, Li Y, Xia HJ (2007) Arabidopsis Inositol Polyphosphate 6-/3-Kinase (AtIpk2β) Is Involved in Axillary Shoot Branching via Auxin Signaling. Plant Physiol 144:942-951.

Tables

Table 1 Potential cis-acting elements identified in the 5′ regulatory sequences of plant *IPK2s*.

Classification	Name	Sequence	Source Organism	Reference
Endosperm	GCN_4 motif	GAAGCCA, TGAGTCA,	*Oryza sativa*	Takaiwa et al., 1996; Onodera et al., 2001
	Skn-1 motif	CAAGCCA, TGTGTCA, GTCAT	*Oryza sativa*	Washida et al., 1999
ABA	ABRE	CACGTG, TACGTG	*Arabidopsis thaliana*	Yamaguchi-Shinozaki et al., 1989
Gibberellin	GARE	AAACAGA, TCTGTTG,	*Brassica oleracea*	Mongkolsiriwatana et al., 2009;
		TATCCAC/T	*Oryza sativa*	Jun Niu et al., 2016
Light	Box 4	ATTAAT	*Petroselinum crispum*	Lois et al., 1989
	CATT-motif	GCATTC	*Zea mays*	Arguello-Astorga et al., 1996
	G-Box	CACGTG	*Pisum sativum, Arabidopsis thaliana, Zea mays*	Sessa et al., 1995; López-Ochoa et al., 2007
	ATCC/T-motif	AATCTAATCC/T	*Pisum sativum, Arabidopsis thaliana*	Arguello-Astorga et al., 1996
	TCT-motif	TCTTAC	*Arabidopsis thaliana*	Arguello-Astorga et al., 1996
	Box-I	TTTCAAA	*Pisum sativum*	Arguello-Astorga et al., 1996
	ACE	AAAACGTTTA	*Petroelinum crispum*	Feldbrugge et al., 1996
	GA-motif	AAGGAAGA	*Glycine max*	Arguello-Astorga et al., 1996
MeJA	CGTCA-motif	CGTCA	*Hordeum vulgare*	Kim et al., 1993; Rouster et al., 1997
	TGACG-motif	TGACG	*Hordeum vulgare*	
Drought	MBS	T/CAACTG	*Arabidopsis thaliana*	Shinozaki and Yamaguchi-Shinozaki, 2000; Abe et al., 2003
Anaerobic	ARE	TGGTTT	*Zea mays*	Nguyen et al. (2003)

Table 2 Physiochemical parameters of shortlisted plant IPK2 sequences computed using the ProtParam tool.

Organism	Accession No.	Sequence length	MW	pI	EC	Ii	Ai	GRAVY	-R	+R
Glycine max	AGW99177.1	279	30979.2	6.29	29910	29.72	89.07	-0.18	35	30
Glycine soja	KHN19419.1	178	19898.87	7.10	23950	29.73	95.73	-0.059	21	21
Vigna radiata	XP_022634636.1	286	31482.10	6.59	28420	38.17	93.67	0.045	29	27
Cajanus cajan	XP_020232596.1	267	29463.71	6.86	28420	33.70	91.31	-0.132	31	30
Phaseolus vulgaris	XP_007133857.1	264	28951.0	6.59	26930	38.90	90.38	-0.017	27	25
Lotus japonicus	AFK39224.1	283	31348.76	5.66	32430	42.36	87.74	-0.139	39	30
Cicer arietinum	XP_004510840.1	296	33061.70	6.12	26930	31.59	91.11	-0.250	38	33
Medicago truncatula	XP_003627882.1	426	47961.8	8.57	29910	31.89	85.02	-0.344	49	53
Arachis hypogaea	ALT56981.1	297	32684.01	6.45	28420	36.41	84.01	-0.227	34	32
Arachis duranensis	XP_015937330.1	297	32883.25	6.45	28420	37.35	84.01	-0.248	35	33
Corchorus capsularis	OMO94774.1	292	32498.97	5.84	28420	40.75	92.43	-0.159	38	33
Theobroma cacao	EOY22693.1	304	34213.77	6.32	31400	37.61	84.61	-0.292	39	36
Herrania umbratica	XP_021285810.1	305	34392.01	6.56	36900	40.38	83.38	-0.307	38	36
Durio zibethinus	XP_022738659.1	295	32954.54	6.71	28420	38.72	87.83	-0.237	36	35
Solanum lycopersicum	XP_004235863.1	294	32476.9	5.93	28420	34.32	83.23	-0.261	35	30
Solanum tuberosum	NP_001335929.1	408	44295.0	5.96	28420	20.21	93.19	-0.254	50	43
Brassica rapa	XP_009112000.1	274	30640.7	6.25	34380	30.25	80.66	-0.283	36	33
Capsicum baccatum	PHT32205.1	374	40995.36	5.97	32890	33.47	90.19	-0.297	50	44
Nicotiana attenuata	XP_019263239.1	368	40865.55	5.51	31400	30.30	93.51	-0.179	48	40
Capsicum annuum	PHT66064.1	374	41015.41	5.91	32890	32.27	92.01	-0.277	50	43
Capsicum chinense	PHU00941.1	374	40986.37	5.81	32890	32.27	92.27	-0.271	51	43
Brassica napus	XP_013749343.1	285	31961.0	5.87	35870	30.23	75.86	-0.387	40	34
Arabidopsis lyrata	XP_020870073.1	300	33671.08	5.82	31400	29.76	83.47	-0.324	42	35
Prunus avium	XP_021815221.1	281	31100.28	6.20	31400	26.98	86.41	-0.209	34	30
Trifolium pratense	PNY08557.1	273	30275.83	5.50	23950	29.99	96.70	-0.054	37	26
Arabidopsis thaliana	NP_200984.1	300	33486.7	5.72	31400	25.96	80.23	-0.329	40	31
Prunus persica	XP_007209445.1	282	31303.57	6.50	31400	28.39	85.39	-0.252	35	33
Lepidium latifolium	ACK86969.2	297	33212.5	6.59	34380	29.25	81.95	-0.285	31	29
Zea mays	XP_008649440.2	240	26665.45	10.48	29450	65.74	68.79	-0.517	22	37
Aegilops tauschii	XP_020147020.1	287	30607.86	6.13	26930	38.85	88.43	-0.052	33	29
Sorghum bicolor	XP_002452184.1	322	34550.64	8.26	28420	39.34	90.93	-0.013	32	34

Molecular weight (MW) (g/mol); isoelectric point (pI); extinction coefficient (EC) (M-1cm-1); instability index (Ii); aliphatic index (Ai); grand average hydropathy (GRAVY); number of negative residues (-R); number of positive residues (+R).

Table 3 Three-state description of secondary structure content and disulfide pattern prediction of IPK2 sequences. The secondary structure data was generated by joint prediction with SOPMA and PHD while disulfide bonding pattern was determined using DiANNA (DiAminoacid Neural Network Application) 1.1 server.

Organism	α-Helix (%Hh)	Extended strands (%Ee)	Random coil (%Cc)	Disulphide bridge prediction
Glycine max	26.88	13.26	31.54	94-121, 123-276, 158-186
Glycine soja	34.27	29.78	35.96	20-117, 22-85, 161-175
Vigna radiata	29.72	17.13	24.48	127-190, 263-281
Cajanus cajan	25.84	13.11	29.21	119-177
Phaseolus vulgaris	24.62	21.21	28.79	116-176, 208-249
Lotus japonicus	25.44	14.49	26.86	94-215, 149-187, 261-283
Cicer arietinum	21.96	14.53	28.72	122-185, 217-279, 260-274
Medicago truncatula	18.54	15.02	31.69	98-262, 325-357
Arachis hypogaea	24.58	14.48	33.67	120-188, 223-288, 269-292
Arachis duranensis	24.58	14.48	33.33	223-288, 269-292
Corchorus capsularis	19.52	20.21	25.34	120-187, 222-267
Theobroma cacao	19.74	16.78	39.14	50-120, 187-222, 267-298, 288-296
Herrania umbratica	18.69	14.75	42.95	120-296, 222-298, 267-288
Durio zibethinus	18.98	22.37	26.10	93-120, 149-222, 267-288
Solanum lycopersicum	18.03	18.37	30.27	120-187, 222-267
Solanum tuberosum	17.40	17.65	26.23	120-187, 222-283
Brassica rapa	13.50	21.17	22.63	144-186, 221-264
Capsicum baccatum	17.65	17.38	30.75	120-187, 222-283
Nicotiana attenuata	21.47	15.76	25.82	156-223, 258-303
Capsicum annuum	19.52	16.58	30.48	120-187, 222-267
Capsicum chinense	20.32	16.31	30.21	120-187, 222-267
Brassica napus	18.25	18.25	24.56	120-220, 144-185, 264-281
Arabidopsis lyrata	22.00	17.00	23.33	120-187, 222-272
Prunus avium	17.44	17.44	27.05	120-222, 156-187
Trifolium pratense	21.61	15.75	31.87	173-205, 203-250
Arabidopsis thaliana	22.00	16.67	24.67	120-187, 197-272
Prunus persica	17.02	17..02	29.79	120-222, 156-187
Lepidium latifolium	17.85	17.17	28.62	78-269, 120-187, 222-285
Zea mays	16.25	10.42	44.17	101-203, 192-200
Aegilops tauschii	28.57	7.67	31.36	110-188, 124-267
Sorghum bicolor	30.12	12.42	26.40	13-300, 137-160

Table 4 Validation parameters computed for the energy minimized 3D models of GmIPK2 protein built using different programs.

Model	RAMPAGE Percentage of residues in favoured region	ProSA Z-Score	Verify3D Percentage of residues with 3D-1D score >= 0.2
GmIPK2-S	92.5	-6.56	97.82
GmIPK2-P	82.2	-6.28	84.95
GmIPK2-M4	87.7	-5.84	79.93

Table 5 Hydrogen bonds between the active site residues of GmIPK2 and its substrates **(A)** 5GUG-I3P and **(B)** 4A69-I0P along with their distances and angles measured using Accelrys Discovery Studio Visualizer 4.1.

(A)

GmIPK2			5GUG-I3P		Distance	Angle
Residue	Atom	Chemistry	Atom	Chemistry	(A°)	(Degree°)
Lys105	HZ3	H-Donor	O7	H-Acceptor	2.074	49.9029
Lys126	HZ3	H-Donor	O14	H-Acceptor	1.852	44.2482
Arg153	O	H-Acceptor	H39	H-Donor	2.168	45.6080
Gln157	CA	H-Donor	O18	H-Acceptor	3.348	52.1869

(B)

GmIPK2			4A69-I0P		Distance	Angle
Residue	Atom	Chemistry	Atom	Chemistry	(A°)	(Degree°)
Lys105	HZ2	H-Donor	O18	H-Acceptor	2.693	47.8602
Lys105	HZ3	H-Donor	O24	H-Acceptor	2.078	28.0171
Lys126	HZ3	H-Donor	O28	H-Acceptor	2.179	35.7367
Lys126	O	H-Acceptor	H40	H-Donor	2.121	39.3959
Pro140	CA	H-Donor	O4	H-Acceptor	3.732	60.3277
Arg153	O	H-Acceptor	H36	H-Donor	2.187	16.0359
Ser217	HG	H-Donor	O27	H-Acceptor	2.899	22.5987

Figures Legends

Fig 1 Agarose gel showing PCR amplification product of *GmIPK2* gene isolated from developing seeds (8-10 mm) of *Glycine max cv*. Pusa-16. Lane M: 500 bp DNA ladder, lanes 1-3: ~910 bp *GmIPK2* fragment.

Fig 2 (a) RT-PCR expression analysis of *GmIPK2* gene in different plant tissues of Pusa-16 cultivar using soybean housekeeping gene *PEPCo* as an internal control. **(b)** Relative quantification of *GmIPK2* transcript levels in the samples analyzed above by qRT-PCR, normalized to soybean housekeeping gene *PEPCo*. The leaf tissue was taken as calibrator. The data is mean of technical triplicates of each of the three biological replicates with error bars indicating standard deviation (SD).

Fig 3 (a) RT-PCR expression analysis of *GmIPK2* gene in developing seeds of Pusa-16 cultivar using soybean housekeeping gene *PEPCo* as an internal control. **(b)** Relative quantification of *GmIPK2* transcript levels in the samples analyzed above by qRT-PCR, normalized to soybean housekeeping gene *PEPCo*. The 0-2 mm seed stage was taken as calibrator. The data is mean of technical triplicates of each of the three biological replicates with error bars indicating SD.

Fig 4 M-Coffee *multiple* sequence alignment diagram of selected plant IPK2 protein sequences rendered with Jalview. The sequence motifs shared amongst all the representatives are coloured according to percentage identity. The consensus row at the bottom shows the most frequent residue at each column or a '+' if two or more residues are equally abundant.

Fig 5 Topology map of GmIPK2 generated using ProMotif. There are a total of 8 α-helices (1-8) and 3 β-sheets (β-sheet 1: strands A1 and A2; β-sheet 2: strands B1, B2, B3 and B4; β-sheet 3: strands C1 and C2).

Fig 6 Phylogenetic tree showing evolutionary relationship of 31 plant IPK2 sequences divided into different clades, colour coded to indicate the plant family to which they belong. The posterior probability values are indicated corresponding to every node.

Fig 7 Homology model of GmIPK2_S protein rendered using PyMOL.

Fig 8 (a) Root mean square deviations and **(b)** Root mean square fluctuations of the Cα backbone atoms in GmIPK2_S model over 50 ns MD simulation.

Fig 9 Molecular docking of substrates to the GmIPK2_S homology model. Hydrogen bonding interactions of **(a)** 5GUG-I3P and **(b)** 4A69-I0P substrates with residues in the active site of GmIPK2_S protein. The substrate is depicted in stick representation with carbon atoms coloured turquoise, oxygen atoms red, hydrogen atoms yellow and phosphorus atoms blue. The interacting protein side chains are represented as maroon sticks in (a) and green sticks in (b). **(c)** & **(d)** 2D-schematic representation of the interactions shown in a & b respectively drawn using Discovery Studio Visualizer.

Fig 10 Time evolution of main hydrogen bonds formed between the GmIPK2_S protein and the 5GUG-I3P and 4A69-I0P ligands over 50 ns MD simulation.

Figure Click here to download Figure Figures.doc

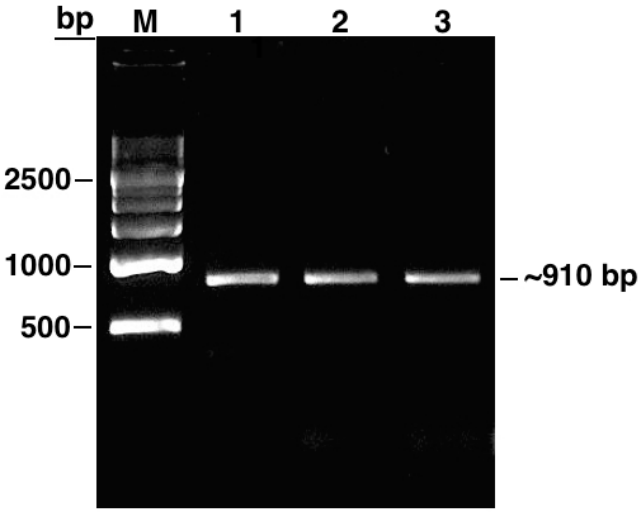


Fig 1

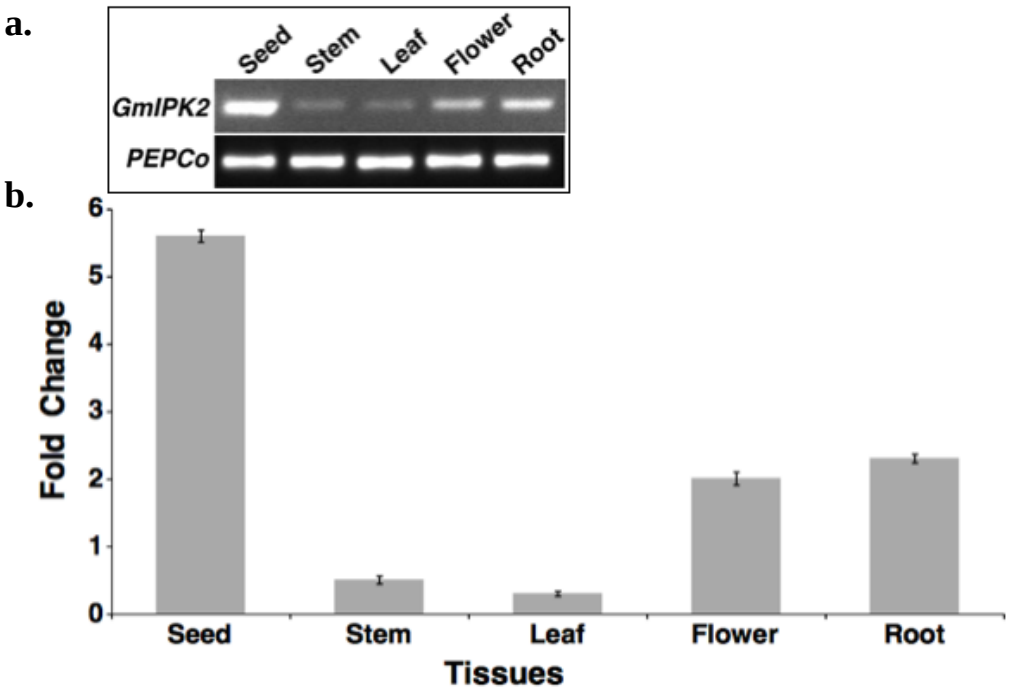

a.
Seed
Stem
Leaf
Flower
Root
GmIPK2
PEPCo
b.
Fold Change
0
1
2
3
4
5
6
Seed
Stem
Leaf
Flower
Root
Tissues

Fig 2

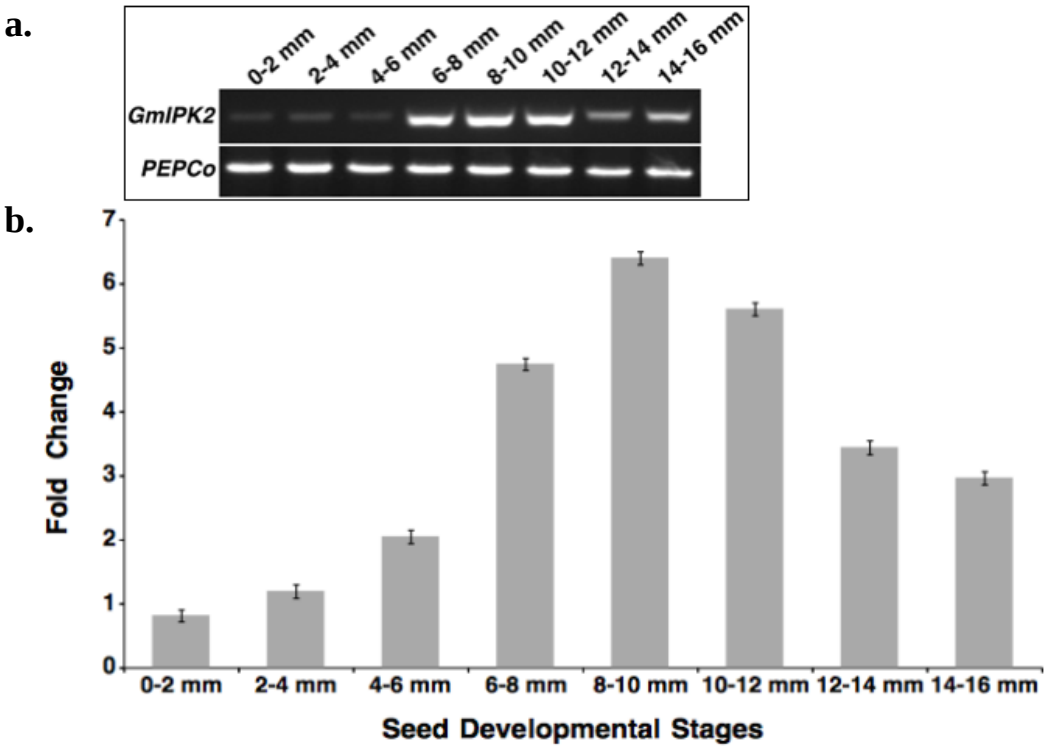
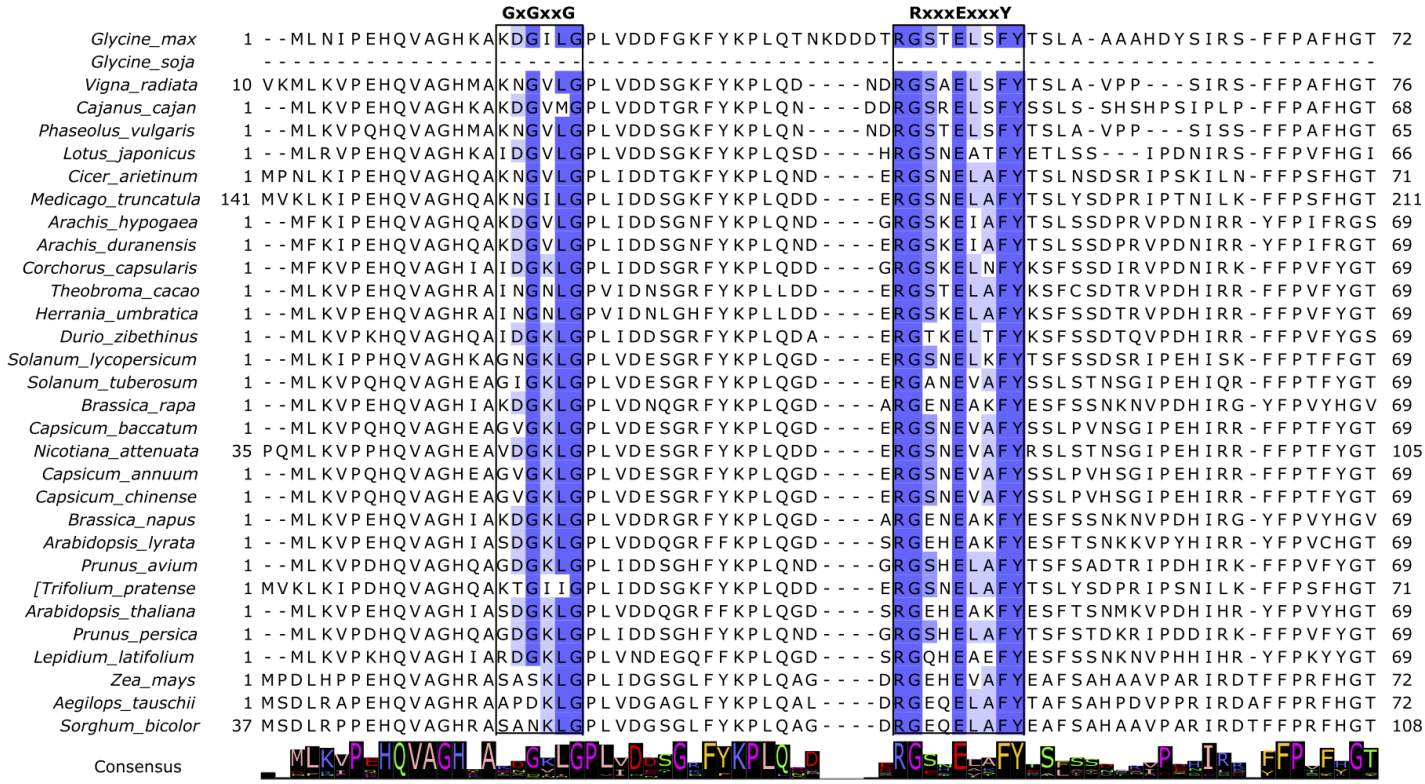
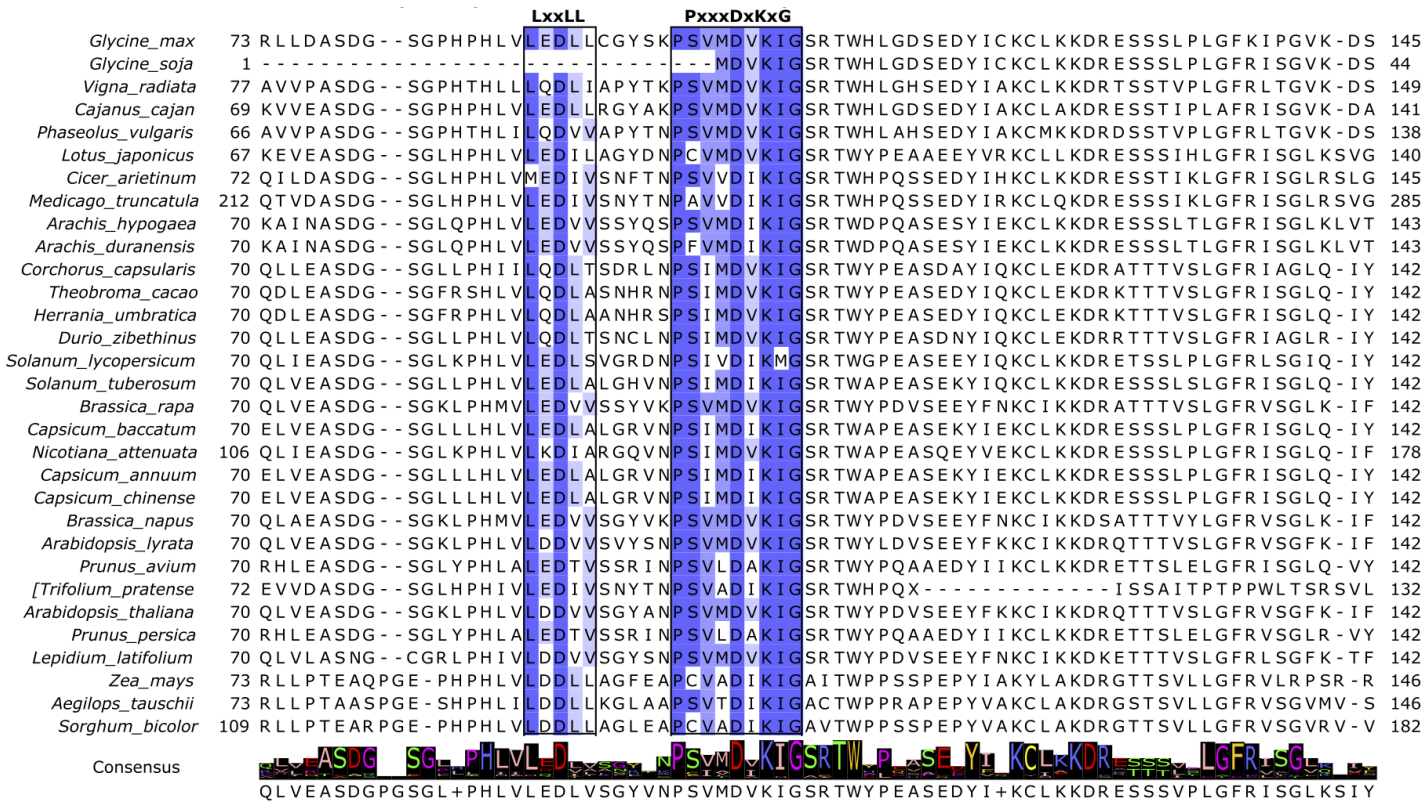

a.
0-2 mm
2-4 mm
4-6 mm
6-8 mm
8-10 mm
10-12 mm
12-14 mm
14-16 mm
GmIPK2
PEPCo
b.
Fold Change
0
1
2
3
4
5
6
7
0-2 mm
2-4 mm
4-6 mm
6-8 mm
8-10 mm
10-12 mm
12-14 mm
14-16 mm
Seed Developmental Stages

Fig 3

a.


GxGxxG
RxxxExxxY
Glycine_max 1 --MLNIPEHQVAGHKAKDGILGPLVDDFGKFYKPLQTNKDDDTRGSTELSFYTSLA-AAAHDYSIRS-FFPAFHGT 72
Glycine_soja
Vigna_radiata 10 VKMLKVPEHQVAGHMAKNGVLGPLVDDSGKFYKPLQD----NDRGSAELSFYTSLA-VPP---SIRS-FFPAFHGT 76
Cajanus_cajan 1 --MLKVPEHQVAGHKAKDGVMGPLVDDTGRFYKPLQN----DDRGSRELSFYSSLS-SHSHPSIPLP-FFPAFHGT 68
Phaseolus_vulgaris 1 --MLKVPQHQVAGHMAKNGVLGPLVDDSGKFYKPLQN----NDRGSTELSFYTSLA-VPP---SISS-FFPAFHGT 65
Lotus_japonicus 1 --MLRVPEHQVAGHKAIDGVLGPLVDDSGKFYKPLQSD----HRGSNEATFYETLSS---IPDNIRS-FFPVFHGI 66
Cicer_arietinum 1 MPNLKIPEHQVAGHQAKNGVLGPLIDDTGKFYKPLQND----ERGSNELAFYTSLNSDSRIPSKILN-FFPSFHGT 71
Medicago_truncatula 141 MVKLKIPEHQVAGHQAKNGILGPLIDDSGKFYKPLQDD----ERGSNELAFYTSLYSDPRIPTNILK-FFPSFHGT 211
Arachis_hypogaea 1 --MFKIPEHQVAGHQAKDGVLGPLIDDSGNFYKPLQND----GRGSKEIAFYTSLSSDPRVPDNIRR-YFPIFRGS 69
Arachis_duranensis 1 --MFKIPEHQVAGHQAKDGVLGPLIDDSGNFYKPLQND----ERGSKEIAFYTSLSSDPRVPDNIRR-YFPIFRGT 69
Corchorus_capsularis 1 --MFKVPEHQVAGHIAIDGKLGPLIDDSGRFYKPLQDD----GRGSKELNFYKSFSSDIRVPDNIRK-FFPVFYGT 69
Theobroma_cacao 1 --MLKVPEHQVAGHRAINGNLGPVIDNSGRFYKPLLDD----ERGSTELAFYKSFCSDTRVPDHIRR-FFPVFYGT 69
Herrania_umbratica 1 --MLKVPEHQVAGHRAINGNLGPVIDNLGHFYKPLLDD----ERGSKELAFYKSFSSDTRVPDHIRR-FFPVFYGT 69
Durio_zibethinus 1 --MLKVPKHQVAGHQAIDGKLGPLIDDSGRFYKPLQDA----ERGTKELTFYKSFSSDTQVPDHIRR-FFPVFYGS 69
Solanum_lycopersicum 1 --MLKIPPHQVAGHKAGNGKLGPLVDESGRFYKPLQGD----ERGSNELKFYTSFSSDSRIPEHISK-FFPTFFGT 69
Solanum_tuberosum 1 --MLKVPQHQVAGHEAGIGKLGPLVDESGRFYKPLQGD----ERGANEVAFYSSLSTNSGIPEHIQR-FFPTFYGT 69
Brassica_rapa 1 --MLKVPEHQVAGHIAKDGKLGPLVDNQGRFYKPLQGD----ARGENEAKFYESFSSNKNVPDHIRG-YFPVYHGV 69
Capsicum_baccatum 1 --MLKVPQHQVAGHEAGVGKLGPLVDESGRFYKPLQGD----ERGSNEVAFYSSLPVNSGIPEHIRR-FFPTFYGT 69
Nicotiana_attenuata 35 PQMLKVPPHQVAGHEAVDGKLGPLVDESGRFYKPLQDD----ERGSNEVAFYRSLSTNSGIPEHIRR-FFPTFYGT 105
Capsicum_annuum 1 --MLKVPQHQVAGHEAGVGKLGPLVDESGRFYKPLQGD----ERGSNEVAFYSSLPVHSGIPEHIRR-FFPTFYGT 69
Capsicum_chinense 1 --MLKVPQHQVAGHEAGVGKLGPLVDESGRFYKPLQGD----ERGSNEVAFYSSLPVHSGIPEHIRR-FFPTFYGT 69
Brassica_napus 1 --MLKVPEHQVAGHIAKDGKLGPLVDDRGRFYKPLQGD----ARGENEAKFYESFSSNKNVPDHIRG-YFPVYHGV 69
Arabidopsis_lyrata 1 --MLKVPEHQVAGHIASDGKLGPLVDDQGRFFKPLQGD----SRGEHEAKFYESFTSNKKVPYHIRR-YFPVCHGT 69
Prunus_avium 1 --MLKVPDHQVAGHQAGDGKLGPLIDDSGHFYKPLQND----GRGSHELAFYTSFSADTRIPDHIRK-FFPVFYGT 69
[Trifolium_pratense 1 MVKLKIPDHQVAGHQAKTGIIGPLIDDSGKFYKPLQDD----ERGSNELAFYTSLYSDPRIPSNILK-FFPSFHGT 71
Arabidopsis_thaliana 1 --MLKVPEHQVAGHIASDGKLGPLVDDQGRFFKPLQGD----SRGEHEAKFYESFTSNMKVPDHIHR-YFPVYHGT 69
Prunus_persica 1 --MLKVPDHQVAGHQAGDGKLGPLIDDSGHFYKPLQND----GRGSHELAFYTSFSTDKRIPDDIRK-FFPVFYGT 69
Lepidium_latifolium 1 --MLKVPKHQVAGHIARDGKLGPLVNDEGQFFKPLQGD----SRGQHEAEFYESFSSNKNVPHHIHR-YFPKYYGT 69
Zea_mays 1 MPDLHPPEHQVAGHRASASKLGPLIDGSGLFYKPLQAG----DRGEHEVAFYEAFSAHAAVPARIRDTFFPRFHGT 72
Aegilops_tauschii 1 MSDLRAPEHQVAGHRAAPDKLGPLVDGAGLFYKPLQAL----DRGEQELAFYTAFSAHPDVPPRIRDAFFPRFHGT 72
Sorghum_bicolor 37 MSDLRPPEHQVAGHRASANKLGPLVDGSGLFYKPLQAG----DRGEQELAFYEAFSAHAAVPARIRDTFFPRFHGT 108
Consensus
M+MLKVPEHQVAGHQAKDGKLGPLVDDSGRFYKPLQGDKDD+ERGSNELAFYTS+SSD+RVPDHIRRTFFPVFHGT

b.


LxxLL
PxxxDxKxG
Glycine_max 73 RLLDASDG--SGPHPHLVLEDLLCGYSKPSVMDVKIGSRTWHLGDSEDYICKCLKKDRESSSLPLGFKIPGVK-DS 145
Glycine_soja 1 ----------MDVKIGSRTWHLGDSEDYICKCLKKDRESSSLPLGFRISGVK-DS 44
Vigna_radiata 77 AVVPASDG--SGPHTHLLLQDLIAPYTKPSVMDVKIGSRTWHLGHSEDYIAKCLKKDRTSSTVPLGFRLTGVK-DS 149
Cajanus_cajan 69 KVVEASDG--SGPHPHLVLEDLLRGYAKPSVMDVKIGSRTWHLGDSEDYIAKCLAKDRESSTIPLAFRISGVK-DA 141
Phaseolus_vulgaris 66 AVVPASDG--SGPHTHLILQDVVAPYTNPSVMDVKIGSRTWHLAHSEDYIAKCMKKDRDSSTVPLGFRLTGVK-DS 138
Lotus_japonicus 67 KEVEASDG--SGLHPHLVLEDILAGYDNPCVMDVKIGSRTWYPEAAEEYVRKCLLKDRESSSIHLGFRISGLKSVG 140
Cicer_arietinum 72 QILDASDG--SGLHPHLVMEDIVSNFTNPSVVDIKIGSRTWHPQSSEDYIHKCLKKDRESSTIKLGFRISGLRSLG 145
Medicago_truncatula 212 QTVDASDG--SGLHPHLVLEDIVSNYTNPAVVDIKIGSRTWHPQSSEDYIRKCLQKDRESSSIKLGFRISGLRSVG 285
Arachis_hypogaea 70 KAINASDG--SGLQPHLVLEDVVSSYQSPSVMDIKIGSRTWDPQASESYIEKCLKKDRESSSLTLGFRISGLKLVT 143
Arachis_duranensis 70 KAINASDG--SGLQPHLVLEDVVSSYQSPFVMDIKIGSRTWDPQASESYIEKCLKKDRESSSLTLGFRISGLKLVT 143
Corchorus_capsularis 70 QLLEASDG--SGLLPHIILQDLTSDRLNPSIMDVKIGSRTWYPEASDAYIQKCLEKDRATTTVSLGFRIAGLQ-IY 142
Theobroma_cacao 70 QDLEASDG--SGFRSHLVLQDLASNHRNPSIMDVKIGSRTWYPEASEDYIQKCLEKDRKTTTVSLGFRISGLQ-IY 142
Herrania_umbratica 70 QDLEASDG--SGFRPHLVLQDLAANHRSPSIMDVKIGSRTWYPEASEDYIQKCLEKDRKTTTVSLGFRISGLQ-IY 142
Durio_zibethinus 70 QLLEASDG--SGLLPHLVLQDLTSNCLNPSIMDVKIGSRTWYPEASDNYIQKCLEKDRRTTTVSLGFRIAGLR-IY 142
Solanum_lycopersicum 70 QLIEASDG--SGLKPHLVLEDLSVGRDNPSIVDIKMGSRTWGPEASEEYIQKCLKKDRETSSLPLGFRLSGIQ-IY 142
Solanum_tuberosum 70 QLVEASDG--SGLLPHLVLEDLALGHVNPSIMDIKIGSRTWAPEASEKYIQKCLKKDRESSSLSLGFRISGLQ-IY 142
Brassica_rapa 70 QLVEASDG--SGKLPHMVLEDVVSSYVKPSVMDVKIGSRTWYPDVSEEYFNKCIKKDRATTTVSLGFRVSGLK-IF 142
Capsicum_baccatum 70 ELVEASDG--SGLLLHLVLEDLALGRVNPSIMDIKIGSRTWAPEASEKYIEKCLKKDRESSSLPLGFRISGLQ-IY 142
Nicotiana_attenuata 106 QLIEASDG--SGLKPHLVLKDIARGQVNPSIMDVKIGSRTWAPEASQEYVEKCLKKDRESSSLPLGFRISGLQ-IF 178
Capsicum_annuum 70 ELVEASDG--SGLLLHLVLEDLALGRVNPSIMDIKIGSRTWAPEASEKYIEKCLKKDRESSSLPLGFRISGLQ-IY 142
Capsicum_chinense 70 ELVEASDG--SGLLLHLVLEDLALGRVNPSIMDIKIGSRTWAPEASEKYIEKCLKKDRESSSLPLGFRISGLQ-IY 142
Brassica_napus 70 QLAEASDG--SGKLPHMVLEDVVSGYVKPSVMDVKIGSRTWYPDVSEEYFNKCIKKDSATTTVYLGFRVSGLK-IF 142
Arabidopsis_lyrata 70 QLVEASDG--SGKLPHLVLDDVVSVYSNPSVMDVKIGSRTWYLDVSEEYFKKCIKKDRQTTTVSLGFRVSGFK-IF 142
Prunus_avium 70 RHLEASDG--SGLYPHLALEDTVSSRINPSVLDAKIGSRTWYPQAAEDYIIKCLKKDRETTSLELGFRISGLQ-VY 142
[Trifolium_pratense 72 EVVDASDG--SGLHPHIVLEDIVSNYTNPSVADIKIGSRTWHPQX----------ISSAITPTPPWLTSRSVL 132
Arabidopsis_thaliana 70 QLVEASDG--SGKLPHLVLDDVVSGYANPSVMDVKIGSRTWYPDVSEEYFKKCIKKDRQTTTVSLGFRVSGFK-IF 142
Prunus_persica 70 RHLEASDG--SGLYPHLALEDTVSSRINPSVLDAKIGSRTWYPQAAEDYIIKCLKKDRETTSLELGFRVSGLR-VY 142
Lepidium_latifolium 70 QLVLASNG--CGRLPHIVLDDVVSGYSNPSVMDVKIGSRTWYPDVSEEYFNKCIKKDKETTTVSLGFRLSGFK-TF 142
Zea_mays 73 RLLPTEAQPGE-PHPHLVLDDLLAGFEAPCVADIKIGAITWPPSSPEPYIAKYLAKDRGTTSVLLGFRVLRPSR-R 146
Aegilops_tauschii 73 RLLPTAASPGE-SHPHLILDDLLKGLAAPSVTDIKIGACTWPPRAPEPYVAKCLAKDRGSTSVLLGFRVSGVMV-S 146
Sorghum_bicolor 109 RLLPTEARPGE-PHPHLVLDDLLAGLEAPCVADIKIGAVTWPPSSPEPYVAKCLAKDRGTTSVLLGFRVSGVRV-V 182
Consensus
QLVEASDGPGSGL+PHLVLEDLVSGYVNPSVMDVKIGSRTWYPEASEDYI+KCLKKDRESSSVPLGFRISGLKSIY

c.


LxxLL
Glycine_max 146 I-------S---------SWEPTRK-SLQCLSAHGVALVLNKFVSSNNINHDDHHPDCAFATEVYG---AVLERLQK 202
Glycine_soja 45 I-------S---------SWEPTRK-SLQCLSAHGVALVLNKFVSSNNINHDDHHPDCAFATEVYG---AVLERLQK 101
Vigna_radiata 150 L-------S---------SWEPSRT-FLQTLSAEGVALVLRKFVSSSNSDDSRSDPDCDFAAEVFG---AVLERLVE 206
Cajanus_cajan 142 L-------S---------AWEPPRK-SLQSLSADGALFILRKFVSSNA-----SDPDCAFASEVYA---AVLERLLE 193
Phaseolus_vulgaris 139 V-------S---------SWEPSRT-FLQSLSAEGVSLILRKFVSSAADSD-RSD--CDFAAEVLG---AVLERLVE 192
Lotus_japonicus 141 P-------S--KQDPSSCWQPHKKF-LQSLSADEVKLVLSKFVSSDGN---A-DPDSAFACEVFM---PVLKQLLE 199
Cicer_arietinum 146 S-------N--NQ----LWQPHKK-LLMDLNSEETMLILRKFVSSDGD---SDEPDCVFASRVYV---SVLEDLLE 201
Medicago_truncatula 286 P-------T--NQ----LWQPHKK-FLMDLSVDDVMLVLRKFVSSDGN---VDEPDCVFGSRVFV---PILEELLE 341
Arachis_hypogaea 144 S-------S--KD--ASVWQPGRK-FLQNLAANDAKLVLSRFVSSNVSSNDNIHPDCSFASKVFGGPSGILEQLLE 207
Arachis_duranensis 144 S-------S--KD--ASVWQPGRK-FLQNLAANDAKLVLSRFVSSNVSSNDNIHPDRSFASKVFGGPSGILEQLLE 207
Corchorus_capsularis 143 E-------S--KE--SGFWKPARK-EVQSFTVDSVRSVLRKFVSSNVSVGSDVNPDCSFASRVYGGSAGILEQLLE 206
Theobroma_cacao 143 E-------S--KE--SGFWKPEKK-QVQSFTADNVRLVLRKFVSSNSSIGSNENPDCSLASIVYGGSAGILEQLLE 206
Herrania_umbratica 143 E-------S--KE--SGFWKPEKK-QVQSFTADNVRLVLRKFVSSNSSIGSNENPDCSLAPIVYGGSAGILEQLLE 206
Durio_zibethinus 143 E-------S--KE--SGCWKPGKK-EVQSFTADNVRSVLRKFVSSNSSIGSNVNPDCSFASIVYGGSAGILEQLLE 206
Solanum_lycopersicum 143 G-------N--KE--SGYWKPERT-SVQNLSAEEVKLILKRFVSSNTSTDMDMQPDCAFAATVYGGSTGILSQLLE 206
Solanum_tuberosum 143 R-------S--KE--LGFWKPGKK-AAQKLSTEEVKLVLRRFVSSNTLNDLDLRPDCAFASTVYGGSTGILSQLLE 206
Brassica_rapa 143 D-------C--QE--SRFWRPDKR-VVLAYKADGARTALRKFVSSNSLAD-DSVPDCAYASEVYGGSDGILVKLLE 205
Capsicum_baccatum 143 R-------S--KE--LGFWKPGKK-AAQKLSTEEVKLVLRRFVSSNTLNDSDSKPDCAFASTVYGGSTGILSQLLE 206
Nicotiana_attenuata 179 R-------S--KE--LGFWKPGKK-AAQKFSTEEVKLVLRRFVSSNTLNDLDLEPDCALASTVYGGSAGILSQLLE 242
Capsicum_annuum 143 R-------S--KE--LGFWKPGKK-AAQKLSTEEVKLVLRRFVSSNTLNDLDSKPDCAFASTVYGGSTGILSQLLE 206
Capsicum_chinense 143 R-------S--KE--LGFWKPGKK-AAQKLSTEEVKLVLRRFVSSNTLNDLDSKPDCAFASTVYGGSTGILSQLLE 206
Brassica_napus 143 D-------C--QE--SRFWRPDKR-VVLAYKVDGARMALRKFVSSNPSAD--SVPDCAYASEVYGGSNGILEKLLE 204
Arabidopsis_lyrata 143 D-------H--QE--SSFWRAEKK-LVLGYNADGARLALRKFVSSNSPADSNLTPNCAFASEVYGGSNGILAQLLE 206
Prunus_avium 143 G-------N--KE--TGFWKPDKK-CVLNFSIEDTRLALRKFVSSNPSTGSDIKPDCSFASTVYGGPAGILAQLLE 206
[Trifolium_pratense 133 E-------H--GT----HNPPKITFVLMDLSAEDAMLILRKFVSSDGN---VDEPDCVFASRVFV---HVLEELLE 189
Arabidopsis_thaliana 143 D-------H--QE--SSFWRAEKK-LVLGYNADGARLALRKFVSSNSPADSNLTPNCAFASEVYGGCNGILAQLLE 206
Prunus_persica 143 G-------N--KE--TGFWKPDKK-CVLNFSIEDTRLALRKFVSSNPSTDSDIKPDCSFASTVYGGPAGILAQLLE 206
Lepidium_latifolium 143 D-------H--QE--SSFWRPEKK-LVLGYNVDSVRLALRKYVSSNSPTDSNSIPNCAFAFEVYGGSNGVLAQLLV 206
Zea_mays 147 PRGRRVADGAPGG--EG-YGHRRR-PPRAPALRVIRLPTRGWTARSRRRCTEEK--VESCHSCASSRHGWRSRLC- 215
Aegilops_tauschii 147 D-------P--SG---AVWRPDRS-ELKGTDIPGVRRMLRRYVSSAGGD--GGGEDCALAAAVYGGEGGVLAQLRE 207
Sorghum_bicolor 183 G-------P--GG---AVWRTERP-EVKALDTAGVRRVLRRYVSSEG-------MDCALAAAVYGGKGGVLSQLRE 238
Consensus
DRGRRVADSAPKEDPSGFWKPGKKF+VQSLSA+GVRLVLRKFVSSNSSNDSDSKPDCAFASEVYGGSAGILEQLLE

Fig 4

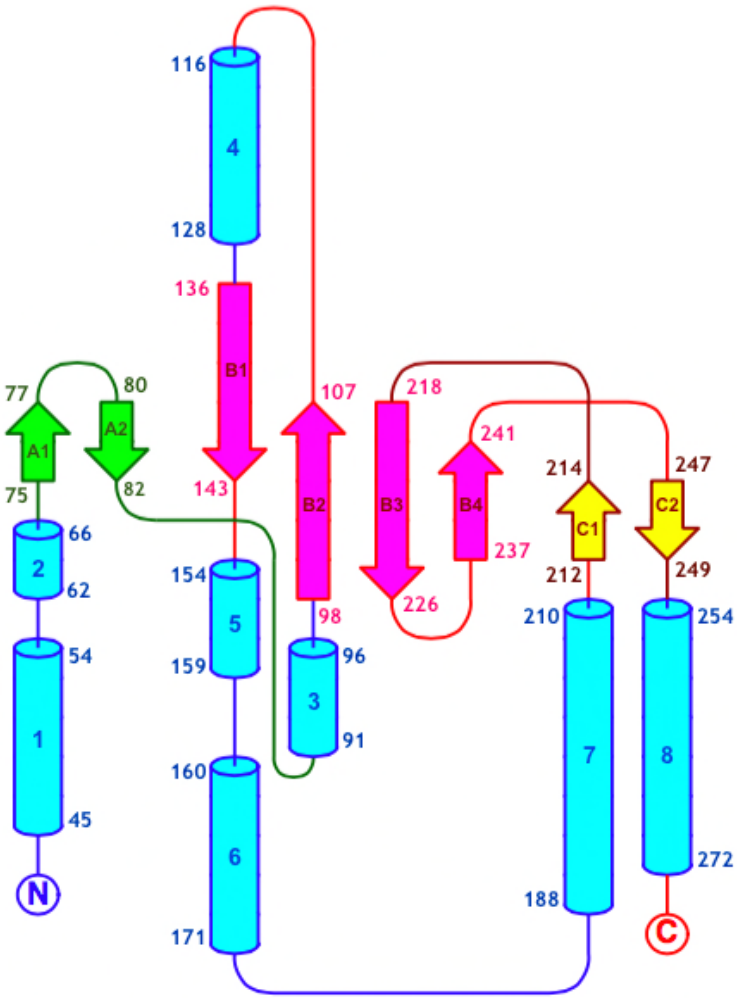

116
4
128
136
B1
107
B2
143
77
80
A1
A2
75
82
66
2
62
54
1
45
N
154
5
159
98
96
3
91
160
6
171
218
B3
241
B4
237
226
214
C1
212
210
7
188
247
C2
249
254
8
272
C

Fig 5

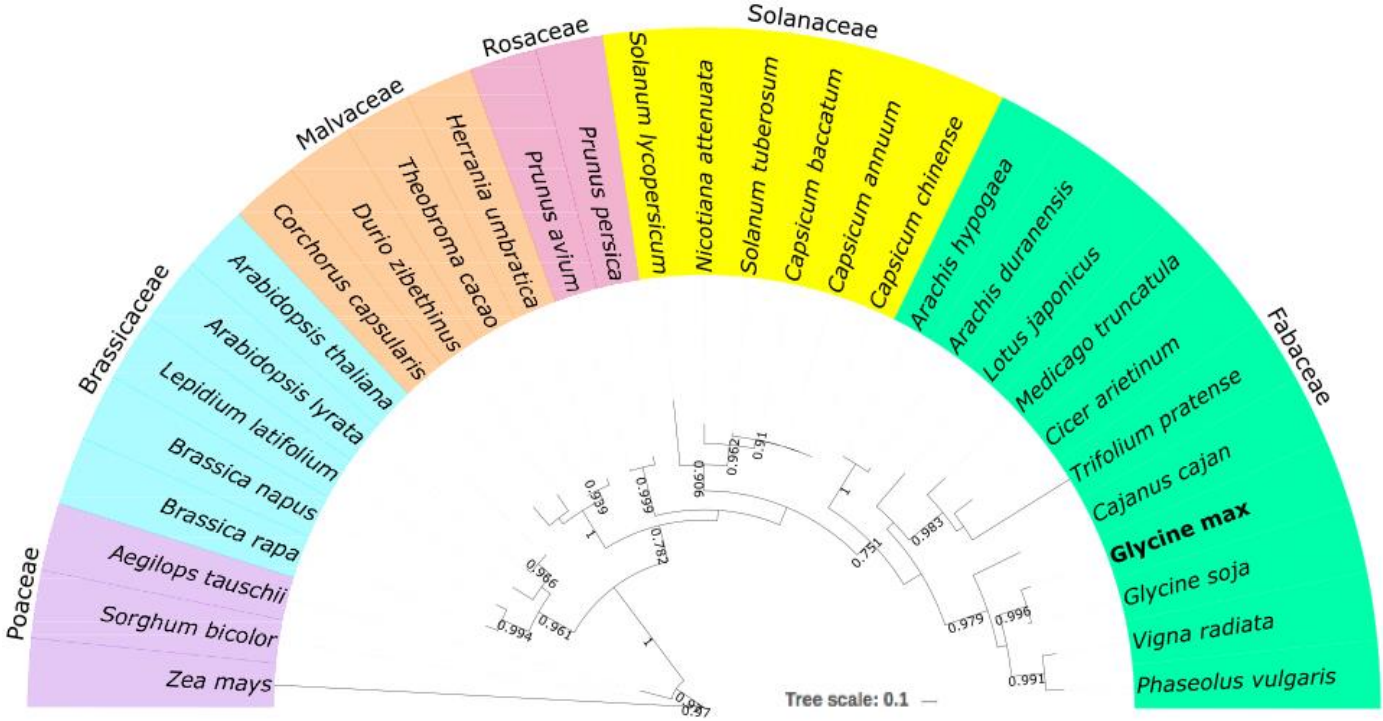

Poaceae
Brassicaceae
Malvaceae
Rosaceae
Solanaceae
Fabaceae
Zea mays
Sorghum bicolor
Aegilops tauschii
Brassica rapa
Brassica napus
Lepidium latifolium
Arabidopsis lyrata
Arabidopsis thaliana
Corchorus capsularis
Durio zibethinus
Theobroma cacao
Herrania umbratica
Prunus avium
Prunus persica
Solanum lycopersicum
Nicotiana attenuata
Solanum tuberosum
Capsicum baccatum
Capsicum annuum
Capsicum chinense
Arachis hypogaea
Arachis duranensis
Lotus japonicus
Medicago truncatula
Cicer arietinum
Trifolium pratense
Cajanus cajan
Glycine max
Glycine soja
Vigna radiata
Phaseolus vulgaris
Tree scale: 0.1

Fig 6

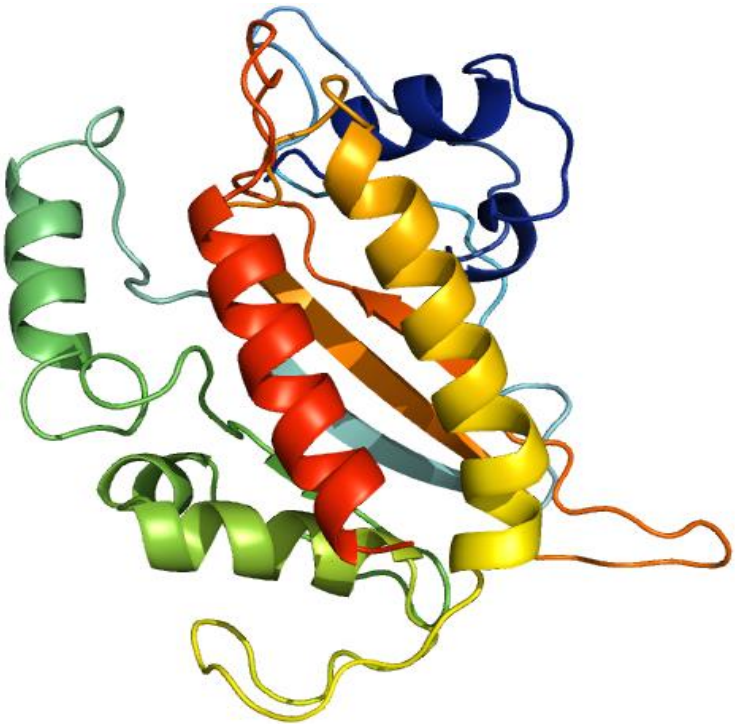
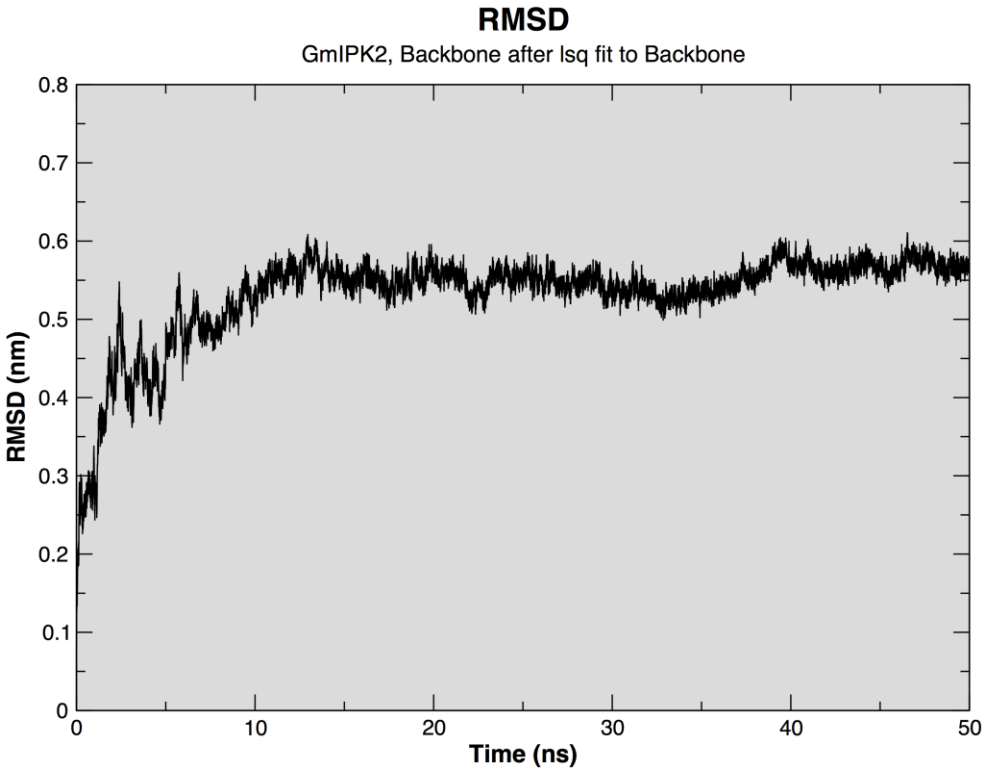


Fig 7

a.


RMSD
GmIPK2, Backbone after lsq fit to Backbone
RMSD (nm)
Time (ns)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
10
20
30
40
50

b.

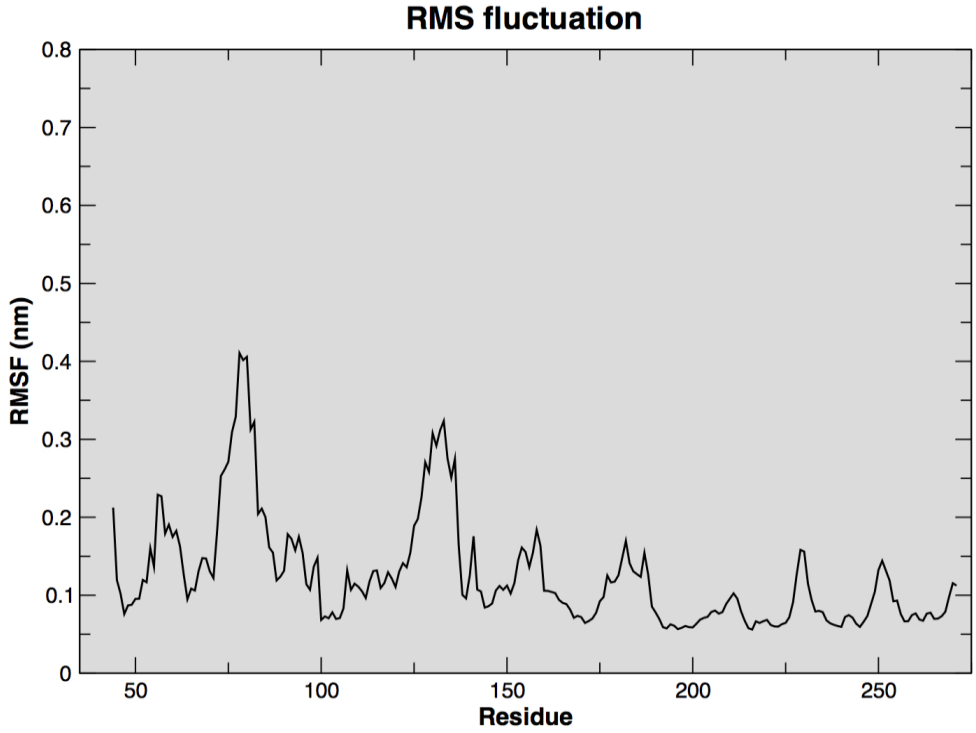
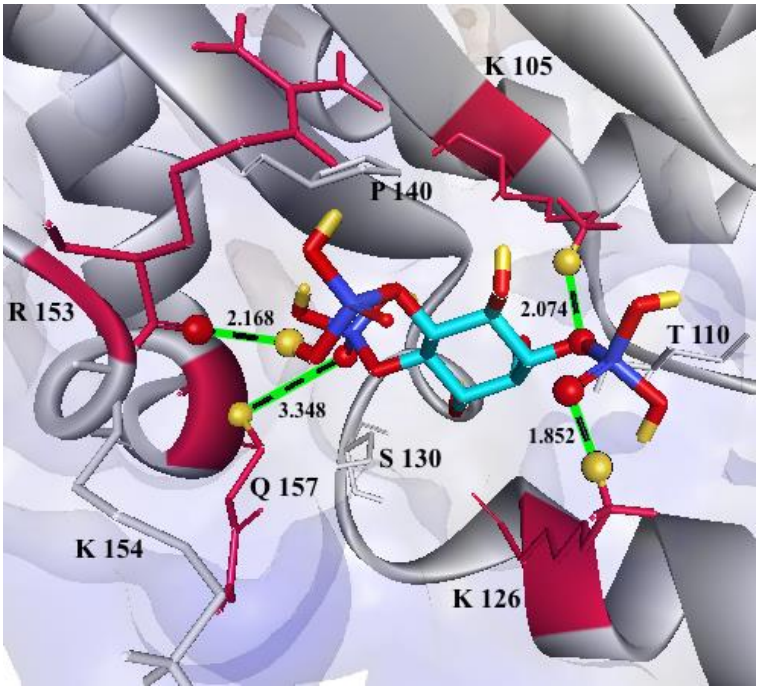
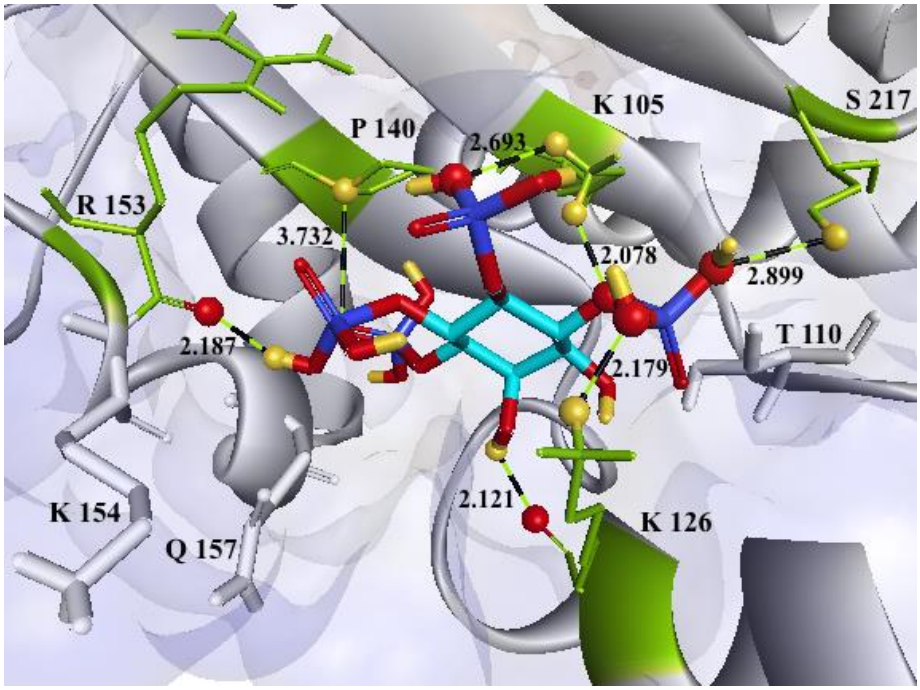

RMS fluctuation
RMSF (nm)
Residue
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
50
100
150
200
250

Fig 8

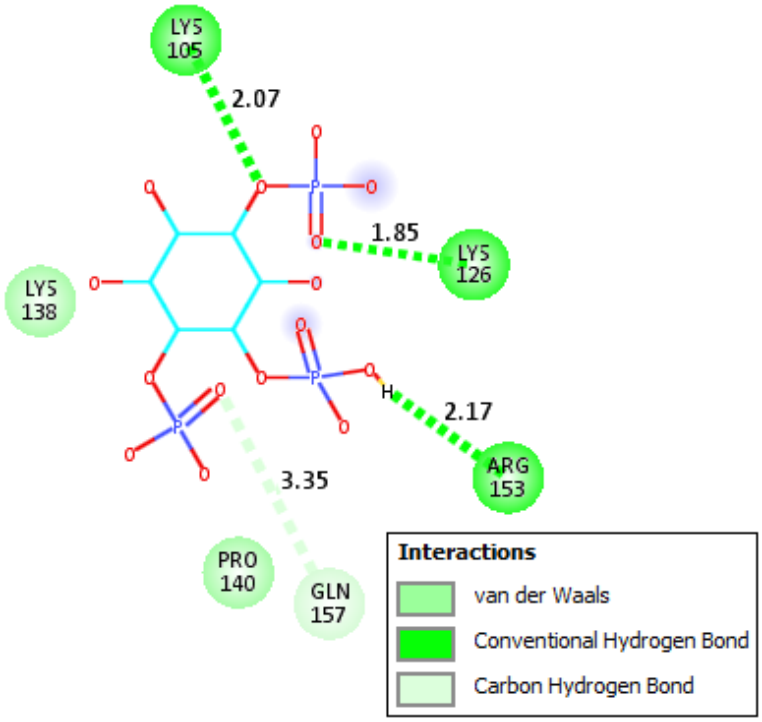
a.


K 105
P 140
R 153
2.168
2.074
T 110
3.348
1.852
S 130
Q 157
K 154
K 126

b.


K 105
S 217
P 140
2.693
R 153
3.732
2.078
2.899
2.187
T 110
2.179
2.121
K 154
K 126
Q 157

c.


LYS 105
2.07
1.85
LYS 126
LYS 138
2.17
ARG 153
3.35
PRO 140
GLN 157
Interactions
van der Waals
Conventional Hydrogen Bond
Carbon Hydrogen Bond

d.

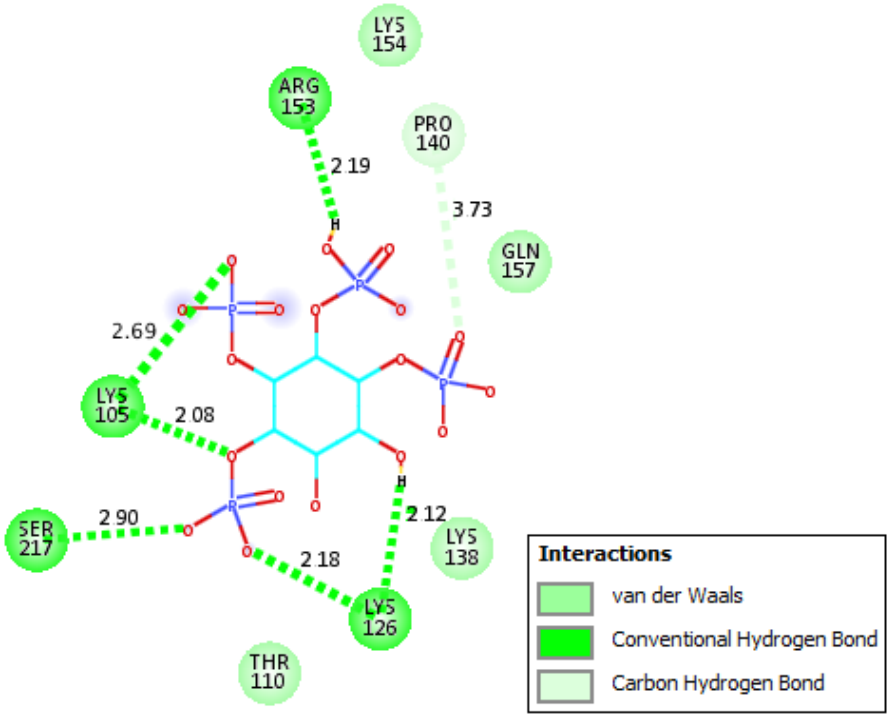

LYS 154
ARG 153
2.19
PRO 140
3.73
GLN 157
2.69
LYS 105
2.08
SER 217
2.90
2.18
2.12
LYS 138
LYS 126
THR 110
Interactions
van der Waals
Conventional Hydrogen Bond
Carbon Hydrogen Bond

Fig 9

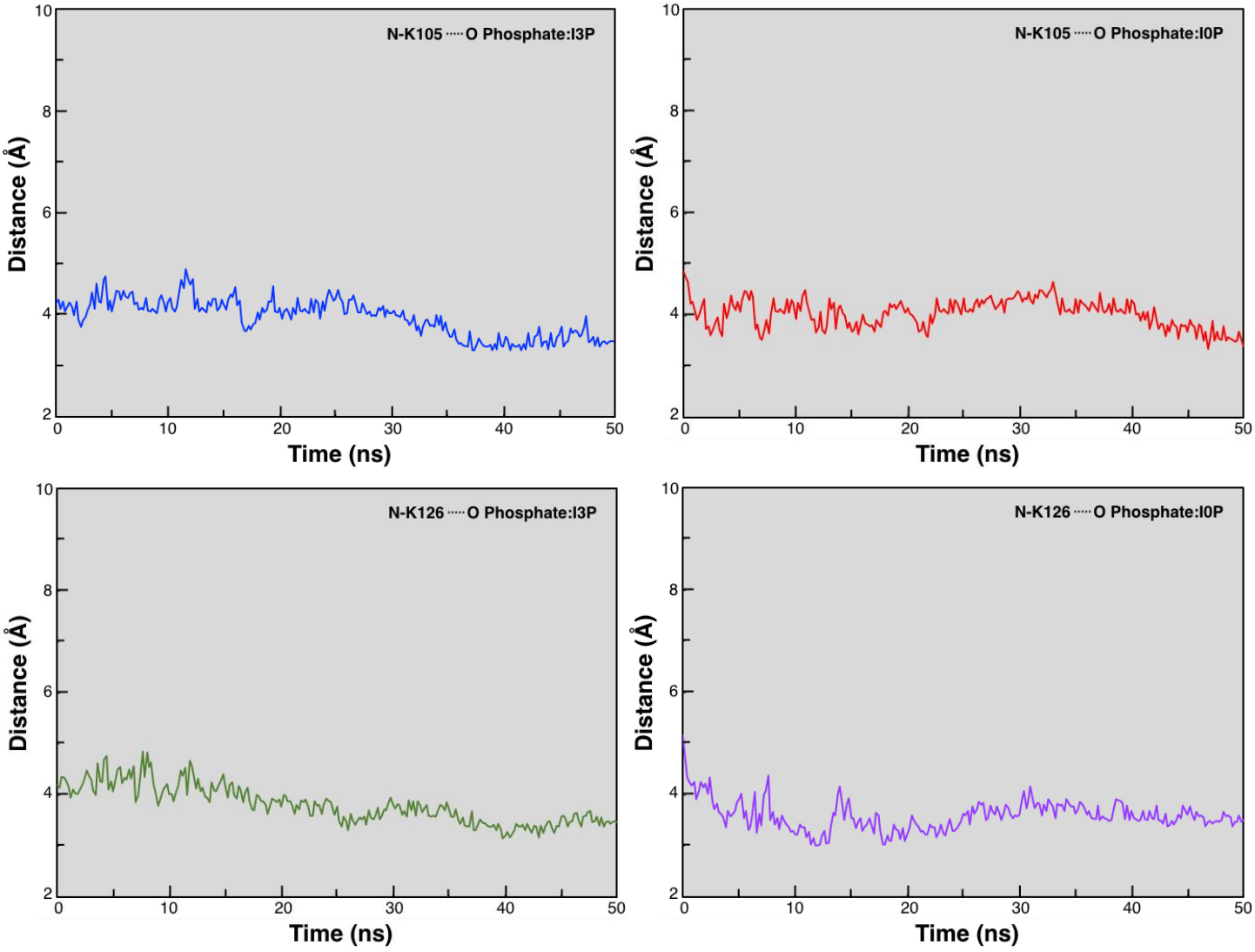

N-K105 ····· O Phosphate:I3P
N-K105 ····· O Phosphate:I0P
N-K126 ····· O Phosphate:I3P
N-K126 ····· O Phosphate:I0P
Distance (Å)
Time (ns)
2
4
6
8
10
0
10
20
30
40
50

Fig 10

Supplementary Material
Click here to download Supplementary Material
Supplementary Material.doc

Supplementary Data

Molecular characterization, modeling and docking analysis of late phytic acid biosynthesis pathway gene, *inositol polyphosphate 6-/3-/5-kinase*, a potential candidate for developing low phytate crops

3 BIOTECH

MANSI PUNJABI[1,2], NAVNEETA BHARADVAJA[1], ARCHANA SACHDEV*[2], VEDA KRISHNAN[2]

[1] *Department of Biotechnology, Delhi Technological University (Formerly Delhi College of Engineering), New Delhi - 110042, India. Email: mansipunjabi@gmail.com, navneetab@dce.ac.in*

[2] *Division of Biochemistry, Indian Agricultural Research Institute, New Delhi - 110012, India*

**Corresponding author: Dr. Archana Sachdev, Tel: 91-11-25842038, Email: arcs_bio@yahoo.com*

TABLES

Table S1 Consensus subcellular localization prediction of GmIPK2 through 4 different servers.

Tools	Subcellular localization	Value	Reliability index	Method
CELLO v2.5	Cytoplasmic	4	5	SVM
SubLoc v1.0	Cytoplasmic	2.281	Higher Value	SVM
WoLF PSORT	Cytoplasmic	7	14	K nearest neighbors classifier
MemType-2L	Cytoplasmic	Not Defined	Not Defined	Pre-PSS

Pre-PSSM and SVM stands for Pseudo Position-Specific Score Matrix and support vector machines respectively.

Table S2 Sequence motifs identified in GmIPK2 protein using MyHits Motif Scan.

Motif information	No. of sites	Amino acid residues
Inositol Polyphosphate Kinase	1	16-269
Tyrosine kinase phosphorylation site	1	42-50
Protein kinase C phosphorylation site	3	35-37, 61-63, 152-154
Casein kinase II phosphorylation site	3	35-38, 100-103, 147-150
c-AMP and c-GMP dependent protein kinase phosphorylation site	1	231-234
N-myristoylation site	2	141-146, 259-264

Table S3 Summary of GmIPK2 models produced using Modeller 9.16.

Filename	molpdf	DOPE score	GA341 score
GmIPK2-M1	1904.96094	-26781.98828	1.00000
GmIPK2-M2	1823.05139	-27677.49023	1.00000
GmIPK2-M3	1818.68884	-26874.69922	1.00000
GmIPK2-M4	1734.48291	-27846.58203	1.00000
GmIPK2-M5	2397.21655	-26359.73828	1.00000

FIGURES

Fig S1 Prediction of transmembrane segments in GmIPK2 protein using WaveTM server. Graph shows that amino acid residues 259 to 274 belong to the transmembrane helix category towards C-terminal.

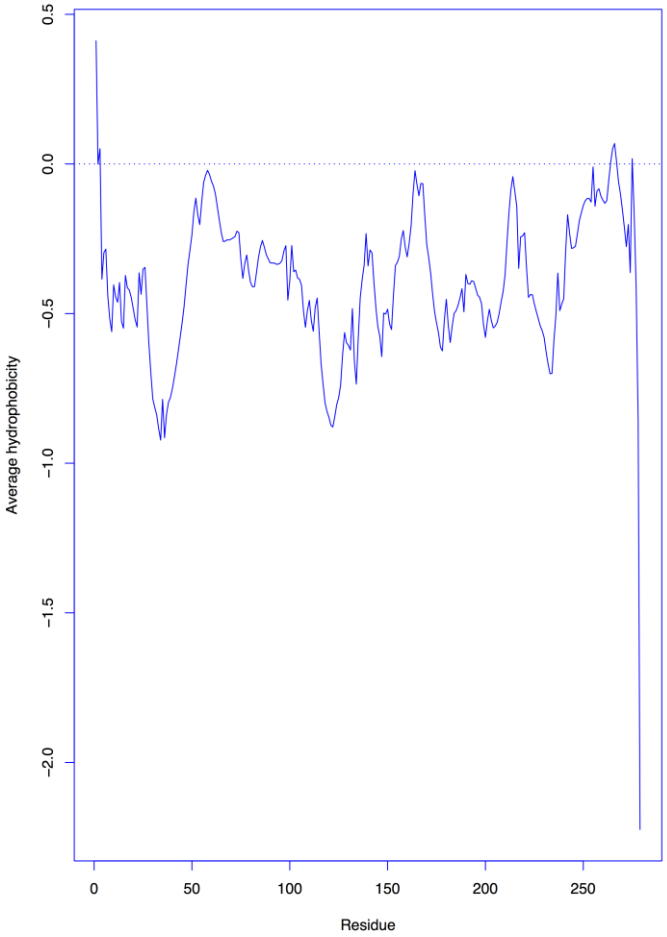


Fig S2 MEME suite motif analysis of selected IPK2 proteins showing eleven identified conserved ungapped motifs.

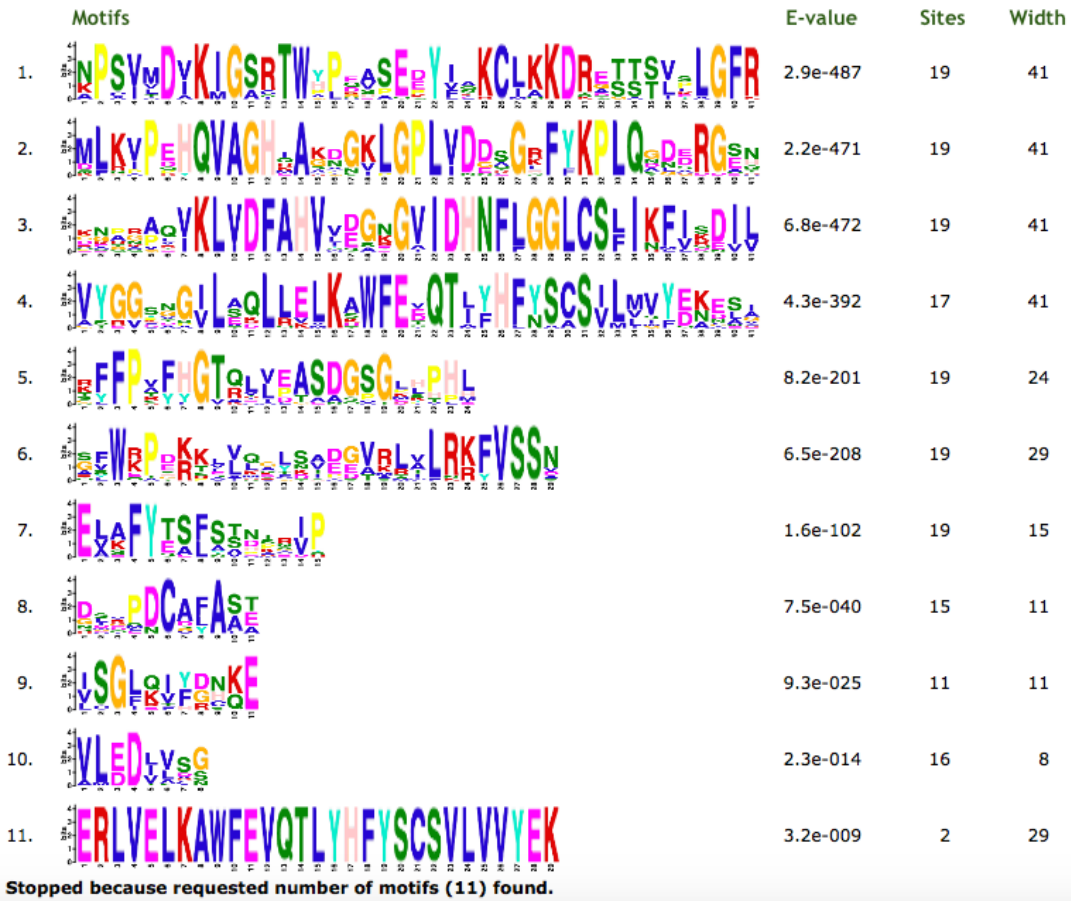
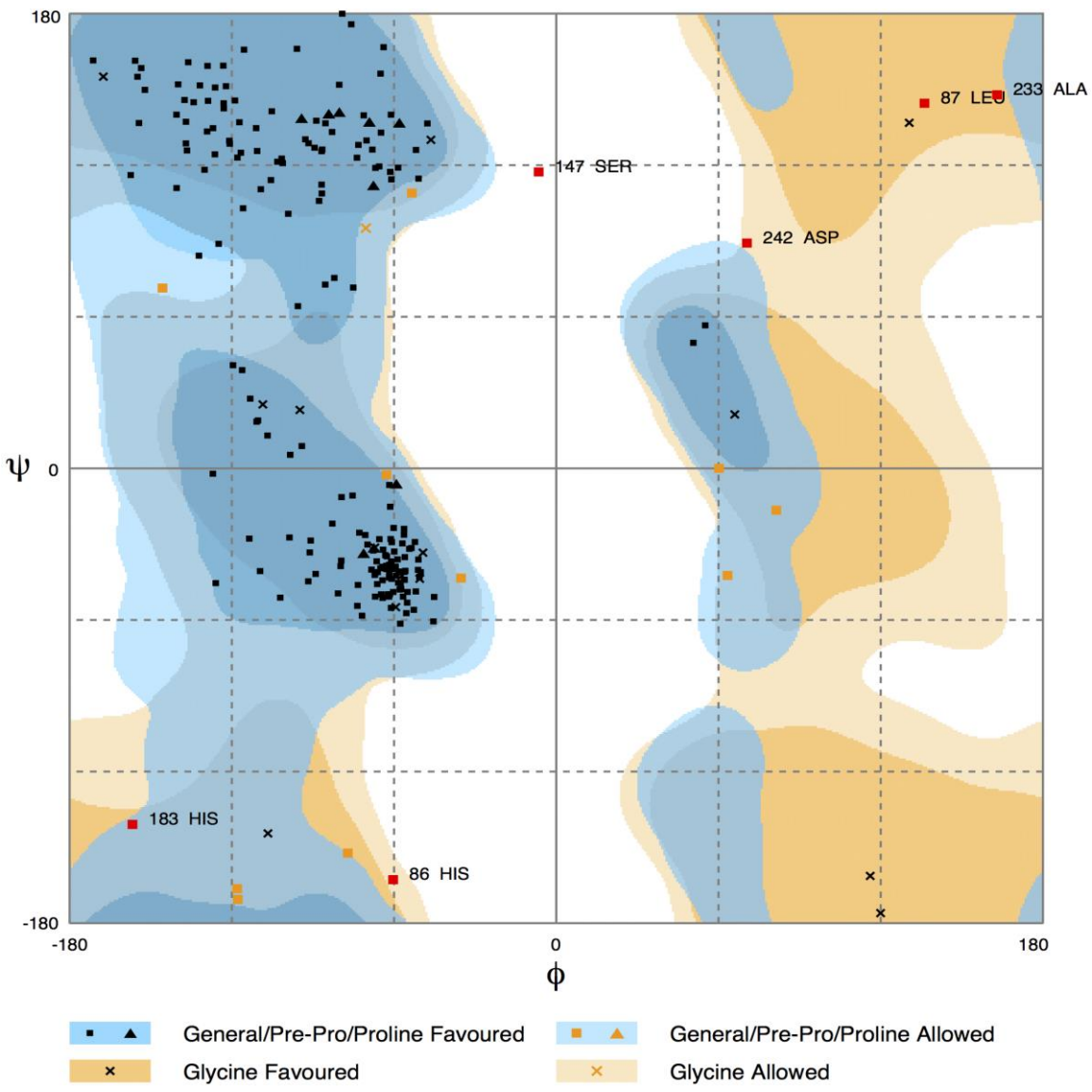
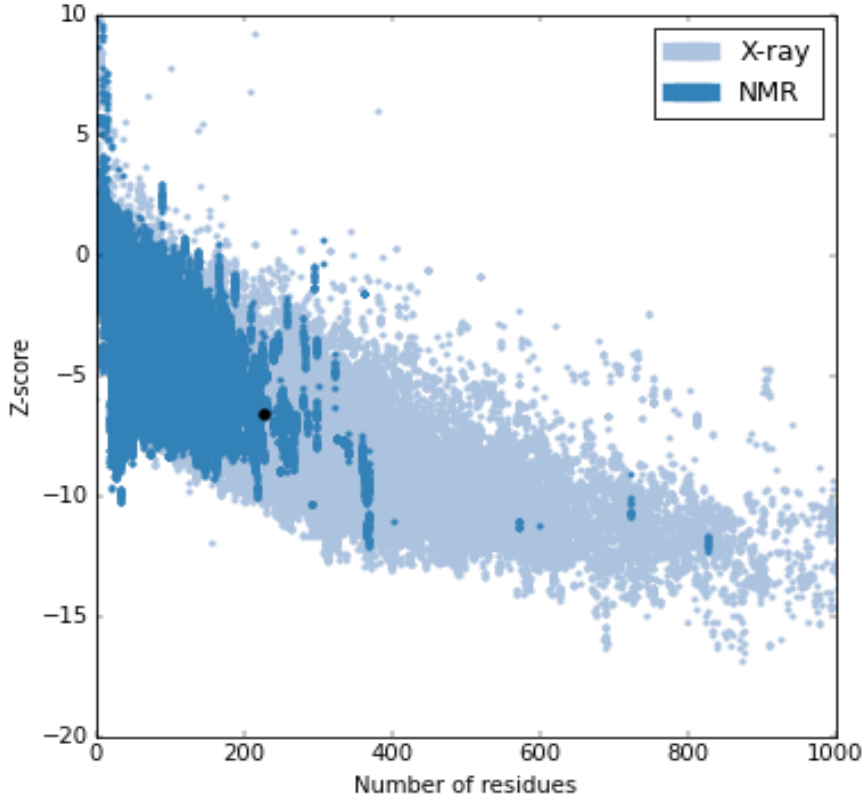


Fig S3 (a) Ramachandran plot generated by RAMPAGE server, validating backbone dihedral angels of the energy minimized model. **(b)** Z-score plot showing a z-score of -6.56 indicating the overall quality is within the range of scores typically found for native proteins of similar size. **(c)** Energy plot showing negative energy values throughout the sequence indicating a good local model quality. **(d)** The 3D-1D profile drawn to verify compatibility of model structure with its amino acid sequence. Residues showing positive compatibility scores are reasonably folded.

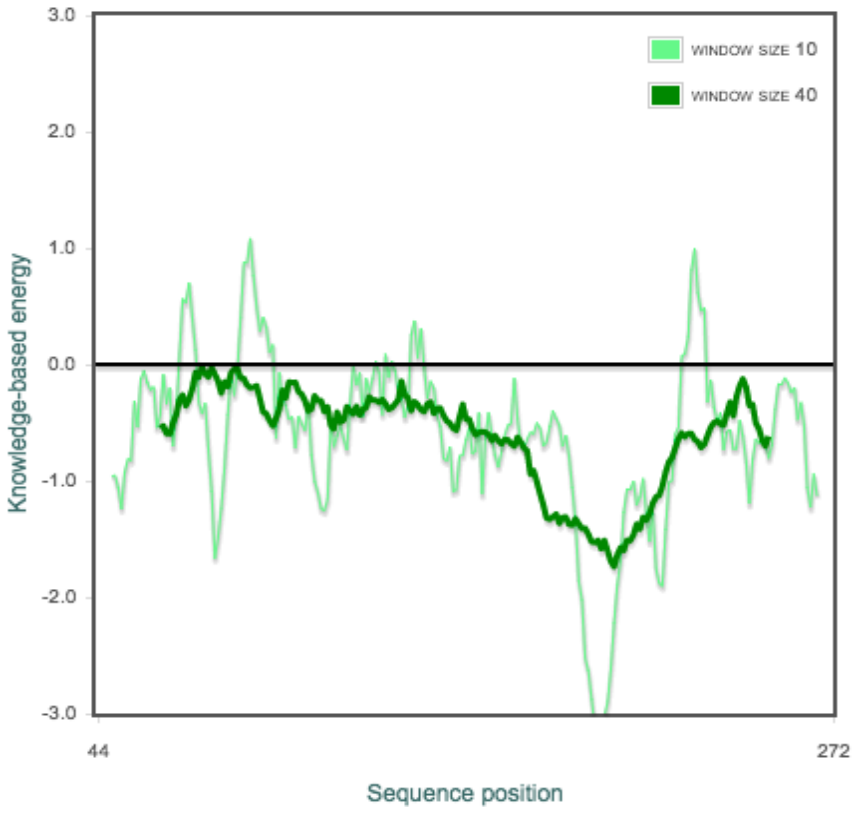
a.



b.


X-ray
NMR
Z-score
Number of residues
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

c.


WINDOW SIZE 10
WINDOW SIZE 40
Knowledge-based energy
Sequence position
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
44
272

d.

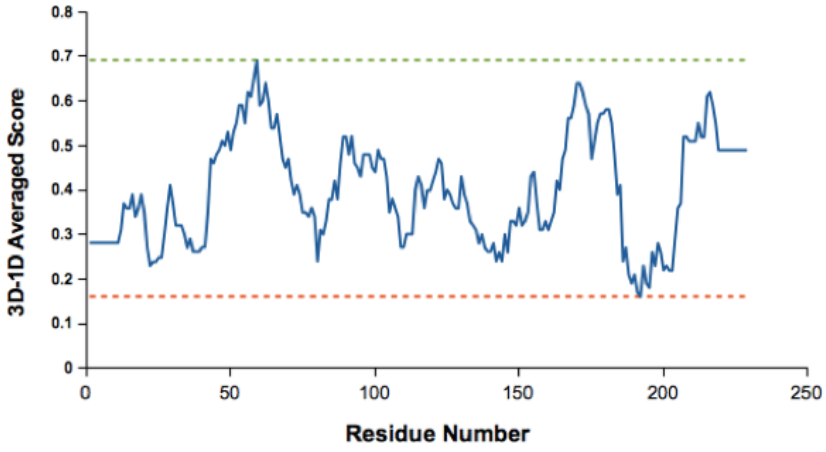

3D-1D Averaged Score
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250
Residue Number

Fig S4 (a) 3D structural alignment between model (GmIPK2-S) and the template protein (4FRF_A) generated using MATRAS server showing 91.3% secondary structure identity. **(b)** Superimposed 3D structures of template 4FRF_A protein (red) and refined GmIPK2_S model (blue).

a.

```
[BEGIN ALIGNMENT]
     :  H1           -                    E1  E2                  E3
     :  HHHHHHHHHHS -  STTGGGS  B      EETTEE SS    B  GGGG SSEEEE
  44 :STELSFYTSLAAAA-HDYSIRSFFPAFHGTRLLDASDGSGPHPHLVLEDLLCGYSKPSVM: 102
        *  ** *          *   **  ***               *** **  * *****
  41 :EIEVKFYESFSSNTEVPEHIHRYFPVYHGT-------------MMVLENLLAEYTKPSVM:  97
     :  HHHHHHHHHH TTS HHHHTTS  B   -------------   B  GGGG SSEEEE
     :  H1             H2           -------------            E1

     :              H2                   E4        ----      H3
     :EEEE SSS  TTS HHHHHHHHHHHTTSTTTTSSEEEEEEEBTT----TTB   HHHHTT
 103 :DVKIGSRTWHLGDSEDYICKCLKKDRESSSLPLGFKIPGVKDSI----sSWEPTRKSLQC: 158
      *** *****    ** ** ******        ** * *           * * ** *
  98 :DVKMGSRTWYPDASEEYIQKCLKKDTGTTTVSSGFRISGFEVYDHKESSFWKPERKLLRG: 157
     :EEEE SSS  TTS HHHHHHHHHHHTTSHHHHSSEEEEEEEEEETTTTEEE   HHHHHT
     :              H3            H4    E2            E3    H5

     :  H4                           H5   - --                  E5
     :  HHHHHHHHHHTT S   STT        HHHHH-H--HHHHHHHHHHHHHHHH  SEE
 159 :LSAHGVALVLNKFVSSNNINHDDHHPDCAFATEVY-G--AVLERLQKLKDWFEVQTVYHF: 215
      *   *  * * ******        ** ***  ** *    *  *  ** *** ** ***
 158 :LDVDGARLTLRKFVSSNSL------PDSAFASSVYGGSHGILTQLLELKTWFENQTLYHF: 217
     :  HHHHHHHHHHTT S   ------     HHHHTSSTTSHHHHHHHHHHHHHH  SEE
     :  H6              ------     H7        H8               E4

     :  E6                 E7        E8      H6
     : SEEEEEEEE    SS     EEEEE  TT EE  S   HHHHHHHHHHHHHHHHH
 216 :YSCSVLVVYEKDLGERKATNPLVKLVDFAHVVDGNGVIDHNFLGGLCSFIKFLKDIL: 272
       *** * ***          * ********* ****************** *   **
 218 :NSCSILMVYENES----DARPQVKLVDFAHVLDGNGVIDHNFLGGLCSFINFIREIL: 277
     : SEEEEEEEE   ----    EEEEE  TT EE  S   HHHHHHHHHHHHHHHHH
     :  E5         ----    E6        E7      H9
```

b.

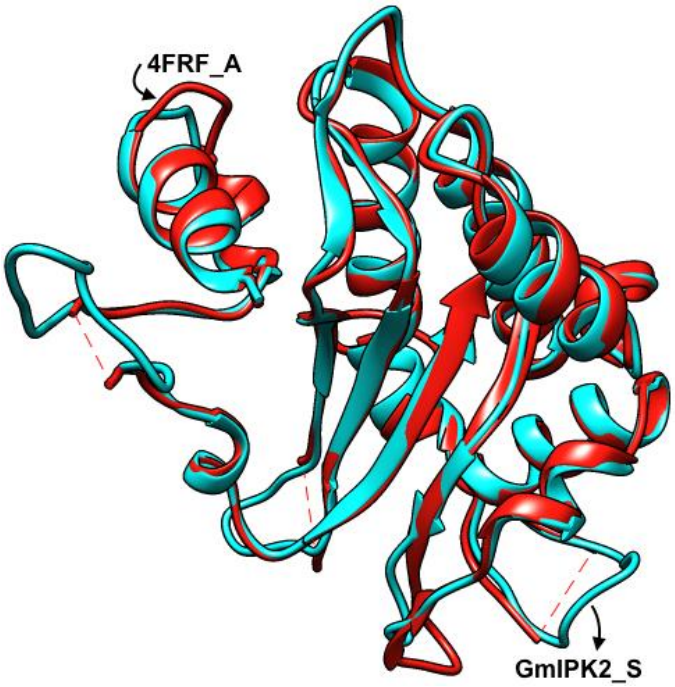
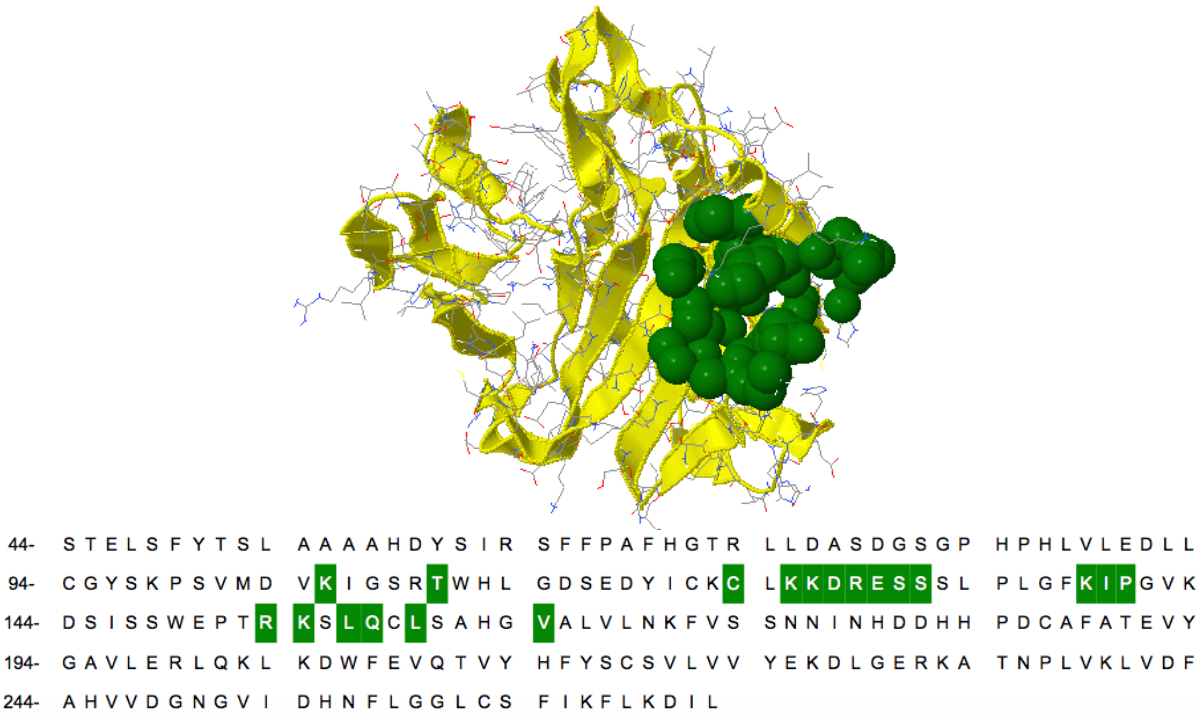
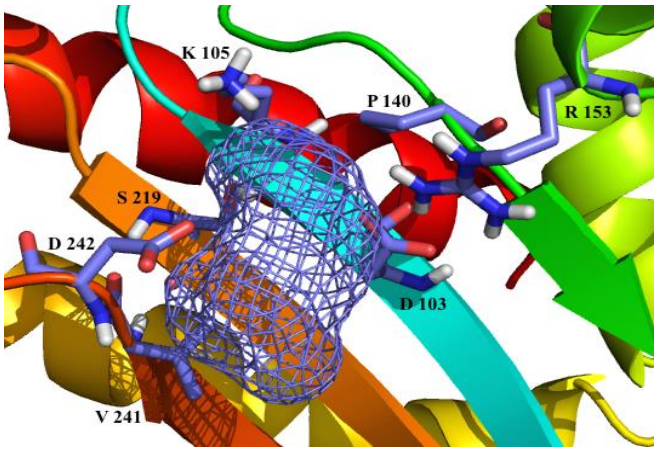


Fig S5 GmIPK2 ligand binding site predicted by online servers **(a)** CASTp (in green) **(b)** FTSite (in purple) **(c)** FunFOLD2 (in grey). The most likely ligand binding residues are marked in each model.

a.



b.



c.

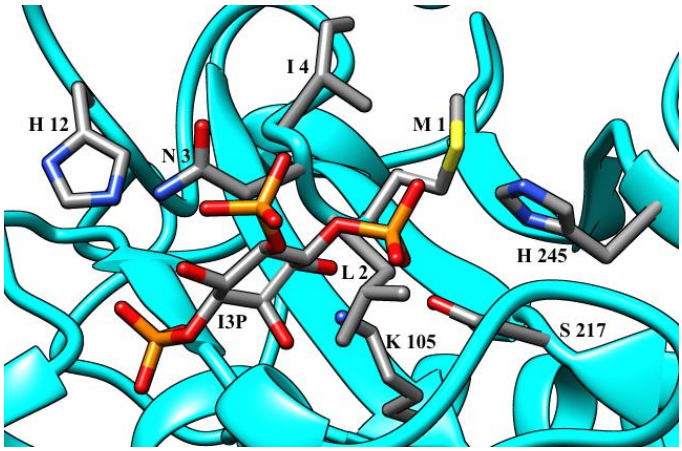
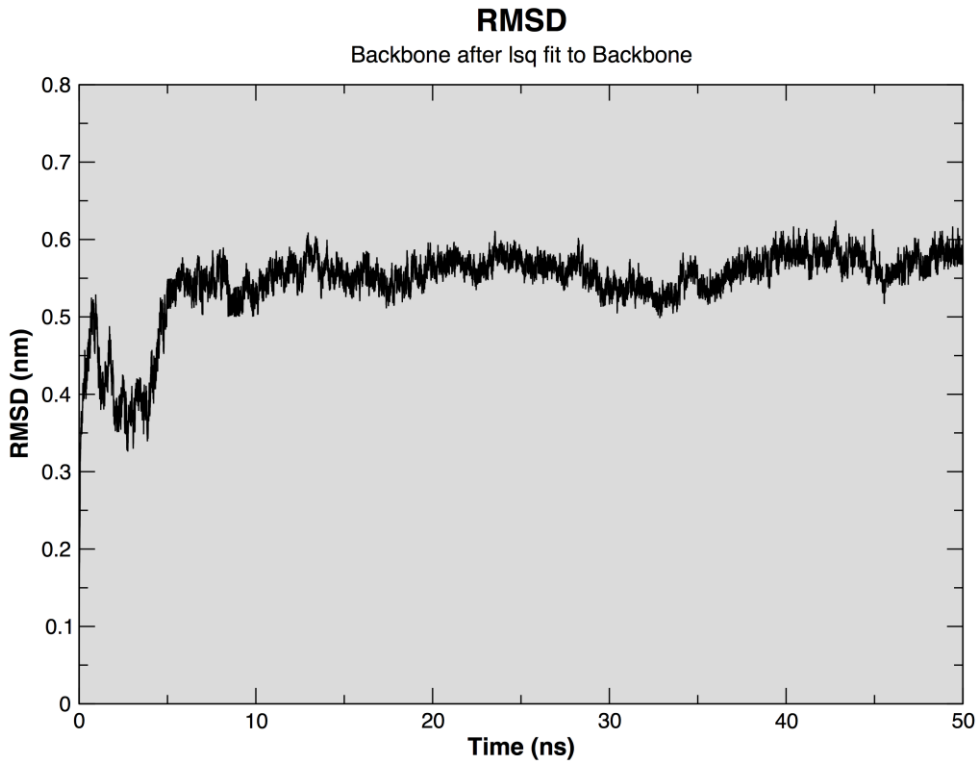


Fig S6 Simulation behavior of GmIPK2 complexed with I3P over 50 ns MD simulation. (a) Root mean square deviations and (b) Root mean square fluctuations.

a.



b.

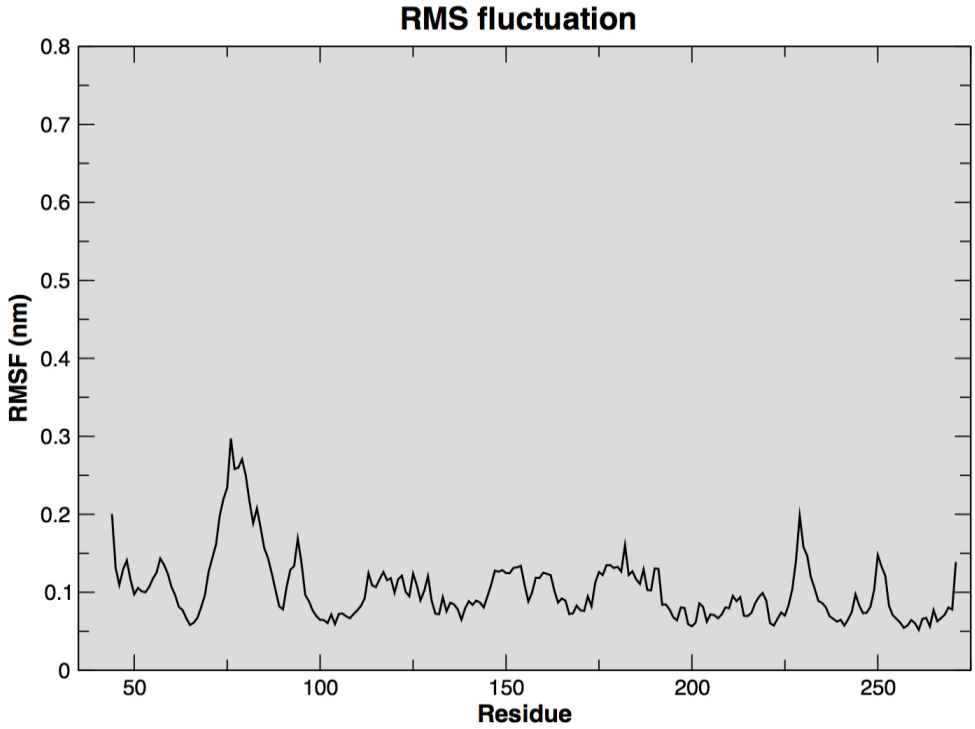
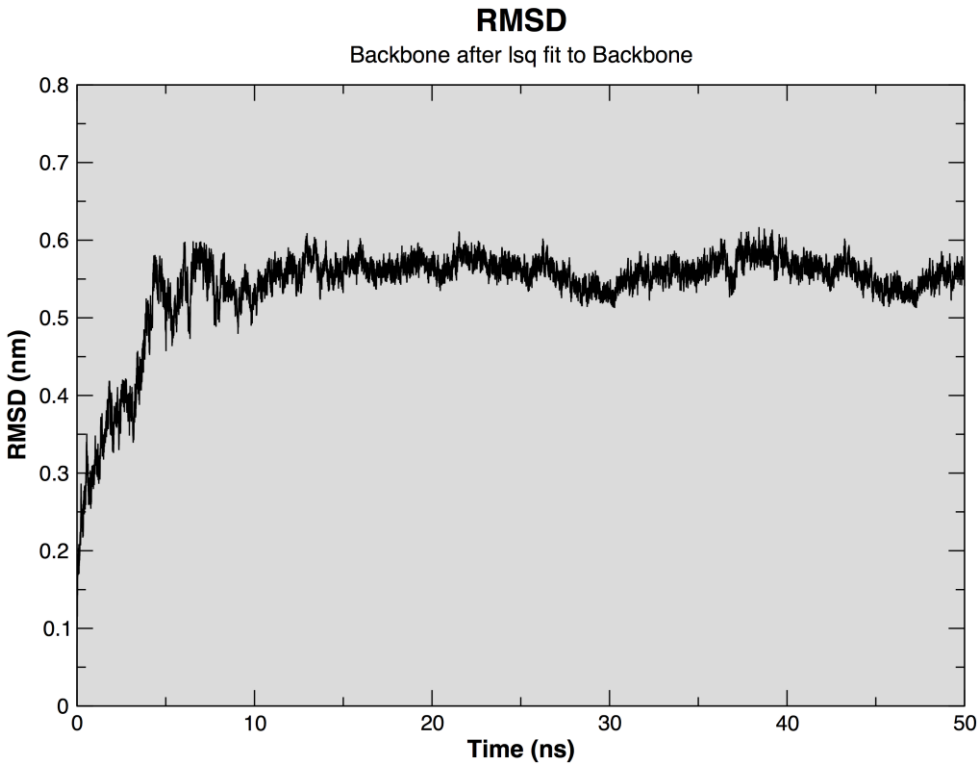
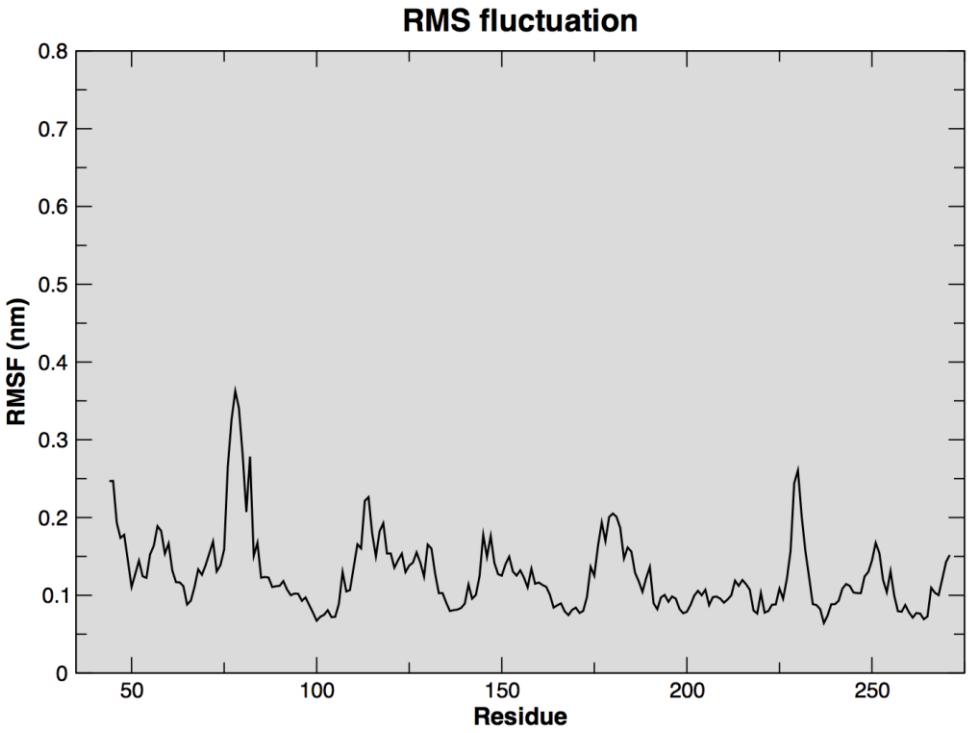


Fig S7 Simulation behavior of GmIPK2 complexed with I0P over 50 ns MD simulation. (a) Root mean square deviations and (b) Root mean square fluctuations.

a.



b.



View publication stats

J. Inst. Eng. India Ser. A
https://doi.org/10.1007/s40030-019-00391-x

ORIGINAL CONTRIBUTION

Optimal Design of Gravity-Fed Sewer Lines Using Linear Programming

Deepak Singh[1] · **P. S. Mahar**[2] · **R. P. Singh**[3]

Received: 9 June 2018 / Accepted: 23 July 2019
© The Institution of Engineers (India) 2019

Abstract The cost of a sewerage system is mainly governed by the size of the sewer pipe, excavation depth and manhole spacing. A linear programming model is developed to minimize the total cost comprising of the pipeline cost, excavation cost and manhole cost of the sewer line. The constraints of the optimization model are related to the distance between two consecutive manholes, and slope of the sewer line to maintain the self-cleansing velocity. The nonlinearity due to the pipe size is eliminated by considering only those available diameters that satisfy the self-cleansing velocity constraint. The model selects the combination of pipe sizes and slope of the sewer line between different manholes maintaining the self-cleansing velocity, which results in the minimum value of the total cost of the entire sewer line. The application of the developed model is illustrated with the help of an existing design problem, and the results are compared with the available solution using forward recursive dynamic programming. It is found that the linear programming model results in lesser value of the total cost of the sewer line.

Keywords Optimal design · Sewer line · Linear programming · Manholes · Self-cleansing velocity

List of Symbols

A	Area of cross section of sewer pipe
C_e	Excavation cost per unit volume of the sewer system between two manholes
C_m	Manhole cost per unit depth of upstream manhole
C_p	Pipe cost per unit length
C_{ijk}	Cost per meter length of the ith sewer link laid between jth upstream node and kth downstream node
d	Inside diameter of pipe
E_c	Excavation cost
ΔH_{ijk}	Elevation difference between jth upstream node and kth downstream node of ith link
i	Link number
j	Upstream node number for ith link
k	Downstream node number for ith link
L	Length of pipe between two consecutive manholes
L_i	Horizontal length between two manholes connected by link
M_c	Manhole cost
M_i	Manhole no. (i = 1, 2, 3,…)
n	Manning's roughness coefficient
nd(i)	Total number of downstream nodes in the ith link
nl	Total number of links in the sewer line
nu(i)	Total number of upstream nodes of the ith link
P	Wetted perimeter
P_c	Pipe cost
Q	Peak discharge through sewer pipe

✉ Deepak Singh
kumaindeepaksingh@gmail.com

P. S. Mahar
pooransmahar@yahoo.co.in

R. P. Singh
rajprasin1@rediffmail.com

[1] Department of Civil Engineering, Delhi Technological University, New Delhi, Delhi 110042, India

[2] Civil Engineering Department, College of Technology, GBPUA&T Pantnagar, Distt. Udham Singh Nagar, Uttarakhand 263145, India

[3] Irrigation and Drainage Engineering Department, College of Technology, GBPUA&T Pantnagar, Distt. Udham Singh Nagar, Uttarakhand 263145, India

Published online: 01 August 2019

R Hydraulic mean radius of channel
S_0 Longitudinal slope of the sewer pipe between two consecutive manholes
T Total cost of sewer line between two consecutive manholes
V Velocity of flow in the pipe
V_e Volume of excavation of the sewer system between two manholes
X_{ijk} Length of ith sewer link laid between jth upstream node and kth downstream node
Y Depth of upstream manhole for a link
y Depth of flow in the sewer pipe
Z Total cost of the sewer line
α_{ijk} Angle of ith link laid between jth u/s node and kth d/s node
θ Half of angle subtended at the center of the pipe by water surface

Introduction

A sewerage system involves the collection, treatment and disposal of the sewage. The sewage is collected from individual homes and carried to the treatment plant through a network of sewer pipes. The network of sewer pipes may consist of house sewers, main/trunk sewers and out fall sewers. Main sewer lines are generally provided with manholes at suitable interval for facilitating their cleaning and inspection. The sewer pipes are laid at downward gradient to carry sewage under gravity up to the outfall point. The design, construction, modification, operation and maintenance of sewerage systems involve a large capital. The sewer lines being the basic unit of a sewerage system, any saving in its cost due to proper design may affect the overall cost of the sewerage system. In spite of the growing concern for the urban environment and the future expenses and effort involved, conventional design methods for storm sewers fail to explicitly account for the cost interactions of the various components of storm sewer systems [18].

The earliest effort to consider hydraulic design of sewer lines was made by Camp [3]. Since then, a large number of research workers contributed to this subject. Some of the researchers employed heuristic methodologies [4, 5, 9, 10, 15, 17, 21]. Many researchers used optimization methods for design of sewer lines. Dynamic programming approach was used by Argaman et al. [2], Walsh and Brown [31], Mays and Yen [18], Mays and Wenzel [19], Gupta et al. [12] and Kulkarni and Khanna [14], to obtain optimal design of sewer networks. Gupta et al. [13] used Powell's method of conjugate directions to optimize the cost function. Swamee [29] applied the Lagrange multiplier method for minimization of the cost function. Afshar and Rohani [1] proposed a hybrid method by considering two sub-optimization problems.

Dajani et al. [6] addressed the problem using a separable convex programming model. Dajani and Hasit [7] compared three models based on separable-convex and mixed integer programming for the optimization of drainage networks. Dajani et al. [8] minimized the cost of a wastewater collection network using separable convex, dynamic and geometric programming. Swamee and Sharma [30] addressed the optimal design of the sewer system by incorporating commercially available pipe sizes directly in the linear programming model. In the recent past, different multi-objective optimization approaches have been developed mainly to minimize the pollution load in combined sewer water and treatment costs by Rathnayake and Tanyimboh [24–26] and Rathnayhke [22, 23].

Theoretical Considerations

Sewer pipes are designed to carry sewage such that the solid particles remain in suspension to avoid occurrence of clogging due to settlement of particles. The particles are kept in suspension by laying the sewer pipes at suitable gradients to achieve self-cleansing velocity at different possible discharges. A trunk sewer line collects sewage from branched sewer pipes at manholes causing variation of flow in its different sections. This requires use of different pipe sizes and gradients to maintain self-cleansing velocity from section to section. The cost of a trunk sewer line includes the costs of sewer pipes, excavation and manholes. The cost of sewer pipes depends on the size and material of the pipe. The cost of the excavation depends upon the gradient adopted, and the cost of the manhole depends upon the layout of sewer system, size and materials used. The aspects related to hydraulics of flow through sewer pipes and cost estimation are described in the following subsections.

Hydraulics of Sewer Pipe

The flow in a sewer pipe under gravity is considered as open-channel flow for designing the pipe size. The velocity of flow through a sewer pipe can be calculated using the Manning's formula as:

$$V = \frac{1}{n} R^{2/3} S_0^{1/2} \tag{1}$$

where V is the velocity of flow in the channel (m/s), n is the Manning's roughness coefficient, R is the hydraulic mean radius of channel (m) and S_0 is the longitudinal slope of the

sewer pipe between two consecutive manholes. The hydraulic mean radius is calculated as ratio of area of cross section A (m^2) to wetted perimeter, P (m). The value of A can be calculated as [28]:

$$A = \frac{d^2}{8}(2\theta - \mathrm{Sin}2\theta) \tag{2}$$

where d is inside diameter of pipe (m) and 2θ is the angle (radian) subtended at the center of the pipe by water surface as shown in Fig. 1. The value of P can be calculated for any depth of flow y (m) as:

$$P = d\theta. \tag{3}$$

The discharge through a sewer pipe can be calculated as:

$$Q = VA \tag{4}$$

where Q is the peak discharge through the sewer pipe (m^3/s).

Using the value of θ as 2.094395 and for y as $0.75d$, and arranging Eqs. (1), (2), (3) and (4), the value of d can be obtained as:

$$d = \left(3.5184\frac{n}{\sqrt{S_0}}Q\right)^{3/8} \tag{5}$$

The value of V should be such that neither the suspended particles in sewage get silted up nor the pipe material gets scoured out. Therefore, the velocity in the sewer pipe is maintained within a permissible range from 0.6 to 3 m/s [11].

Pipe Cost

The cost of a sewer pipe line can be calculated according to diameter and length as:

$$P_c = C_p L \tag{6}$$

where P_c is the pipe cost (\$), C_p is the pipe cost per unit length (\$/m) that depends upon the diameter and L is the length of pipe between two consecutive manholes (m).

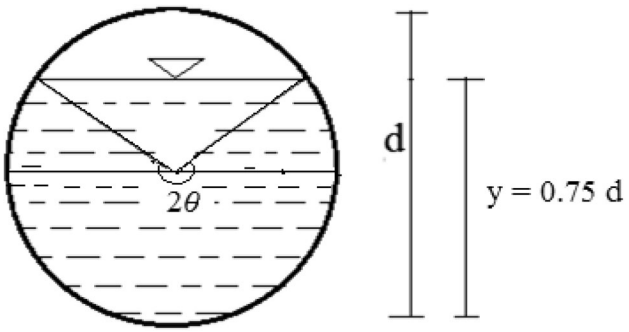


Fig. 1 Circular channel running partially full

Excavation Cost

The excavation cost based on the average depth (sum of pipe diameter and minimum safe excavation cover) and width of excavation can be calculated as:

$$E_c = C_e V_e \tag{7}$$

where E_c is the excavation cost (\$), C_e is the excavation cost per unit volume of the sewer system between two manholes (\$/m^3) and V_e is the volume of excavation of the sewer system between two manholes (m^3).

Manhole Cost

Manholes are generally provided at every bend, junction, change of gradient and change of sewer pipe size. The discharge between different pairs of two consecutive manholes may vary according to the addition of flow from lateral lines at the upstream manhole. Manhole cost of a sewer line between two consecutive manholes can be expressed as:

$$M_c = C_m Y \tag{8}$$

where M_c is the cost of manhole (\$), C_m is the cost per unit depth of upstream manhole (\$/m) and Y is the depth of upstream manhole (m).

Total Cost

The total cost of a sewer line between two consecutive manholes, T (\$), can be expressed by combining Eqs. (6), (7) and (8) as:

$$T = P_c + E_c + M_c \tag{9}$$

Linear Programming Model

A linear programming model is formulated for minimizing the total cost of the sewer line. The sewer line is designed to carry the design discharge satisfying the requirements of the length, velocity and slope in different sections. The sewer line between two consecutive manholes is termed as a link and the manholes as nodes. The lengths of the available discrete pipe sizes satisfying the permissible values of the minimum and maximum velocities in each link are treated as decision variables. For each link, the sewer pipe can be laid at different slopes and accordingly, there will be as many possible diameters as the number of possible slopes to carry the design discharge for the selected pipe material. The preliminary value of diameter corresponding to each possible slope for each link is first calculated using Eq. (5) and then rounded to the next higher available pipe size. If the velocity for the rounded

pipe size is within the permissible range, the diameter is selected as candidate diameter for the link. This process eliminates direct use of nonlinear cost functions in the optimization model.

Objective Function

The objective function of the linear programming model to minimize the total cost of the sewer line can be written as:

$$\text{Minimize } Z = \sum_{i=1}^{\text{nl}} \sum_{j=1}^{\text{nu}(i)} \sum_{k=1}^{\text{nd}(i)} C_{ijk} X_{ijk} \tag{10}$$

where Z is the optimal total cost of the sewer line ($), i is link number, nl is the total number of links in the sewer line, j is upstream node number of the ith link, nu(i) is the total number of upstream nodes of the ith link, k is the downstream node number of the ith link, nd(i) is the total number of downstream nodes of the ith link, C_{ijk} is the total cost per meter length of the ith sewer link laid between the jth and the kth nodes ($/m) and X_{ijk} is the length of the ith sewer link laid between the jth and the kth nodes (m). For each link, the combination of upstream and downstream nodes is chosen in such a way that downstream node is at a lower elevation. For each set of i, j, k, there will be a single diameter for which C_{ijk} is calculated from Eq. (9).

Constraints

The length constraint for each link is imposed by making the horizontal length between the two consecutive manholes equal to the sum of the horizontal equivalent of lengths of different candidate pipe sizes as:

$$\sum_{j=1}^{\text{nu}(i)} \sum_{k=1}^{\text{nd}(i)} X_{ijk} \text{Cos}\alpha_{ijk} = L_i \quad \forall,\ i = 1, nl \tag{11}$$

where L_i is the horizontal length between two manholes connected by link i and α_{ijk} is the angle in degree for X_{ijk} such that:

$$\tan \alpha_{ijk} = \frac{\Delta H_{ijk}}{L_i} \tag{12}$$

where ΔH_{ijk} is the elevation difference between jth upstream node and kth downstream node of the ith link. The non-negativity constraints of the decision variables are included as:

$$X_{ijk} \geq 0 \quad \forall\, i, j, k \tag{13}$$

Design Example

To illustrate the applicability of the proposed linear programming model, a design problem is taken for Goodwin Avenue sewer from Mays and Tung [20]. The data related to the length and peak design inflows for the sewer links named 1, 2 and 3 are given in Table 1. The sewer line along with the ground elevation of manholes and possible elevations of the u/s and d/s ends of different links considered between two consecutive manholes are shown in Fig. 2. The manholes are named M_1, M_2, M_3 and M_4, respectively, from upstream (u/s) to downstream (d/s). The minimum cover depth over the sewer pipe is 1.0668 m (3.5 ft). At each manhole, the crown elevation of the downstream pipe draining out the manhole is equal to or less than the crown elevation of the upstream pipe draining into the manhole. Three possible elevations are considered for both upstream and downstream ends of each sewer link between two consecutive manholes. Thus, total number of possible paths for each sewer link is nine. The difference between two consecutive possible elevations of the sewer link at upstream (or downstream) end is taken as 0.3048 m. The unit costs of excavation and manholes are $ 6.00/yd^3 and $ 100.00/ft depth, respectively. The average excavation width for all pipes is taken as 1.524 m (5 ft). The unit cost of available commercial pipe sizes is given in Table 2. The value of Manning's coefficient is assumed as 0.014 for all pipe sizes.

Table 1 Data related to Goodwin Avenue sewer line [20]

U/S–D/S manholes	Sewer link	Sewer length (m)	Peak inflow (m^3/s)
M_1–M_2	1	70.104	1.0423
M_2–M_3	2	48.768	1.1896
M_3–M_4	3	76.505	1.3312

Solution

The cost coefficients of Eq. (10) for different possible paths of each link of the sewer line are computed in Table 3. The slope for all possible paths of each link is taken using the elevation of the u/s and d/s ends. The values of slope and peak flow rate are used in Eq. (5) to calculate the value of diameter for all possible paths of each link. The pipe diameters are then rounded to just larger available pipe size, and their corresponding costs are taken from Table 2 to obtain the pipe cost from Eq. (6). To compute the excavation cost using Eq. (7), V_e is obtained by multiplying the average width of excavation with the depth of excavation (equal to sum of pipe diameter and minimum safe excavation cover). The computation of total cost of the sewer line using Eq. (9) is given in Table 3. The diameters given in Table 3 for each link are satisfying the permissible velocity limits. Further, the cases of adverse slope (i.e., d/s crown elevation higher than u/s) are considered infeasible as represented by dashed lines in

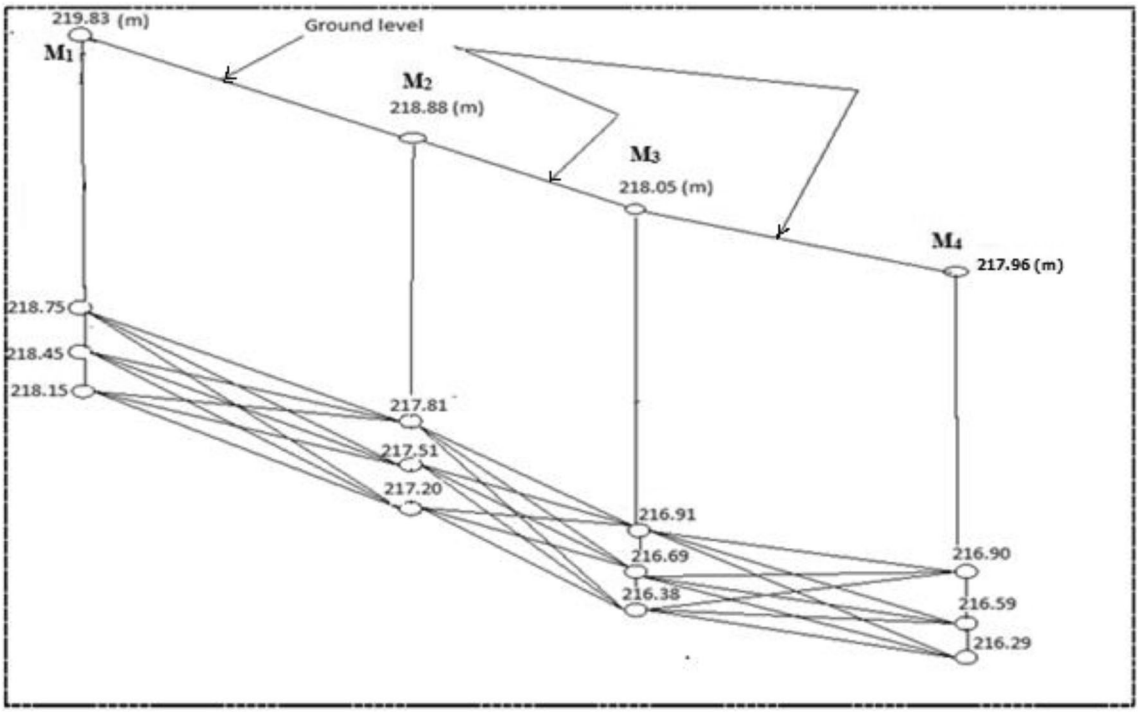


Fig. 2 Schematic diagram showing elevations and possible paths of sewer links

Table 2 Cost of different pipe sizes of sewer line

Pipe size (cm)	30.48	38.1	45.72	53.34	60.96	68.58	76.2	91.44	106.68	121.92	137.16
Cost ($/m)	11.155	14.6	19.357	24.279	30.184	36.254	46.588	62.5	82.021	127.461	129.43

Table 3. Therefore, the cost per unit length of the links corresponding to adverse slope is taken as a large number B in Table 3 using the concept of big-M method [27]. The values of total cost per unit length and slope from Table 3 with B as 2×10^{20} and the lengths from Table 1 are substituted in the linear programming model (Eqs. 10–13) for all the three links resulting in Eqs. (14–20).

$$\begin{aligned}\text{Minimize } Z = {} & 77.0227X_{111} + 67.2390X_{112} + 69.017X_{113} \\ & + 98.77205X_{121} + 82.0906X_{122} \\ & + 72.31086X_{123} + 101.9619X_{131} \\ & + 103.7878X_{132} + 87.1625X_{133} + 80.6871X_{211} \\ & + 70.7543X_{212} + 72.5769X_{213} + 103.3010X_{221} \\ & + 86.3701X_{222} + 76.4373X_{223} \\ & + 129.5241X_{231} + 108.9844X_{232} \\ & + 92.05303X_{233} + 169.0491X_{311} \\ & + 118.5110X_{312} + 98.3365X_{313} \\ & + 2 \times 10^{20}X_{321} + 174.0015\,X_{322} \\ & + 123.4636X_{323} + 2 \times 10^{20}X_{331} \\ & + 2 \times 10^{20}X_{332} + 178.9475X_{333}\end{aligned} \tag{14}$$

subject to:

$$\begin{aligned}& 0.9999X_{111} + 0.9998X_{112} + 0.99989X_{113} + 0.9999X_{121} \\ & + 0.9998X_{122} + 0.9997X_{123} + 0.9999X_{131} + 0.9998X_{132} \\ & + 0.9998X_{133} \\ & = 70.104\end{aligned} \tag{15}$$

$$\begin{aligned}& 0.9998X_{211} + 0.9997X_{212} + 0.9996X_{213} + 0.9999X_{221} \\ & + 0.9998X_{222} + 0.9997X_{223} + 0.9999X_{231} + 0.9999X_{232} \\ & + 0.9998X_{233} \\ & = 49.0728\end{aligned} \tag{16}$$

$$\begin{aligned}& 0.9999X_{311} + 0.9999X_{312} + 0.9999X_{313} + 0.0X_{321} \\ & + 0.9999X_{322} + 0.9999X_{323} + 0.0X_{331} + 0.0X_{332} \\ & + 0.9999X_{333} \\ & = 76.5048\end{aligned} \tag{17}$$

$$X_{111}, X_{112}, X_{113}, X_{121}, X_{122}, X_{123}, X_{131}, X_{132}, X_{133} \geq 0 \tag{18}$$

$$X_{211}, X_{212}, X_{213}, X_{221}, X_{222}, X_{223}, X_{231}, X_{232}, X_{233} \geq 0 \tag{19}$$

$$X_{311}, X_{312}, X_{313}, X_{321}, X_{322}, X_{323}, X_{331}, X_{332}, X_{333} \geq 0 \tag{20}$$

Table 3 Computation of cost of sewer system

Links no.	U/s manhole	D/s manhole	U/s elevation	Crown elevation u/s (m)	D/s elevation	Crown elevation d/s (m)	Slope m/m	Diameter (cm)	Pipe cost (P) ($)	Excavation cost (E) at u/s manhole ($)	U/s manhole cost (M) ($)	Total cost (T) ($)	Cost/ length
I	M_1	M_2	1	218.75	1	217.81	0.01341	76.20	3266.00	1533.60	600.00	5399.60	77.0227
			1	218.75	2	217.51	0.01768	68.58	2541.50	1597.22	575.00	4713.72	67.2390
			1	218.75	3	217.20	0.02211	68.58	2541.50	1735.00	575.00	4841.50	69.0617
			2	218.45	1	217.81	0.00913	91.44	4381.50	1788.89	750.00	6920.39	98.7205
			2	218.45	2	217.51	0.01341	76.20	3266.00	1788.89	700.00	5754.89	82.0906
			2	218.45	3	217.20	0.01768	68.58	2541.50	1852.78	675.00	5069.28	72.3109
			3	218.15	1	217.81	0.00485	91.44	4881.50	1916.66	850.00	7148.16	101.9619
			3	218.15	2	217.51	0.00913	91.44	4381.50	2044.44	850.00	7275.94	103.7878
			3	218.15	3	217.20	0.01355	76.20	3266.00	2044.44	800.00	6110.40	87.1620
II	M_2	M_3	1	217.81	1	216.99	0.01671	76.20	2286.20	1073.33	600.00	3959.53	80.6871
			1	217.81	2	216.69	0.02282	68.58	1779.05	1118.06	575.00	3472.11	70.7543
			1	217.81	3	216.38	0.02914	68.58	1779.05	1207.50	575.00	3561.55	72.5769
			2	217.51	1	216.99	0.01060	91.44	3067.05	1252.22	750.00	5069.27	103.3010
			2	217.51	2	216.69	0.01671	76.20	2286.20	1252.22	700.00	4238.42	86.3701
			2	217.51	3	216.38	0.02303	68.58	1779.05	1296.94	675.00	3750.99	76.4373
			3	217.20	1	216.99	0.00428	106.68	4025.00	1431.11	900.00	6356.11	129.5241
			3	217.20	2	216.69	0.01040	91.44	3067.05	1431.11	850.00	5348.16	108.9844
			3	217.20	3	216.38	0.01671	76.20	2286.20	1431.11	800.00	4517.31	92.05303
III	M_3	M_4	1	216.99	1	216.90	0.00118	137.16	9901.95	2231.12	800.00	12,933.07	169.0491
			1	216.99	2	216.59	0.00523	106.68	6275.00	2091.66	700.00	9066.66	118.5110
			1	216.99	3	216.29	0.00915	91.44	4781.55	2091.66	650.00	7523.21	98.3365
			2	216.69	1	216.90	Adverse	–	–	–	–	–	*B*
			2	216.69	2	216.59	0.00131	137.16	9901.95	2510.00	900.00	13,311.95	174.0015
			2	216.69	3	216.29	0.00523	106.68	6275.00	2370.56	800.00	9445.56	123.4636
			3	216.38	1	216.90	Adverse	–	–	–	–	–	*B*
			3	216.38	2	216.59	Adverse	–	–	–	–	–	*B*
			3	216.38	3	216.29	0.00118	137.16	9901.95	2788.88	1000.00	13,690.33	178.9475

Optimal Design

The optimization model formulated as per Eqs. 14–20 is solved using LINGO 8.0 software [16] in demo mode. Table 4 shows the optimal lengths for the first, second and third links as 70.115 m, 49.086 m and 76.508 m having diameters as 68.52 cm (27 inch), 68.58 cm (27 inch) and 91.44 cm (36 inch) with their corresponding paths as 1–2, 1–2 and 1–3, respectively. The optimal cost of sewer line is obtained as $ 15,712.25. The optimal solution of the same problem obtained by Mays and Tung [20] using the forward recursive dynamic programming optimization technique is also given in Table 4. The comparison of the optimal results shows that the linear programming model resulted in lesser cost of the sewer line with the same values of optimal diameters due to different optimal path obtained for link II.

Effect of Slope on Optimal Results

The developed linear programming model is solved for two additional sets of slopes, set I and set II, as given in Table 5. The set I and set II of slopes are taken by lowering the downstream elevation of each link by 0.1524 m (half a foot) and 0.3048 m (one foot), respectively, from the values for the case shown in Fig. 2. The computations of total cost for set I and set II of slopes of the sewer line are given in Tables 6 and 7, respectively. The optimal solutions for these two sets of slopes using the developed linear programming model are given in Table 8. On comparing the results of Tables 4 and 8, it is observed that the increase in optimal costs for set I and set II of slopes is $ 253.53 and $ 491.64, respectively, showing nonlinear nature of increase. Further, the results show there is no change in optimal diameters and optimal lengths for both the sets of slope whereas there is change in optimal path of link III for the

Table 4 Comparison of results by linear programming and dynamic programming models

S.no.	Optimal results from linear programming model				Optimal results from dynamic programming model			
Link no.	Optimal length (m)	Optimal path	Optimal diameter (cm)	Optimal cost ($)	Optimal length (m)	Optimal path	Optimal diameter (cm)	Optimal cost ($)
I	70.115	1–2	68.58	15,712.25	70.104	1–2	68.58	16,882.35
II	49.086	1–2	68.58		49.073	2–3	68.58	
III	76.508	1–3	91.44		76.104	1–3	91.44	

Table 5 Set of slopes for sensitivity analysis of sewer line design

Set of slopes	Path	Link I		Link II		Link III	
		U/S elevation	D/S elevation	U/S elevation	D/S elevation	U/S elevation	D/S elevation
I	1–1	218.75	217.66	217.66	216.84	216.84	216.90
	1–2	218.75	217.35	217.66	216.53	216.84	216.59
	1–3	218.75	217.05	217.66	216.23	216.84	216.29
	2–1	218.45	217.66	217.35	216.84	216.53	216.90
	2–2	218.45	217.35	217.35	216.53	216.53	216.59
	2–3	218.45	217.05	217.35	216.23	216.53	216.29
	3–1	218.15	217.66	217.05	216.84	216.23	216.90
	3–2	218.15	217.35	217.05	216.53	216.23	216.59
	3–3	218.15	217.05	217.05	216.23	216.23	216.29
II	1–1	218.75	217.51	217.51	216.68	216.68	216.59
	1–2	218.75	217.20	217.51	216.38	216.68	216.29
	1–3	218.75	216.90	217.51	216.07	216.68	215.98
	2–1	218.45	217.51	217.20	216.68	216.38	216.59
	2–2	218.45	217.20	217.20	216.38	216.38	216.29
	2–3	218.45	216.90	217.20	216.07	216.38	215.98
	3–1	218.15	217.51	216.90	216.68	216.07	216.59
	3–2	218.15	217.20	216.90	216.38	216.07	216.29
	3–3	218.15	216.90	216.90	216.07	216.07	215.98

Table 6 Computation of cost of sewer system for set I of slopes

Links no.	U/s manhole	D/s manhole	U/p elevation	Crown elevation u/s (m)	D/s elevation	Crown elevation d/s (m)	Slope m/m	Diameter (cm)	Pipe cost (P) ($)	Excavation cost (E) at u/s manhole ($)	U/s manhole cost (M) ($)	Total cost (T) ($)	Cost/ length
I	M_1	M_2	1	218.75	1	217.66	0.01555	76.20	3266.00	1597.22	600.00	5463.22	77.930
			1	218.75	2	217.35	0.01998	68.58	2541.50	1661.11	575.00	4777.61	68.151
			1	218.75	3	217.05	0.02425	68.58	2541.50	1788.89	575.00	4905.39	69.973
			2	218.45	1	217.66	0.01127	76.20	3266.00	1725.00	700.00	5691.00	81.179
			2	218.45	2	217.35	0.01569	76.20	3266.00	1852.78	700.00	5818.78	83.002
			2	218.45	3	217.05	0.01997	68.58	2541.50	1916.67	675.00	5133.17	73.222
			3	218.15	1	217.66	0.00699	91.44	4881.50	1980.56	850.00	7712.06	110.009
			3	218.15	2	217.35	0.01141	76.20	3266.00	1980.56	850.00	6096.56	86.965
			3	218.15	3	217.05	0.01569	76.20	3266.00	2108.33	800.00	6174.33	88.074
II	M_2	M_3	1	217.66	1	216.84	0.01671	76.20	2286.20	1162.78	650.00	4098.98	83.529
			1	217.66	2	216.53	0.02303	68.58	1779.05	1207.50	625.00	3611.55	73.596
			1	217.66	3	216.23	0.02914	68.58	1779.05	1296.94	625.00	3640.99	74.196
			2	217.35	1	216.84	0.01039	76.20	3067.05	1252.22	750.00	5069.27	103.301
			2	217.35	2	216.53	0.01671	76.20	2286.20	1341.67	750.00	4377.87	89.117
			2	217.35	3	216.23	0.02283	68.58	1779.05	1386.39	725.00	3890.44	79.279
			3	217.05	1	216.84	0.00428	91.44	3067.05	1431.11	900.00	5398.16	110.003
			3	217.05	2	216.53	0.01060	76.20	2286.20	1431.11	850.00	4567.31	93.072
			3	217.05	3	216.23	0.01671	76.20	2286.20	1520.56	850.00	4656.76	94.895
III	M_3	M_4	1	216.84	1	216.90	0.00131	121.92	7743.35	2231.11	800.00	10,774.46	140.834
			1	216.84	2	216.59	0.00523	91.44	4781.55	2091.67	700.00	7573.22	98.991
			1	216.84	3	216.29	0.00928	91.44	4781.55	2231.11	700.00	7712.66	100.813
			2	216.53	1	216.90	adverse	–	–	–	–	–	*B*
			2	216.53	2	216.59	0.00118	121.92	7743.35	2510.00	900.00	11,153.35	145.786
			2	216.53	3	216.29	0.00523	91.44	4781.55	2370.56	750.00	7902.11	103.289
			3	216.23	1	216.90	Adverse	–	–	–	–	–	*B*
			3	216.23	2	216.59	Adverse	–	–	–	–	–	*B*
			3	216.23	3	216.29	0.00131	121.92	7743.35	2788.89	1000.00	11,532.24	150.739

Table 7 Computation of cost of sewer line for set II of slopes

Link no.	U/s manhole	D/s manhole	U/p elevation	Crown elevation u/s (m)	D/s elevation	Crown elevation d/s (m)	Slope m/m	Diameter (cm)	Pipe cost (*P*) ($)	Excavation cost (*E*) at u/s manhole ($)	U/s manhole cost (*M*) ($)	Total cost (*T*) ($)	Cost/ length
I	M_1	M_2	1	218.75	1	217.51	0.01769	68.58	2541.50	1597.26	575.00	4713.52	67.2256
			1	218.75	2	217.20	0.02211	68.58	2541.50	1981.04	575.00	5097.57	72.6968
			1	218.75	3	216.90	0.02625	68.58	2541.50	1853.72	575.00	4970.22	70.8735
			2	218.45	1	217.51	0.01341	76.20	3266.00	1788.05	700.00	5754.05	82.0717
			2	218.45	2	217.20	0.01783	68.58	2541.50	1853.07	675.00	50,769.57	72.3036
			2	218.45	3	216.90	0.02211	68.58	2541.50	1981.036	675.00	5197.54	74.1224
			3	218.15	1	217.51	0.00913	91.44	4381.68	2044.54	850.00	7276.036	103.7849
			3	218.15	2	217.20	0.01355	76.20	3266.00	2044.62	800.00	6110.62	87.1576
			3	218.15	3	216.90	0.01783	68.58	2541.5	2108.66	775.00	5425.16	77.3752
2	M_2	M_3	1	217.51	1	216.69	0.01692	76.20	2286.20	1252.40	700.00	4238.60	86.3611
			1	217.51	2	216.38	0.02303	68.58	1779.05	1297.29	675.00	3751.16	76.4202
			1	217.51	3	216.07	0.029344	68.58	1779.05	1386.98	675.00	3841.03	78.2383
			2	217.20	1	216.68	0.01060	91.44	3067.05	1431.19	850.00	5348.24	108.9788
			2	217.20	2	216.38	0.01671	76.20	2286.20	1431.32	800.00	4517.52	92.0441
			2	217.20	3	216.07	0.02303	68.58	1779.05	1476.23	775.00	4030.28	82.1283
			3	216.90	1	216.68	0.00448	106.68	4025.00	1610.02	1000.00	6635.016	135.1988
			3	216.90	2	216.38	0.01060	91.44	3067.05	1610.09	950.00	5633.14	114.7841
			3	216.90	3	216.07	0.01692	76.20	2286.20	1610.23	900.00	4796.43	97.7268
3	M_3	M_4	1	216.68	1	216.59	0.00118	121.92	7743.35	2370.556	850.00	10,963.906	143.3100
			1	216.68	2	216.29	0.005098	106.68	6275.08	2370.588	900.00	11,013.94	143.9543
			1	216.68	3	215.98	0.00915	91.44	4781.55	2370.66	850.00	8002.21	104.45904
			2	216.38	1	216.59	Adverse	–	–	–	–	–	*B*
			2	216.38	2	216.29	0.00118	121.92	7743.35	2649.44	950.00	11,342.79	148.2625
			2	216.38	3	215.98	0.00523	106.68	6275.08	2649.48	1000.00	11,392.83	148.9064
			3	216.07	1	216.59	Adverse	–	–	–	–	–	*B*
			3	216.07	2	216.29	Adverse	–	–	–	–	–	*B*
			3	216.07	3	215.98	0.00118	121.92	7743.35	2928.33	1050.00	11,721.68	153.2019

Table 8 Optimal solutions for two sets of slopes

Set of slopes	Link	Diameter	Paths	U/S elevation	D/S elevation	Length (m)	Total cost ($)
I	I	68.58	1–2	218.75	217.35	70.12504	15,965.78
	II	68.58	1–2	217.66	216.53	49.08753	
	III	91.44	1–2	216.84	216.59	76.51245	
II	I	68.58	1–2	218.75	217.20	70.118	16,457.42
	II	68.58	1–2	217.51	216.38	49.088	
	III	91.44	1–3	216.68	215.98	76.528	

slopes of set II. However, the optimal cost of the sewer line is increased due to higher excavation cost at larger slope. This indicates that the optimal solution in terms of length and diameter of the sewer line may or may not be affected due to change in slope.

Conclusions

The developed linear programming model can be used to minimize the total cost comprising of pipe cost, excavation cost and manhole cost of a sewer line connecting a series of manholes. The developed model results in optimal values of pipe diameter and slope from different alternatives for carrying the desired discharge and satisfying the constraints of self-cleansing velocity and lengths of sewer line between each pair of successive manholes. The optimal solution of the developed linear programming model is found economical over the optimal solution obtained by forward recursive dynamic programming model. The optimal cost of the sewer line increases nonlinearly with an increase in slope. The developed model can also be used to obtain the optimal range of discharge for an existing sewer line. The developed linear programming model will be useful for wastewater management.

References

1. M.H. Afshar, M. Rohani, Optimal design of sewer networks using cellular automata-based hybrid methods: discrete and continuous approaches. Eng. Optim. **44**(1), 1–22 (2012)
2. Y. Argaman, U. Shamir, E. Spivak, Design of optimal sewer system. J. Environ. Eng. Div. **99**(5), 705–716 (1973)
3. T.R. Camp, Design of sewers to facilitate flow. Sewage Works J. **18**(1), 3–16 (1946)
4. C. Charalambous, A.A. Elimam, Heuristic design of sewer networks. J. Environ. Eng. **116**(6), 1181–1199 (1990)
5. L.A. Cook, B. Lockwood, The investigation of sewer network by computer. Proc. Inst. Civ. Eng. **63**(2), 481–494 (1977)
6. J.S. Dajani, R.S. Gemmel, E.K. Morlok, Optimal design of urban wastewater collection networks. J. Sanit. Eng. Div. **98**(6), 853–867 (1972)
7. J.S. Dajani, Y. Hasit, Capital cost minimization of drainage networks. J. Environ. Eng. Div. **100**(2), 325–337 (1974)
8. J.S. Dajani, Y. Hasit, S.D. McCullers, Mathematical programming in sewer network design. J. Eng. Optim. **3**(1), 27–35 (1977)
9. D.P. Dasher, P.K. Davis, Designing sanitary sewers with microcomputers. J. Environ. Eng. **112**(6), 993–1007 (1986)
10. A.A. Elimam, C. Charalambous, F.H. Ghobrial, Optimum design of large sewer networks. J. Environ. Eng. **115**(6), 1171–1190 (1989)
11. G.G.S. (Greater Grand Sudbury), *Engineering Design Manual: Engineering Services Division, Ontario, Canada* (2012)
12. A. Gupta, S.L. Mehndiratta, P. Khanna, Gravity wastewater collection systems optimization. J. Environ. Eng. **109**(5), 1195–1209 (1983)
13. J.M. Gupta, P. Khanna, S.K. Agarwal, Optimal design of wastewater collection systems. J. Environ. Eng. Div. **102**(5), 1029–1041 (1976)
14. V.S. Kulkarni, P. Khanna, Pumped wastewater collection systems optimization. J. Environ. Eng. **111**(5), 589–601 (1985)
15. J.C. Liebman, A heuristic aid for the design of sewer networks. J. Sanit. Eng. Div. **93**(4), 81–90 (1967)
16. LINDO Systems, Inc., *Demo LINGO, Release 8.0, Chicago* (2003)
17. J.M. Lornic, Sanitary sewer design. Civ. Eng. **52**(6), 66–67 (1982)
18. L.W. Mays, B.C. Yen, Optimal cost design of branched sewer systems. Water Resour. Res. **11**(1), 37–47 (1975)
19. L.W. Mays, H.G. Wenzel Jr., Optimal design of multilevel branching sewer systems. Water Resour. Res. **12**(5), 913–917 (1976)
20. L.W. Mays, Y.-K. Tung, *Hydrosystems Engineering and Management* (McGraw-Hill Inc, New York, 1992)
21. S.W. Miles, J.P. Heaney, Better than "optimal" method for designing drainage systems. J. Water Resour. Plan. Manag. **114**(5), 477–499 (1988)
22. U. Rathnayake, Migrating storms and optimal control of urban sewer networks. Hydrology **2**(4), 230–241 (2015)
23. U. Rathnayke, Enhanced water quality modelling for optimal control of drainage systems under SWMM constraint handling approach. Asian J. Water Environ. Pollut. **12**(2), 81–85 (2015)
24. U.S. Rathnayake, T. Tanyimboh, Integrated optimal control of urban wastewater systems, in *IWA-WCE Conference, Dublin, Republic of Ireland* (2012)
25. U.S. Rathnayake, T.T. Tanyimboh, Multi-objective optimization of urban wastewater systems, in *10th International Conference on Hydroinfomatics, HIC 2012, Hamburg, Germany* (2012)
26. U.S. Rathnayake, T.T. Tanyimboh, Evolutionary multi-objective optimal control of combined sewer overflows. Water Resour. Manag **29**(8), 2715–2731 (2015)
27. A. Ravindran, D.T. Phillips, J.J. Solberg, *Operations Research: Principles and Practice* (Wiley, New York, 1987)

28. K. Subramanya, *Flow in Open Channels* (Tata McGraw Hill Education Private Limited, New Delhi, 2010)
29. P.K. Swamee, Design of sewer line. J. Environ. Eng. **127**(9), 776–781 (2001)
30. P.K. Swamee, A.K. Sharma, Optimal design of a sewer line using linear programming. J. Appl. Math. Model. **37**(6), 4430–4439 (2013)
31. S. Walsh, L.C. Brown, Least cost method for sewer design. J. Environ. Eng. Div. **99**(3), 333–345 (1973)

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

Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Performance and Emission Characteristics of Ternary Fuel Blends in Diesel Engines

To cite this article: Naveen Kumar and Sidharth 2019 *J. Phys.: Conf. Ser.* **1240** 012095

View the article online for updates and enhancements.



IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 14.139.251.100 on 13/08/2019 at 11:31

IOP Conf. Series: Journal of Physics: Conf. Series **1240** (2019) 012095 doi:10.1088/1742-6596/1240/1/012095

Performance and Emission Characteristics of Ternary Fuel Blends in Diesel Engines

Naveen Kumar*[1], Sidharth[1,2]

[1]Centre for Advanced Studies and Research in Automotive Engineering, Delhi Technological University, Main Bawana Road, Delhi, India
[2]Department of Mechanical and Automation Engineering, Maharaja Agrasen Institute of Technology, Rohini Sector-22, Delhi, India

naveenkumardce@gmail.com

Abstract. The unprecedented use of petroleum fuels in recent years has created an imbalance in the supply and demand. Moreover, fast depletion of these fuels, price perturbation and multi-fold increase in greenhouse gas emissions are debatable issues amongst the scientific fraternity. A sustainable solution to the above problems is to shift towards the renewable alternative fuels. In the Indian context, diesel substitute have special relevance. Ethers have some promising properties similar to diesel, out of which Diethyl ethers (DEE) are most promising alternatives. The present investigation deals with the effect of DEE blends on the performance and emission characteristics of unmodified compression ignition (CI) engines. During the present investigation, ternary blends of diethyl ether, jatropha biodiesel (JB) in diesel were prepared in different proportions. The physico-chemical properties and spray characteristics of the test fuels including density, viscosity, calorific value and Sauter mean diameter (SMD) were also calculated. The experimental results reveal that there is no significant effect of the addition of DEE and JB on the brake thermal efficiency. However, the ternary blend carrying 7.5 % of DEE and JB yields the maximum efficiency of 27.9% which is nearly equal to that of peak efficiency of diesel fuel i.e. 28.4%. The exhaust emissions studies reveal that the percentage of CO, HC, NO_X and smoke opacity were reduced significantly. Therefore usage of ternary fuel blends of diesel, JB and DEE is a viable approach for the partial replacement of diesel.

1. Introduction

A global crisis of crude oil and increasing threat to the environment due to emissions from prime movers like internal combustion engines have inhibited the attention of researcher fraternity towards the promising domain of alternative transportation fuels. As compression ignition engine has been used as a commercial prime mover in the Indian economy since a number of decades, so, the researchers are adhering to the field of alternative fuels for this stallion of Indian economy. Energy sources and alternative transportation fuels are a concern of utmost significance not only among the scientific and engineering area but also in framing economic and public policy. Elective energizes and vitality sources give a superb chance to presenting an assortment of science points and expanding understudy enthusiasm for those themes. Science and designing fields are progressively disciplinary - exercises on biodiesel can exhibit that unmistakably, by demonstrating the covering of science,

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

science, and material science in concentrate this and other elective powers. Some of the promising effective alternatives are obtained from biodiesel, alcohols, Di ethyl ether (DEE) [1–7]. These can be used in neat form or with blends.

There are several issues found with biodiesel and vegetable oils in diesel engine [8–10]. However, higher alcohols are gaining limelight for being its greater compatibility in diesel engines. DEE is also a viable alternative because of its lower density and lower viscosity compared to diesel. Moreover, their easy blending and comparable calorific value have made them as a suitable substitute in a diesel engine. Their higher cetane number and higher oxygen content is a very beneficial aspect for its working in a diesel engine. Limited literature is available on the usage of various ternary blends of DEE-diesel-biodiesel in a diesel engine which is an important aspect for this study.

Ali et al. [11] studied the effect of adding DEE (upto 8%) to biodiesel-diesel blended fuel. The authors found better density and acid value with the DEE additive. The authors also observed lower viscosity, cloud point, pour point and calorific value as compared to the blends without DEE.

Ibrahim [12] has experimentally investigated the effect of blending the DEE (upto 15%) with the diesel fuel and found that DEE helped in better engine performance for most of the engine loadings.

Ashok [13] added DEE in the diesel-ethanol emulsion. The author found better performance, combustion and emission of a diesel engine owing to higher cetane number of DEE. Brake thermal efficiency (BTE) of the blends was also found to be comparable to diesel.

Roy et al. [14] used biodiesel-diesel-DEE blends and found lower carbon monoxide (CO) and hydrocarbon (HC) emissions and comparable nitric oxide (NO) and nitrogen dioxide (NO_2) emissions.

Lee et al. [15] has used DEE with diesel and found that the combustion of DEE blended fuels produced lower particulate matter (PM), CO and HC emissions than diesel due to the high oxygen content of DEE. The authors also found a bit higher oxides of nitrogen (NO_X) from the blended fuels.

Tudu et al. [16] studied the effect of DEE in tyre pyrolysis oil-diesel blends. The authors found that the addition of DEE helped in lowering the CO, carbon dioxide (CO_2), HC and smoke emissions with some penalty on NO_X emissions.

Rakopoulos et al. [17] conducted experiments on a diesel engine using blends of diesel fuel with ethanol, butanol and diethyl ether at different volume percentages for emission analysis. The authors found that by increasing the percentage of all biofuels in the blends, a significant reduction of smoke opacity (mainly higher for butanol blend), a reduction of NO_X emissions (mainly higher for DEE blends), as well as a reduction of CO emissions compared to diesel fuel.

Venu et al. [18] used (alcohol-biodiesel-diesel) with DEE as ignition enhancer. The authors found that the test fuels results in better engine performance with minimal emissions. The authors also concluded that the fuel blends can be used directly in the engine as a potential alternative for diesel fuel without any modifications.

Based on the literature it is found that the usage of ternary fuel blends of diesel, biodiesel and DEE is not completely explored. In this study, usage of different blends of diesel, jatropha biodiesel and DEE were formulated. Several physico-chemical properties of the blends were determined. Emission and performance characteristics were determined for analyzing the usage of the blends in a diesel engine

2. Material and Methods

Jatropha biodiesel (JB) is prepared from the potassium hydroxide, methanol, sulphuric acid (analytical grade) purchased from the local vendor in New Delhi. Jatropha seeds were collected from the campus of DTU and the oil was extracted from the seeds with the help of an expeller. The free fatty acid content of the oil was 5.8%. It is converted to biodiesel using two-step transesterification process discussed in [19–21]. Table 1 describes the nomenclature of the test fuels.

Table 1. Nomenclature of the test fuels

S.NO.	BLENDS	DEE(% v/v)	BIODIESEL(%v/v)	DIESEL(%v/v)
1	**D100**	0	0	100
2	**DEE2.5B**	2.5	2.5	95
3	**DEE5B5**	5.0	5.0	90
4	**DEE7.5B**	7.5	7.5	85
5	**DEE10B1**	10	10	80



Figure 1. Test samples

Table 2 describes the equipment's used for determining density, kinematic viscosity calorific value, and Sauter mean diameter of the fuels.

Table 2. Equipment's used for determining different properties

Property	Equipment
Density	Anton Parr DMA 4500
Kinematic Viscosity	Capillary tube
Calorific Value	Parr 6100 Oxygen Bomb
Sauter Mean	Malvern Spraytec

Table 3 shows the properties of the test fuels determined using the equipment.

Table 3. Properties of the test fuels

BLENDS	Density (g/cm^3)	Kinematic viscosity (cSt)	Calorific Value (MJ/Kg)	Sauter Mean diameter (µm)
DEE2.5B2.5	0.82456	2.65	42.57	21.1
DEE5B5	0.8235	2.82	41.94	21.7
DEE7.5B7.5	0.82444	2.99	41.72	22.2
DEE10B10	0.82667	3.14	40.64	22.9
D100	0.825	2.6	44.50	20.8

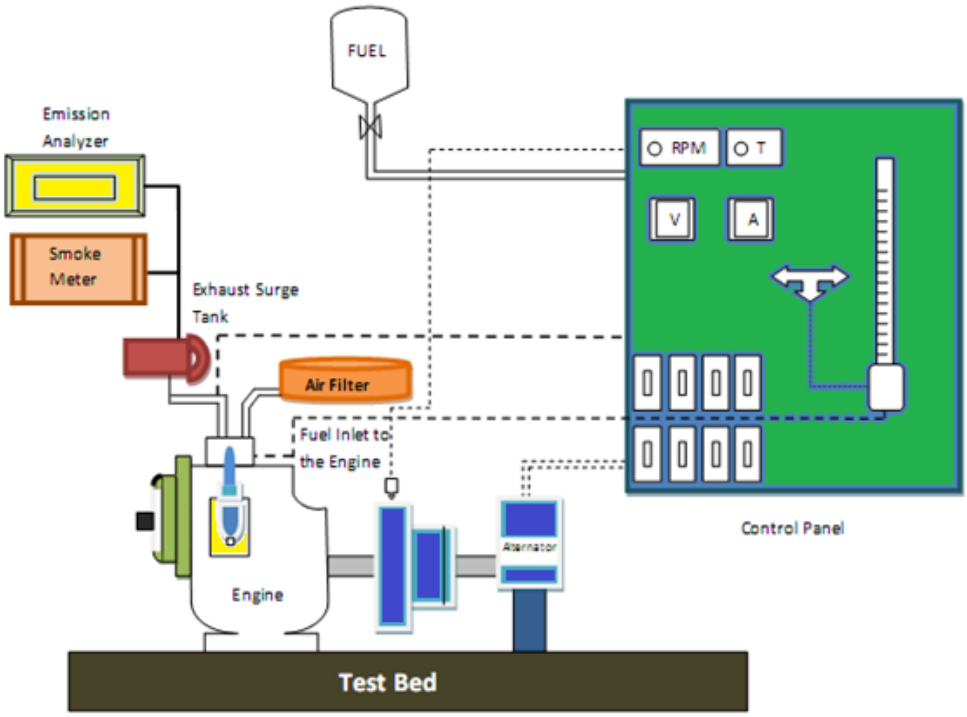


Figure 2. Setup of diesel engine test rig

Table 4 gives about the typical specification of the diesel engine used for the testing. The engine used was a Kirloskar make unmodified air cooled diesel engine.

Table 4. Specifications of the diesel engine

Make	Kirloskar
Model	DAF8
Type	Single cylinder, vertical. DI diesel engine
Speed	1500 rpm, constant speed
Bore (mm)	95
Stroke (mm)	110
Compression Ratio	17.5
Cooling System	Air cooled
Cycle	4 stroke
Maximum output power (KW/HP)	5.9/8.0

All the tests were performed on the developed test rig at varying loads. The results were taken thrice and an average of the three was taken for the better accuracy of the results. Emission tests were carried on AVL Digas analyzer and AVL smoke meter.

3. Performance characteristics

3.1. Brake Thermal Efficiency

From Figure 3 the brake thermal efficiency with BMEP variation at different loading conditions is seen. The efficiency shows an increasing trend to near about 60% load and decreases afterward. BTE decreases sharply after 80% load.

Maximum efficiency of 28.4% is witnessed with 100% diesel at about 60% load and for blends in which diesel is replaced by DEE and Jatropha biodiesel shows a maximum efficiency of 27.9% with the DEE7.5B7.5 blend.
The blend that resulted in most inferior efficiency is for DEE2.5B2.5. In overall prospect, the performance of the engine in terms of BTE is quite comparable to that with diesel, no significant hampering to engine performance is witnessed with ternary blends of DEE and biodiesel in diesel. The most superior fuel among the blends was DEE7.5B7.5 giving highest efficiency among all blends because of its comparable viscosity and energy content. Moreover, added oxygen from DEE and JB is helping in better combustion. DEE10B10 is having slightly lower BTE because of higher viscosity, lower energy content and less Sauter mean diameter (SMD) than DEE7.5B7.5. All other test fuels have lower BTE owing to the lower oxygen content of the fuels. DEE also has higher cetane number which also promotes better combustion and hence shall help in achieving better BTE. SMD of the blends were comparable to diesel which exhibits similar spray characteristics and helps in promoting combustion.

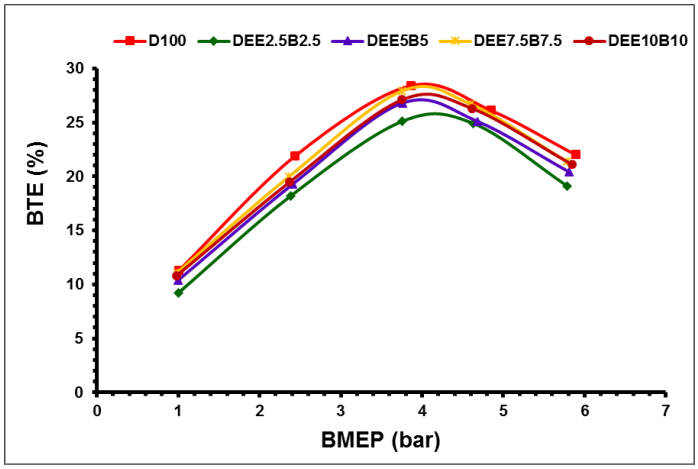


Figure 3. BTE vs BMEP

3.2. Brake Specific Energy Consumption

From Figure 4 the brake specific energy consumption (BSEC) variation at different loading conditions is seen. BSEC in general for all blends and diesel as well decreases sharply with increasing load. Maximum BSEC is observed for DEE2.5B2.5 because of lowest BTE among all prepared test fuels. It is observed that on adding DEE and JB to diesel, the BSEC values decrease. However, for DEE10JB10, owing to low energy content and high viscosity, BSEC is lower as compared to DEE7.5B7.5. Calorific value of DEE and JB is quite lower as compared to petroleum diesel. It therefore needs more energy input for producing the same efficiency. Therefore, BSEC of the blends are higher as compared to diesel fuel.

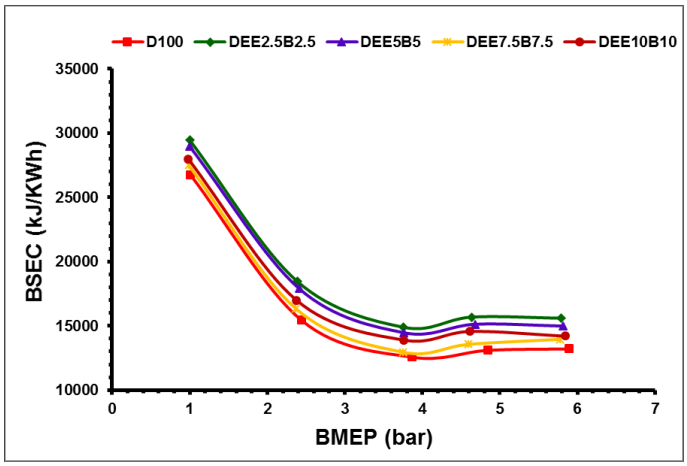


Figure 4. BSEC vs BMEP

4. Emission Characteristics

4.1. Exhaust Gas Temperature

From Figure 5 the exhaust gas temperature variation at different loading conditions is plotted against BMEP. Exhaust temperature shows an increasing trend with the load. Maximum exhaust temperature reached is with diesel fuel i.e. 332˚C at full load. For the blends, DEE2.5B2.5 shows a maximum exhaust temperature. Greater the percentage of DEE in the blends lower is the exhaust gas temperature. DEE helps in lowering the temperature of the blends due to quenching effect produced in the cylinder. Diesel shows overall higher exhaust temperature as compared to ternary blends for respective loads and BMEP.

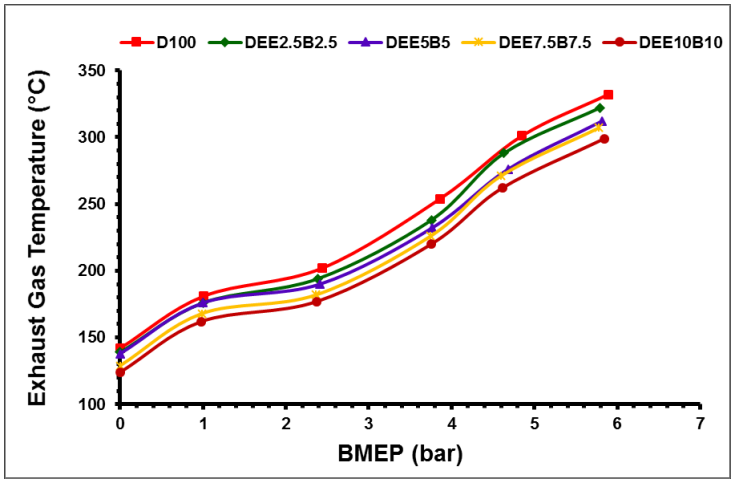


Figure 5. Exhaust gas temperature vs BMEP

4.2. CO Emissions

From Figure 6 shows the CO emissions variation at different loading conditions is plotted against BMEP. CO emissions are steady till 60% load conditions after 60% load CO emissions rises sharply. CO emissions for diesel fuel is highest because of the absence of oxygen. The ternary fuel blends prepared were having lower CO emissions owing to the presence of oxygenated compounds. Moreover, better combustion of the ternary fuel blends was observed. It can be because of the addition of DEE and JB which have better cetane number than diesel which shall exhibit better combustion and therefore lower CO emissions. However, DEE10B10 has slightly higher CO emissions then DEE7.5B7.5 because of its higher viscosity, lower heating values and higher SMD of the DEE10B10 blend as compared to DEE7.5B7.5.

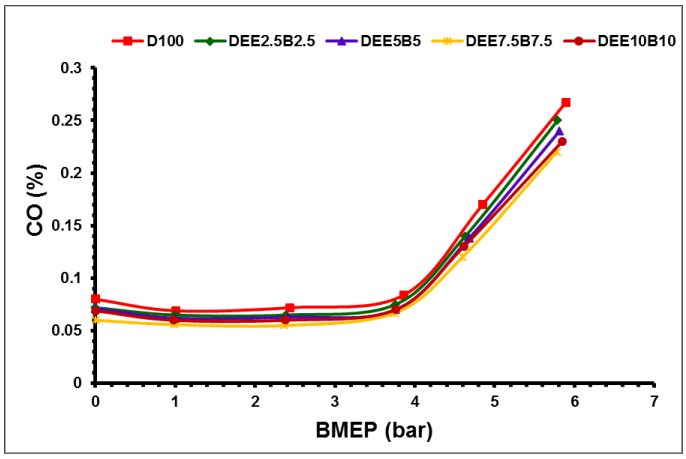


Figure 6. CO vs BMEP

4.3. HC Emissions

From Figure 7 the HC emission variation at different loading conditions is plotted against BMEP. HC emissions increase as the load increases for all the fuel taken into consideration for engine trials. DEE2.5B2.5 shows lower HC emission characteristics at part load up to 40% load, after that emissions increases sharply. HC emission for DEE10B10 is higher at all loads as compared to other blends. Owing to lower energy content, higher viscosity and higher SMD of DEE10B10 blend, more HC emissions were liberated in the atmosphere. However, HC emissions of diesel fuel are higher than the blends carrying DEE 7.5% or lower because of complete combustion of the blends and more prolonged effect of biodiesel in them.

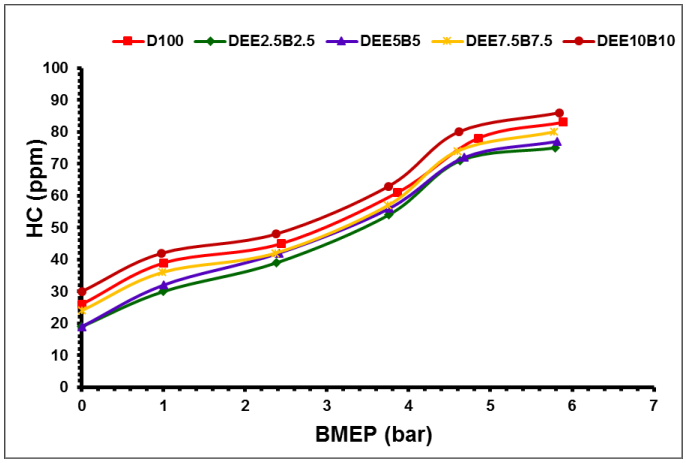


Figure 7. HC vs BMEP

4.4. NO_X Emissions

From Figure 8 the NO_X emissions variation at different loading conditions is plotted against BMEP. NO_X emissions for diesel and DEE2.5B2.5 are quite comparable and are on the higher side in comparison to other blends. DEE5B5, DEE7.5B7.5 and DEE10B10 shows lower NO_X emissions. The reason for lower NO_X can be attributed to the greater quantity of DEE in the blends which leads to more quenching effect in the cylinder. Also NO_X depends on the exhaust gas temperature of the engine. Lower the exhaust gas temperature, lower is the NO_X. Therefore DEE helps in cooling the cylinder and lowering the temperature which results in lower thermal NO_X.

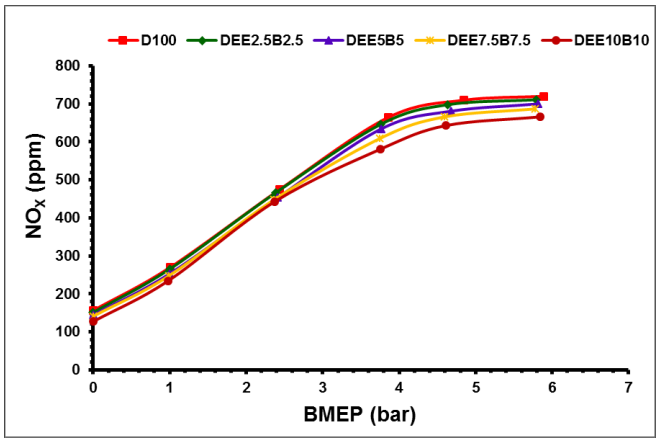


Figure 8. NO_X vs BMEP

4.5. Smoke Opacity

From figure 9 the smoke opacity emissions variation at different loading conditions is plotted against BMEP. Smoke opacity is nearly the same for all fuels tested, however, usage of biodiesel and DEE both having more oxygen content and more cetane number leads to proper burning and hence produces lower smoke emissions. Therefore DEE2.5B2.5 and DEE5B5 produce lower smoke emissions than diesel. It is also found that DEE7.5B7.5 produces the lowest smoke because of better burning then DEE10B10 owing to lower viscosity and smaller SMD of DEE7.5B7.5 which leads to proper atomisation of the fuel and produces less smoke when compared to diesel.

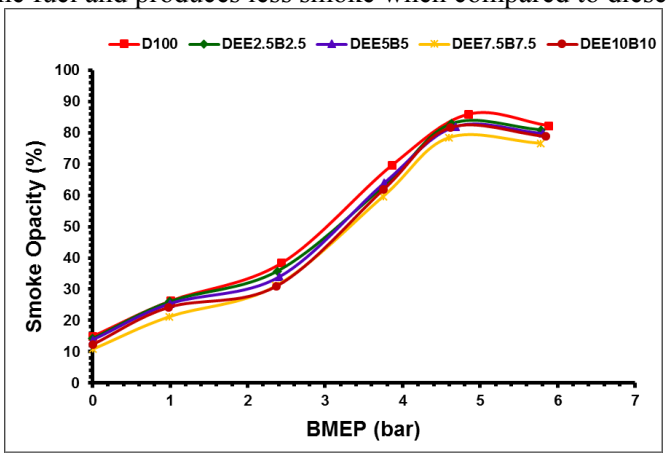


Figure 9. Smoke Opacity vs BMEP

5. Conclusions

The current study focused on using ternary fuel blends of diesel, JB and DEE in varying proportions. The results have shown comparable physico-chemical properties and SMD. An unmodified diesel engine is used for determining the performance and emission characteristics. BTE and BSEC of the test fuels were very similar to diesel which shows that the addition of DEE in diesel and biodiesel gives comparable BTE and BSEC. However, the ternary blend carrying 7.5 % of DEE and JB yields the maximum efficiency of 27.9% which is nearly equal to that of peak efficiency of diesel fuel i.e. 28.4%. Slight lower CO, HC, NO_X and smoke opacity emissions were observed for the blends. However, for DEE10B10 blend, HC emissions were somewhat higher owing to lower energy content, higher viscosity and higher SMD of DEE10B10 blend. The current study has revealed that the ternary blends of diesel, biodiesel and DEE shall be a viable solution for the replacement of diesel because of favorable results.

References

[1] Ma Y, Huang R, Fu J, Huang S and Liu J 2018 Development of a diesel/biodiesel/alcohol (up to n-pentanol) combined mechanism based on reaction pathways analysis methodology *Appl. Energy* **225** 835–47

[2] Suhaimi H, Adam A, Mrwan A G, Abdullah Z, Othman M F, Kamaruzzaman M K and Hagos F Y 2018 Analysis of combustion characteristics, engine performances and emissions of long-chain alcohol-diesel fuel blends *Fuel* **220** 682–91

[3] Emiroğlu A O and Şen M 2018 Combustion, performance and emission characteristics of various alcohol blends in a single cylinder diesel engine *Fuel* **212** 34–40

[4] Yilmaz N and Atmanli A 2017 Experimental evaluation of a diesel engine running on the blends of diesel and pentanol as a next generation higher alcohol *Fuel* **210** 75–82

[5] Sathiyamoorthi R, Sankaranarayanan G and Pitchandi K 2017 Combined effect of nanoemulsion and EGR on combustion and emission characteristics of neat lemongrass oil (LGO)-DEE-diesel blend fuelled diesel engine *Appl. Therm. Eng.* **112** 1421–32

[6] Patil K R and Thipse S S 2015 Experimental investigation of CI engine combustion, performance and emissions in DEE-kerosene-diesel blends of high DEE concentration *Energy Convers. Manag.* **89** 396–408

[7] Kaimal V K and Vijayabalan P 2016 An investigation on the effects of using DEE additive in a di diesel engine fuelled with waste plastic oil *Fuel* **180** 90–6

[8] Agarwal A K, Gupta J G and Dhar A 2017 Potential and challenges for large-scale application of biodiesel in automotive sector *Prog. Energy Combust. Sci.* **61** 113–49

[9] Sorate K A and Bhale P V. 2015 Biodiesel properties and automotive system compatibility issues *Renew. Sustain. Energy Rev.* **41** 777–98

[10] Farfan-Cabrera L I, Pérez-Gonzàlez J and Gallardo-Hernàndez E A 2018 Deterioration of seals of automotive fuel systems upon exposure to straight Jatropha oil and diesel *Renew. Energy* **127** 125–33

[11] Ali O M, Mamat R, Masjuki H H and Abdullah A A 2016 Analysis of blended fuel properties and cycle-to-cycle variation in a diesel engine with a diethyl ether additive *Energy Convers. Manag.* **108** 511–9

[12] Ibrahim A 2016 Investigating the effect of using diethyl ether as a fuel additive on diesel engine performance and combustion *Appl. Therm. Eng.* **107** 853–62

[13] Ashok M P 2012 Identification of best additive using the selected ratio of ethanol-diesel-based emulsified fuel *Int. J. Sustain. Energy* **31** 203–12

[14] Roy M M, Calder J, Wang W, Mangad A and Diniz F C M 2016 Cold start idle emissions from a modern Tier-4 turbo-charged diesel engine fueled with diesel-biodiesel, diesel-biodiesel-ethanol, and diesel-biodiesel-diethyl ether blends *Appl. Energy* **180** 52–65

[15] Lee S and Kim T Y 2017 Performance and emission characteristics of a DI diesel engine operated with diesel/DEE blended fuel *Appl. Therm. Eng.* **121** 454–61

[16] Tudu K, Murugan S and Patel S K 2016 Effect of diethyl ether in a DI diesel engine run on a tyre derived fuel-diesel blend *J. Energy Inst.* **89** 525–35

[17] Rakopoulos D C, Rakopoulos C D, Giakoumis E G, Komninos N P, Kosmadakis G M and Papagiannakis R G 2017 Comparative Evaluation of Ethanol , n -Butanol , and Diethyl Ether Effects as Biofuel Supplements on Combustion Characteristics , Cyclic Variations , and Emissions Balance in Light-Duty Diesel Engine *J. Energy Eng.* **143** 1–8

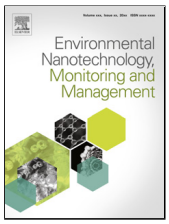
[18] Venu H and Madhavan V 2017 Influence of diethyl ether (DEE) addition in ethanol-biodiesel-diesel (EBD) and methanol-biodiesel-diesel (MBD) blends in a diesel engine *Fuel* **189** 377–90

[19] Kumar N and Sidharth 2018 Some Studies on use of ternary blends of diesel, biodiesel and n-octanol *Energy Sources, Part A Recover. Util. Environ. Eff.* **40** 1721–8

[20] Kumar N and Pali H S 2016 Effects of n-Butanol Blending with Jatropha Methyl Esters on Compression Ignition Engine *Arab. J. Sci. Eng.* **41** 4327–36

[21] Pali H S, Kumar N and Mishra C 2014 Some Experimental Studies on Combustion, Emission and Performance Characteristics of an Agricultural Diesel Engine Fueled with Blends of Kusum Oil Methyl Ester and Diesel *SAE Technical Paper 2014-01-1952 Published*

Accepted Manuscript



Title: PHYTIC ACID DOPED POLYANILINE NANOFIBERS: AN ADVANCED ADSORBENT FOR METHYLENE BLUE DYE

Authors: Monika Duhan, Raminder Kaur

PII: S2215-1532(19)30071-6
DOI: https://doi.org/10.1016/j.enmm.2019.100248
Article Number: 100248

Reference: ENMM 100248

To appear in:

Received date: 25 March 2019
Revised date: 14 July 2019
Accepted date: 16 July 2019

Please cite this article as: Duhan M, Kaur R, PHYTIC ACID DOPED POLYANILINE NANOFIBERS: AN ADVANCED ADSORBENT FOR METHYLENE BLUE DYE, *Environmental Nanotechnology, Monitoring and amp; Management* (2019), https://doi.org/10.1016/j.enmm.2019.100248

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Journal Pre-proof

PHYTIC ACID DOPED POLYANILINE NANOFIBERS: AN ADVANCED ADSORBENT FOR METHYLENE BLUE DYE

Monika Duhan[a] **mduhandeswal@gmail.com**, **Raminder Kaur**[b,*]

rkwalia14@gmail.com

[a]Research Scholar, Department of Applied Chemistry, Delhi Technological University (Formerly Delhi College of Engineering), Delhi-110042, India.

[b]Assistant Professor, Department of Polymer Science and Chemical Technology, Delhi Technological University (Formerly Delhi College of Engineering), Delhi-110042, India.

*Corresponding Author

HIGHLIGHTS

- The phytic acid (PA) doped polyaniline nanofibers (DPANI) has been synthesized and used as an adsorbent for the cationic dye (Methylene Blue).
- The synthesized doped nanofibers were characterized by different techniques to investigate the surface morphology (SEM), atomic scale structure (HRTEM), crystalline or amorphous nature (XRD), functional groups on the surface (FTIR), composition (XPS), surface charge (ZETA Potential) and surface area (BET).
- The effect of contact time and initial MB concentration were investigated and optimized to 60 minutes and 6mg/L respectively.
- Kinetic study of adsorption process revealed that Pseudo-second order model was best fitted to the data and the adsorption occurred through chemisorption.

Journal Pre-proof

Journal Pre-proof

- The isotherm studies carried out using Langmuir model and Freundlich model confirmed that the Freundlich isotherm model fitted best to the adsorption data of the process and the isotherm Langmuir capacity was noted to be 43.4 mg/g.
- The effects of contact time, initial MB concentration, pH and temperature were investigated and optimized to 60 minutes, 6 mg/L, 8.5 and 35^{o}C respectively.
- The Gibb's free energy changes with temperature (-3.466, -3.625 and -3.699 kJ/mol at temperatures 293 K, 303 K and 308 K, respectively). This change in Gibb's free energy indicates that the removal of methylene blue dye by phytic acid doped polyaniline is feasible and spontaneous.

Abstract:

In this paper, the boosted methylene blue (MB) dye adsorption efficiency of polyaniline nanofibers doped with phytic acid (DPANI) is premeditated. DPANI was prepared by simple radical polymerization method using phytic acid (PA) solution. The synthesized doped nanofibers were characterized by different techniques to investigate the surface morphology (SEM), atomic scale structure (HRTEM), crystalline or amorphous nature (XRD), functional groups on the surface (FTIR), composition (XPS), surface charge (ZETA Potential) and surface area (BET). DPANI displayed awfully superior adsorption capacity than undoped polyaniline nanofibers. The presence of the six phosphoric groups on the surface of DPANI is responsible for the enhanced interaction between MB and DPANI. The surface area analyser showed greater surface area of DPANI as compare to conventionally synthesized polyaniline due to the insitu nanofiber formation. The effect of contact time and initial MB concentration were also investigated and optimized to 60 minutes and 6mg/L respectively. Kinetic study of adsorption process was explored by fitting the adsorption data into different kinetic models: Pseudo-first order model, Pseudo-second order model and Intraparticle diffusion model. The

Journal Pre-proof

results of kinetic study revealed that Pseudo-second order model was best fitted to the data and the adsorption occurred through chemisorption. The isotherm studies carried out using Langmuir model and Freundlich model confirmed that the Freundlich isotherm model fitted best to the adsorption data of the process and the isotherm Langmuir capacity was noted to be 43.4 mg/g. The effect of contact time, initial MB concentration, pH and temperature were also investigated and optimized to 60 minutes, 6 mg/L, 8.5 and 35^{o}C respectively. The Gibb's free energy changes with temperature (-3.466, -3.625 and -3.699kJ/mol at temperatures 293 K, 303 K and 308 K, respectively). This change in Gibb's free energy indicates that the removal of methylene blue dye by phytic acid doped polyaniline is feasible and spontaneous. The change in enthalpy was observed to be 1.1kJ/mol, which revealed the existence of physical interaction (physosorption) between the dye molecules and the adsorbent. Surface alteration or doping of polyaniline nanofibers provides an alternate and effective modified adsorbent (DPANI) for wastewater remediation.

Keywords: *Doped Polyaniline, Methylene Blue, Batch Adsorption, Doped Nanofibers, Polyphosphoric Acid.*

1. Introduction:

Since the industrial revolt entered in the human life, discrete things have been changed in the human life style, but it has also affected the nature and ecosystem in different ways. Textile sector is the spot of attention from last few decades, being one of the leading offenders for the deterioration of quality of fresh water bodies. Textile industries employ a large volume of water for distinct processes of dyeing, for good quality fabric production, and release a huge amount of effluent with several pollutants (organic and inorganic substances, dyes, additives

Journal Pre-proof

etc.). The eviction of these coloured compounds or dyes from industrial wastewater is an immensely alarming assignment because these pigments and dyes resist biodegradation and remain in the environment for a long period of time. Hence, establishing a viable method of wastewater management for the textile industries has become an important question for the environmental conservation. The required solution could be achieved by employing a number of techniques such as electrochemical precipitation, membrane separation, adsorption, biological methods, ozonation, oxidation or coagulation (Agrawal and Tratnyek,1996; Han *et.al,*2013;. Sud *et.al,* 2008) etc. From the previously reported literature, adsorption was found to be an effective method for the efficient dye removal from effluent, because, it is convenient, easy to operate, non-destructive, energy-saving, simply designed, economically cheap and affordable (Kannan and Meenakshisundaram, 2001; Wang *et.al,* 2005). In addition, this wastewater treatment technique is also a verdant chemical technology due to its capability to reuse the exhausted adsorbent and retrieve the adsorbate (Lata *et.al,* 2005). As anew, several adsorbents, like nanostructured materials and polymeric compounds (Rao *et.al,* 2007; Kaur and Duhan, 2017) have been utilized for the efficient collection of dyes to remediate the polluted water. Due to the ease of structural modification (doping) of the polymer compound's backbone, there use as adsorbents is growing popular in comparison to conventional adsorbents in wastewater treatment. It is quite interesting to note that the polyaniline (PANI) that was primitively recognized as a conducting polymer (Ali *et.al,* 2006; Ali *et.al,* 2012), is now emerging as a novel adsorbent material with its improved properties such as larger surface area, environmental stability, easy nano-range structure synthesis, modified mechanical properties, low cost, high adsorption capacity, ease of synthesis, controlled electrical properties, exclusive oxidation/reduction and doping/de-doping chemistry. Moreover, nano-range PANI exhibits some specific useful properties as compare to its conventional analogue (e.g., more surface area, added active sites and improved

Journal Pre-proof

conductivity etc.) (Duhan and Kaur, 2019). The molecular structure of PANI, contains nitrogen atom within its backbone to which different functional groups can be attached during doping to modify its surface properties, which may result into enhanced dye removal. Out of the available dopants, phytic acid (PA) can be considered as a good dopant because: i) it is natural (phytates in plant extracts and legumens), low cost, non-polluting and innocuous (Iemma *et.al,* 2008). ii) It can be used as a substituent of HCl for the synthesis of polyaniline nanofibers by oxidation mechanism (Asunskis and Sherwood, 2003). iii) It is a cyclic compound with six phosphate carboxyl groups, which heighten the binding rapport for pollutants (Iemma *et.al,* 2008; Vasca, *et.al,* 2002). Owing to the above properties, the doping of PANI with PA seemed to be more effective for basic dye (cationic dye) removal, because PA acid consolidated as anion to the backbone of PANI structure during *in situ* synthesis (polymerization of aniline) using PA solution. PA doping not only modify the surface properties of PANI nanofibers, but also interact with the dye molecules during adsorptive removal of the same. The reaction between PA and PANI occurred via phosphate carboxyl group preferably at the quinoid site of PANI backbone. Basically, during doping with PA, the nitrogen group of PANI got protonated to form doped PANI. As per the reported literature, one PA molecule can combined minimum with two PANI molecules and act as crosslinker along with a doping agent and form a net-like structure (Gao *et.al,* 2018). Methylene blue is one of the significant dye used in the textile industries; it also has some other applications in different fields like medical science, biology and chemistry etc. Even though it is not very hazardous, some serious diseases (hypertension, dizziness, vomiting, anaemia and nausea etc.) can be induced after its long period exposure.

Already reported literature studies have revealed that PA and PA-salts have been used for the remediation of effluent rich with heavy metals (Cu^{2+}, Cd^{2+}, Zn^{2+}) (Iemma *et.al,* 2008), but till date, no studies have been reported on the dye removal treatment of wastewater by using

Journal Pre-proof

PA doped PANI nanofibers. So, this study, is mainly carried out on the use of PA doped PANI nanofibers for the removal of MB (cationic dye) from its aqueous solution. The synthesized PA doped PANI is characterized by SEM, HRTEM, FTIR, BET, XPS and XRD. UV- visible spectrophotometer was used to investigate the concentration of unadsorbed MB in the residual solution. Isotherm study was carried out to understand the relation between the amount of doped PANI and MB dye by using Langmuir and Freundlich models. The adsorption studies were performed by different adsorption parameters (contact time and initial MB dye concentration). Various kinetic models were used to analyze the mechanism of the reaction. Finally, desorption study was also conducted by using different solutions. The synthesized PA doped PANI nanofibers were found to be novel and efficient adsorbent for MB (cationic dye) removal.

2. Material and Method:

2.1. Materials:

Aniline (CDH, India), Ammonium persulfate (APS)[$(NH_4)_2S_2O_8$] (CDH, India), Phytic Acid (1 mmol, 50 wt % in H2O) (TCI Chemicals, India), Methylene Blue (M.Wt. 319.85 g/mol, λ_{max}: 664 nm) (CDH India) were used as received. All the chemicals used during this work were of AR grade. Distilled water was used to prepare the dye solutions and for washing.

2.2. Synthesis of PA doped PANI nanofibers:

Phytic acid solution is used for the synthesis of PA doped PANI nanofibers by oxidative polymerization of aniline (monomers). In brief, prepare solution-I, by dissolving 0.46mL phytic acid (dopant) and 0.45mL aniline (monomer) in 2mL of distilled water and solution-II, by dissolving 0.285g of APS into 1mL of distilled water. Then add solution-II to solution-I and agitate the mixture for a period of 1 minute. As reaction (polymerization) proceeded, a

Journal Pre-proof

noticeable change in colour of the solution (from brown to dark green) is observed. Then keep the reaction mixture at 4^0C for 24 h undisturbed. Wash the synthesized PA doped PANI nanofibers with distilled water several times and dry the obtained powder in oven at 60^0C for 24 h.

2.3. Characterization of PA doped PANI nanofibers:

Development of PA doped PANI nanofibers was affirmed by ESEM, HRTEM, XRD, FTIR, XPS, ZETA Potential and BET etc. Environmental Scanning Electron Microscope (ESEM) (model FEI Quanta 200F with Oxford-EDS system IE 250 X Max 80) was used to examine the morphology and texture of the resulted doped polyaniline. Size of the synthesized adsorbent was investigated by TEM (Transmission Electron Microscopy) by transferring ethanol dispersed PA doped PANI nanofibers to a copper grid and imaging on F 30 S twin 300 HRTEM at 300 kV. Nicolet 380 FT-IR spectrophotometer was used to obtain the FTIR spectra of synthesized PA doped PANI nanofibers within 400-4000 cm^{-1} region at 25^0 C by preparing KBr pellet. X-ray diffraction (XRD) patterns of PA doped PANI nanofiber was obtained by using Bruker D8 Advance diffractometer by employing CuKα (Kα –1.54056 Å, 40mA, 40 kV) radiation. The scanning step was 0.02^0 in 2θ with a dwell period of 0.5 s per step. Surface area analyser (Micromeritics ASAP 2010) was used to determine the BET surface area of the prepared sample. XPS Analysis was done by using PHI 5000 Versa Probe III to determine the composition of prepared adsorbent. Agilent Technologies Carry Series UV-Vis spectrophotometer was used to study the adsorption of MB dye on PA doped PANI nanofibers within 200 – 800 nm region of wavelength.

Journal Pre-proof

2.4. Determination of pH-ZPC (Zero point charge) of PA doped PANI nanofibers:

Zero point charge (pH-ZPC) of any moiety arbitrates the point at which the adsorbent surface acquires a zero charge density. Salt addition method was used to determine the pH value at which PA doped PANI nanofibers possess zero charge density i.e. pH-ZPC. Briefly, take 10 Erlenmeyer flasks (100mL), each containing 0.01M KNO_3 (10mL) solution. Alter the pH of all solutions from 2 to 12 by using either NaOH (0.1M) or HCl (0.1M) solution. Consider these pH values as initial pH of each flask (ipH). Now, add 0.2 g of PA doped PANI nanofibers into each KNO_3 containing flask and agitate in a shaker incubator for 1h. Keep the reaction mixture for 48 h at room temperature just to equilibrate the mixture. After 48 h, note down the pH of each solution (using pH meter) again and consider it as final pH i.e. fpH. Then calculate the value of ΔpH (ipH - fpH). Finally, calculate the value of ZPC by plotting a graph between ΔpH vs ipH. The ZPC of adsorbent was calculated from the plot between ΔpH versus pHi (Muhammad and Hussain, 2012).

2.5. Batch Adsorption Studies:

An emphatic amount of MB dye was dissolved in distilled water to prepare the stock solution for the experiment. Further, different concentration solutions, as needed, were prepared by diluting the stock solution with distilled water. The adsorption experiment was performed by agitating the appropriate amount of the adsorbent with a definite volume of dye solution in a shaker incubator, at a fixed speed, for an interval of time. Then the reaction mixture was centrifuged at 6000 rpm for 15 minutes. The un-adsorbed dye was estimated by using UV-Visible spectroscope at 664 nm. The standard solutions of methylene blue dye were prepared

Journal Pre-proof

for the calibration of the UV-Visible spectrophotometer. Analogous absorption for individual concentrations was obtained at λ_{max} 664 nm. By plotting the data between concentration and absorbance, a linear regression curve was obtained. The following equations were used to calculate the adsorption capacity and adsorbed percentage of dye onto PA doped PANI nanofibers respectively:

$$Q = \frac{(C_o - C_e)V}{m} \quad (1)$$

$$Percentage\ Adsorption = \frac{C_o - C_e}{m} \times 100\% \quad (2)$$

Where, Q (mg/gm) is the adsorption capacity, C_o is the initial dye concentration (mg/L), C_e is the dye concentration at equilibrium (mg/L), V (L) is the volume of the solution used and m (g) is the mass of PA doped PANI nanofibers used.

2.6. Desorption study of PA doped PANI nanofibers:

Desorption is a crucial step in the adsorption process to make the course money-making. In this study, the MB dye loaded PA doped PANI nanofibers was treated with different solvents (water, 0.1M NaOH, 0.1M HCl and 95% Ethanol) to regenerate the exhausted adsorbent. In brief, add 50 mL of each solvent to four different conical flasks (100mL) and then add 0.03g of MB loaded PA doped PANI nanofibers to each flask and shake for 1h at a speed of 150 rpm in a shaker incubator. Then centrifuge the mixture and analyse the supernatant using UV-Visible spectrophotometer to analyse the recovered dye in the solution.

3. Result and Discussion:

3.1. Synthesis Mechanism:

The radical polymerization method was used for the synthesis of nanofibers of PANI. Basically, the conventional oxidant, APS (radical initiator) was used to initiate the process,

Journal Pre-proof

Journal Pre-proof

by pulling out an electron from each monomer unit (aniline) and form radical cation. This radical cation of aniline can be exists into three resonating structures which combines into a head-to-tail manner to form the PANI Chain (Sapurina and Stejskal, 2008). The secondary amine of PANI backbone, act as a doping site at which different types of ions, molecules and functional groups can be attached to modify the surface of the synthesized PANI nanfibers. In current study, the Phytic acid (PA) used as the dopant, contains six phosphoric groups with hydroxyl and oxygen (PA^-) at terminal sites. These added active sites of PA increase the adsorption capacity of the synthesized doped PANI nanofibers, hence, resulting in more efficient adsorbent for dye removal. A schematic representation of the interaction between the methylene blue dye (MB) and the phytic acid doped polyaniline (DPANI) is as shown in Figure 1.

3.2. Characterization of PA doped PANI Nanofibers:

FTIR spectra of synthesized PA doped PANI nanofibers (Figure 2) explains the molecular structure of the same, between the wavelength range of 4000 cm^{-1} to 500 cm^{-1} (more or less similar to undoped emeraldine salt form of PANI). A characteristic broad absorption band for N-H stretching of the structure was obtained at 3441.4 cm^{-1} (Trchova et al., 2003). The peaks at 1571.8 cm^{-1} and 1496.4 cm^{-1} are attributed to the C=C stretching mode of quinoid and benzenoid ring respectively, indicating that the structure of the polyaniline is in emeraldine form (Furukawa et al., 1988). The absorption peaks at 1230.0 and 1301.5 cm^{-1} are attributed to C=N and C-N stretching vibration (secondary amine) respectively (Furukawa *et al* 1988). The vibration bands at 824.0 and 1145.5 cm^{-1} is attributed to out-of-plane and in- plane bending of C-H bond respectively (Trchova and Stejskal, 2011). The peak at 692.9 cm^{-1} obtained due to out-of-plane deformation of phenylene ring (Trchova and Stejskal, 2011). The vibration peaks at 506.8 cm^{-1} and 754.7 cm^{-1} is due to C-H bond bending vibrations in

Journal Pre-proof

1,4 di-substituted benzene ring (Furukawa, 1988). In addition to these common vibration peaks of emeraldine PANI, some specific peaks at 597.1, 937.2 and 1613.0 cm^{-1} are attributed to the P-O bending vibration, P-O-C symmetrical vibration and P=O stretching vibrations respectively. These peaks indicated that, PANI nanofibers were doped with a consequential amount of PA, without any change in the molecular structure of nanofibers during synthesis.

The nature of PA doped PANI nanofiber (either crystalline or amorphous) and crystallite size was determined by XRD technique (Figure 3).The XRD pattern shows a broad band with peaks at 2θ values of 20^0 and 25^0, which indicated a major amorphous portion of PA doped PANI nanofibers with very little crystalline nature. The crystal size of the prepared PA doped PANI nanofibers was found to be 3.5nm as calculated from Debye Scherer Equation (Eq. 3, below) The literature studies also support the above result (Sasikumar *et.al*, 2015).

$$D = \frac{K\lambda}{(\beta cos\theta)} \qquad (3)$$

Where, β is FWHM i.e. full width half maximum and k is Bragg's constant (0.9), λ is wavelength of X-ray radiation (0.154 nm) and θ is the Bragg's angle.

The result of SEM (Figure 4 a) shows the heterogeneous surface morphology of PA doped PANI nanofibers. This heterogeneous surface morphology of PA doped PANI nanofibers is due to the monomer's unidirectional head-to-tail attachment with each other during polymerization (free radical mechanism). The average diameter of PA doped PANI nanofibers, as prepared in this work, was found to be 60 nm. The synthesized PA doped PANI nanofibers were further analysed by using HRTEM technique (Figure 4 b). HRTEM images confirmed the nano-range structure of doped PANI. In addition, the synthesized

Journal Pre-proof

polyaniline nanofibers were characterized by surface area analyser. As per the results of surface area analyser, the surface area of PA doped PANI nanofibers was found to be higher as compared to conventionally synthesised polyaniline. Table 1 shows the surface area, pore volume and pore diameter of nanostructured and conventional polyaniline. The higher surface area of the synthesized nanostructured doped PANI is quite beneficial for the adsorption of MB.

The ZETA potential measurement of the phytic acid doped polyaniline nanofibers (ethanol dispersion) was carried out by using Malvern Zetameter (Figure 5). It has been noticed that the doped adsorbent possessed negative charge on its surface with negative value of Zeta potential equal to -2.06 mV.

The value of pH-ZPC of PA doped PANI was found to be approximately 7.2 in this study, which indicated that, the surface of the PA doped PANI is positively charged below this pH-ZPC value and negatively charged above it. As, in basic conditions, dihydrogen phosphate (PA) converted to phosphate or hydrogen phosphate (more negative sites), it can be concluded that, more adsorption could be observed at higher pH value, because MB is a cationic dye and get attracts more towards negatively charged moiety.

XPS characterization was done to determine the composition of the synthesised adsorbent (PA doped PANI). The PA doped PANI shows three characteristic peaks (Figure 6 a), which centred at binding energies of 286, 401 and 532 eV, which are attributed to the presence of C1s, N1s and O1s respectively. Figure 6 b shows the presence of phosphorous in the synthesised adsorbent with the clearly observed peak for the P2p, at the binding energy of 135 eV. These results clearly indicated that a substantial amount of phosphoric compound was doped to the polyaniline.

Journal Pre-proof

3.3.Adsorption Studies:

3.3. Adsorption Studies:

3.3.1. Effect of contact time on MB removal:

As the adsorption process is significantly influenced by the contact time, it is quite interesting to study the effect of contact time on the adsorption of dye by varying the time of adsorption from 20-120 minutes (Figure 7 a). The ease of PA doped PANI to remove MB from its aqueous solution was measured by performing the adsorption batch experiments. In brief, 0.08 g of PA doped PANI was added to the 40 mL of MB dye solution. Other adsorption parameters were fixed as pH~7, temperature- 35 ^{o}C, initial MB concentration-7 mg/L and agitation speed-250 rpm. As shown in figure, the removal of MB from its aqueous solution was increased with increase in contact time. It has been observed that about 95% of the dye was adsorbed within first 20 minutes and an equilibrium stage (99% removal) was attained after 60 minutes. So, the optimized equilibrium contact time for the present study was 60 minutes.

3.3.2. Effect of initial dye concentration on MB removal:

The study of the effect of initial dye concentration on adsorption (Figure 7 b) is also a very important factor to understand the adsorption process, which is carried out by varying the initial concentration of MB dye (2-10 mg/L) and keeping all other parameters constant (pH~7,dose- 0.08 g, temp- 35 ^{o}C, and contact time = 60 minutes). It could be seen from the figure that, the adsorption percentage decreases with increase in initial dye concentration. This is due to the availability of more uninhabited active sites on the adsorbent surface in case of lower initial dye concentration. But at higher initial dye concentration, almost complete coverage of active sites has been achieved, hence, a less MB adsorption has been

Journal Pre-proof

observed. On the other hand, the amount of dye adsorbed at equilibrium is increased with an increase in initial dye concentration. This can be explained on the basis of the driving forces provided by the dye molecules that defeats the resistance to the mass transfer (of dye molecules) among solid and aqueous phases. Moreover, there is apparently an increased interaction between dye molecules and the adsorbent with increase in the initial dye concentration. After a specific initial dye concentration i.e. 7 mg/L, no significant change in the adsorption of MB has been observed. So, 7 mg/L value is the equilibrium concentration for the current research work. Some other studies have been reported the similar pattern of results for different dyes such as methyl violet, methyl orange and reactive blue 221 (Chen *et al,* 2010; Karaoglu *et al,* 2010).

3.3.3. Effect of pH on MB adsorption:

The pH of reaction solution in an adsorption process is an important parameter of concern, because the nature of the surface of the adsorbent gets changed with the pH of the solution (Benerjee and Chattopadhaya, 2017). So, the study of the pH variation on the dye removal efficiency of adsorbent is a very important in adsorption study. The effect of solution pH on MB removal by PA doped PANI nanofibers was studied (between pH value 2-11) to optimize the pH value for this study (Figure 7 c).The significant effect of pH on MB adsorption was observed between 6-8 pH value, which also supports the results of pH-ZPC discussed earlier in section 2.4, (pH-ZPC =7.2). In case of PA doped PANI nanofibers, the deprotonation of PA's phosphoric groups occurred, depending upon the pH of the reaction solution. At a medium pH value, some of the phosphoric groups get deprotonated (Hydrogen + dihydrogen phosphates), while at higher pH, more or all phosphoric groups changed to phosphates with

Journal Pre-proof

more negative sites on the adsorbent. Hence, higher MB adsorption is achieved at higher pH value and finally an equilibrium is attained at pH~8.5.

3.3.4. Effect of temperature on MB adsorption

Figure 8 shows the variation of adsorption of methylene blue dye onto phytic acid doped polyaniline with time at different temperatures. The experiment was performed at three different temperatures (293, 303 and 308 K). It has been found that the adsorption percentage increases with increase in temperature. This behaviour is attributed to the increased rate of diffusion of the adsorbate across the external boundary layer and in the internal pores of the adsorbent.

3.3.4. Kinetic Study:

The kinetic study of dye removal by adsorption is important due to the reason of its control over the equilibrium time and the efficiency of the adsorption process. It is also helpful to explain the mechanism of the adsorption process. Three different kinetic models were used to study the rate of dye uptake on the PA doped PANI nanofibers, which are as follows: the pseudo-first order model, the pseudo-second order model and the intra-particle diffusion model.

The following equation represents the pseudo-first order model:

$$\log(Q_e - Q_t) = \log Q_e - \frac{K_1 t}{2.303} \qquad (4)$$

Fitting of adsorption data to the pseudo-first order model was analysed by plotting a graph of log (Q_e-Q_t) vs t relationships as presented in Figure 9 a, where Q_e and Q_t are the amount (mg/g) of MB dye adsorbed on PA doped PANI nanofibers at equilibrium and at time t (min)

Journal Pre-proof

respectively. K_1 (l/min) is the adsorption constant for pseudo-first order model. The slope and the intercept of the graph (Figure 9 a) provided the values of K_1 and Q_e respectively. The values of the above adsorption parameters are given in Table 2. As shown in the table, the value of correlation coefficient is 0.92 for this adsorption process and the values of calculated and experimental Q_e do not coincide with each other. So, it could be concluded that this particular model is not followed by this adsorption system preferably.

Another adsorption model studied for adsorption of MB on PA doped PANI is the pseudo-second order model, which is related with the quantity of dye adsorbed at the equilibrium and on the surface of polyaniline (Figure 9 b). The following equation represents the pseudo-second order model:

$$\frac{t}{Q_t} = \frac{1}{K_2 Q_e^2} + \frac{t}{Q_e} \qquad (5)$$

The plot of t/Q_t vs t is a straight line curve, where Q_t, Q_e and t have their usual meanings and K_2 (g/mg min) is the rate constant for the pseudo-second order model. The values of above parameters are represented in Table 2. As observed, the values of calculated $Q_{e\ cal.}$ (2.92) and experimental $Q_{e\ exp.}$ (2.88) agreed with each other and the R^2 value is reportedly equal to 0.99. So, we can assume that the adsorption data for MB onto PA doped PANI is best fitted in this model. This fitting of adsorption data also indicates that chemisorption played significant role during this experiment.

The intra-particle diffusion model was also studied to understand the adsorption process of MB dye onto PA doped PANI nanofibers. This model assumes that the dye molecules that are adsorbed on the surface of doped PANI nanofibers are diffused into interstitial sites of the adsorbent, but the formation of any bond (physical or chemical) with adsorbent in interstitial sites is a limiting step. This model can be represented by the following equation:

Journal Pre-proof

$$Q_t = K_i t^{0.5} + C \quad (6)$$

Where,K_i (mg/g $min^{0.5}$) is the rate constant for the intra-particle diffusion, C is the boundary layer thickness and Q_t and t have usual meanings. A plot of Q_t against $t^{0.5}$ (Figure 9 c) gave the values of rate constant of and boundary layer thickness from slope and intercept respectively. The values of all the parameters, so obtained, are given in Table 2. But the intra-particle model is not the only model followed by the adsorption of MB on PA doped PANI nanofibers because the best fitted straight line (Figure 9 c) does not pass through the origin. Moreover, the value of R^2 (0.93) also indicated that this model is not preferably followed by the adsorption of MB onto PA doped PANI nanofibers, but plays a significant role in along with other mechanism (s).

3.3.5. Isotherm Study:

Adsorption isotherms are principally used to understand the distribution of the adsorbate over the adsorbent and subsequently for the designing of an adsorption system. The adsorptive behaviour of MB onto PA doped PANI nanofibers have been studied by using two isotherm models: Langmuir and Freundlich isotherm models.

The Langmuir isotherm model is based on the assumption that the adsorption of dye has taken place on a homogenous surface with the formation of a monolayer (Langmuir, 1916). The linear form of Langmuir isotherm is as follows:

$$\frac{C_e}{Q_e} = \frac{1}{bQ_m} + \frac{1}{Q_m} C_e \quad (7)$$

Where, Q_m (mg/g) is the maximal capacity of adsorption analogous to the exhaustive monolayer coverage, b (L/mg) is the Langmuir constant, C_e and Q_e are the dye concentration

Journal Pre-proof

(mg/L) and the adsorbed amount (mg/g) at equilibrium respectively. The liner plot of Langmuir model (Figure 10 a) provides the values of Q_m and b from the slope and the intercept respectively. The values of the all adsorption parameters are tabulated in Table 3. The R^2 value (0.70), undoubtedly indicated that this model is not best fitted to the adsorption data of the MB dye onto PA doped PANI nanofibers. The favourability of the isotherm is also confirmed by the dimensionless factor (R_L), calculated from equation 8 below:

$$R_L = \frac{1}{1+bC_o} \qquad (8)$$

As per the R_L value (0.005), this model shows the feasibility of the adsorption process, but the removal of MB on DPANI cannot be explained by Langmuir's isotherm preferably.

Another isotherm model studied is the Freundlich isotherm model. This model assumes that the adsorption occurs on a heterogeneous surface in spite of a homogeneous surface, and, the active sites on the adsorbent are not equal (Freundlich, 1906). Freundlich isotherm is expressed by the following equation:

$$\log q_e = \log K_f + n_h \log C_e \qquad (9)$$

Where, K_f and n_h are the Freundlich constants and K_f is an indicator of the maximum adsorption capacity; Q_e and C_e have their usual meanings. Figure 10 b shows the $log\ Q_e$vs $logC_e$ relationship for the adsorption of MB onto PA doped PANI nanofibers. This model is seemed to be the best-fitted isotherm model with the value of the regression coefficient, R^2= 0.99. The values of all the parameters are given in Table 3. So, it could be concluded that the multilayer adsorption of MB on the heterogeneous surface of the PA doped PANI has taken place during this adsorption experiment.

Journal Pre-proof

Journal Pre-proof

The PA doped PANI nanofibers have shown a considerably enhanced MB adsorption. The chief focus of this study is to develop a new and novel alternative adsorbent, with better performance characteristics, for the removal of cationic dyes. In the past, different researchers have reported a variety of adsorbents from different precursors such as natural materials, agriculture by-products or the wastes of different industrial products. Table 4 gives a comparison of the different adsorbents used, in past, for the removal of Methylene Blue Dye.

3.3.7. Thermodynamic Studies:

Temperature significantly affects the adsorption capacity of the adsorbent. The effect of temperature on the MB adsorption onto Doped polyaniline nanofibers was studied at 293 K, 303 K and 308 K respectively. The change in free energy (ΔG^o) was calculated using Langmuir constant, b, and, the values of the change in enthalpy (ΔH^o) and entropy (ΔS^o) were evaluated from the slope and intercept of the linear plots of $\Delta G^o/RT$ vs. 1/T (Figure 11), and are given in Table 5. Negative values of ΔG^o indicated feasibility and spontaneity of MB onto doped polyaniline. The ΔH^o value (1.1 kJ/mol) indicated that physical adsorption is also taking place in the system. The positive value of ΔS^o (0.015 kJ/mol) indicated an increase in degree of freedom of the adsorbing system with spontaneity.

3.3.6. Desorption Study:

Desorption study was carried out to make the adsorbent reusable and to make the process cost effective. In this study, the MB dye loaded PA doped PANI nanofibers were treated with different solvents (water, 0.1M NaOH, 0.1M HCl and 95% Ethanol) to regenerate the

Journal Pre-proof

exhausted adsorbent. Among the above solvents, ethanol (95%) was found to be more effective and almost 65% of the dye was recovered. A similar desorption study has also been reported in the literature for the regeneration of the Methyl Orange loaded polyaniline nanofibers (Duhan and Kaur., 2019).

4. Conclusion:

A polymeric based novel adsorbent has been developed for the absorptive removal of cationic (methylene blue) dye. The adsorbent was synthesized by a simple oxidative mechanism (free radical mechanism) using phytic acid solution. The preparation of doped PANI nanofibers was evidenced by different analysis such as: FTIR, XRD, ESEM, HRTEM, XPS and BET etc. UV- visible spectroscopy was used to analyse the un-adsorbed amount of dye. The PA doped PANI nanofibers showed a significantly improved MB removal as compare to its conventional counterpart because of the presence of more active sites on the surface of the concerned adsorbent, due to the phosphoric groups of phytic acid moiety. In addition, the higher MB adsorption at higher pH is also due to the automatic deprotonation of the dihydrogen phosphate groups. This resulted into more negative sites in basic conditions and hence, more MB adsorption occurred. Batch adsorption experiments were used to study the kinetics of the adsorption system. The kinetics of the adsorption of dye was studied using different kinetic models such as Pseudo-first-order, Pseudo-second-order and Intra-particle diffusion models. It has been observed that Pseudo-second-order model was found to be best fitted to the adsorption data of the MB onto PA doped PANI nanofibers, which indicated that the removal of MB dye occurred by chemisorption process. It has been seen that intra-particle diffusion is also playing a significant role in the adsorption along with the chemisorption, as the value of the regression coefficient obtained, for the best fitted straight line for intra-

Journal Pre-proof

particle diffusion model, is significantly high i.e. R^2= 0.93. In addition, both Langmuir and Freundlich isotherm models were also used to analyse the equilibrium data. The best fitting of the data was observed with the Freundlich model. It can be concluded that the adsorptive removal of the MB dye occurred on the heterogeneous surface of the synthesized doped adsorbent and there is a formation of multilayer during this adsorption experiment. The obtained value of isotherm Langmuir capacity (Q_m) was 43.4mg/g. The effect of different adsorption parameters i.e. contact time, initial dye concentration, solution pH and temperature were also studied and the equilibrium values of these parameters are found to be 60 minutes, 7 mg, 7.2 and 35 ^oC respectively. The negative values of Gibbs's free energies at three different temperatures (293 K, 303 K and 308 K) indicated the feasibility and spontaneity of the reaction. Desorption study revealed that almost 65% of MB dye could be recovered by using 95% ethanol as a solvent, from dye loaded PA doped PANI nanofibers. From this study, we concluded that, the surface improvement of PANI nanofibers by doping, resulted in high rates of cationic dye adsorption and can be further explored at commercial scale for dye removal from industrial effluent.

Acknowledgement: The research facilities provided by Delhi Technological University are highly appreciated.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations of Interest: None

Conflict of Interest: The work conducted in this manuscript is for research purpose only. It does not involve any conflict of interest.

Journal Pre-proof

Journal Pre-proof

References:

Agrawal, A., Tratnyek, P.G.,1996. Reduction of nitro aromatic compounds by zero-valent iron metal. Environ. Sci. Technol, 30:153−160.

Al-Bastaki, N., Banat, F.; 2004. Combining ultrafiltration and adsorption on bentonite in a one-step process for the treatment of colored waters. Resour, Conserv. Recy., 41, 103–113.

Ali, V., Kaur, R., Lakshami, G.B.V.S., Kumar, A. Kumari, K., Kumar, S., 2012. Electrical conductivity and dielectric parameters of polyaniline doped with $CuClO_4$.4BN in aqueous DMSO solvent. Adv Polym Tech., 31:374-379.

Ali, V., Kaur, R., Neelkamal, Singh, S., Jain, S.C., Kang, H.P.S., Zulfequer M., Husain M., 2006, Use of Cu^{+1} dopant and its doping effects on polyaniline conducting system in water and tetrahydrofuran. J Phys Chem Solids., 67: 659-664.

Ansari, R., Mohammad, Z.M., Keivani, M.B., Khah, A.M.; 2011. Adsorption of Cationic Dyes from Aqueous Solutions using Polyaniline Conducting Polymer as a Novel Adsorbent. J. Adv. Sci. Res., 2, 27-34.

Asunskis, D.J., Sherwood, P.M.A., 2003. Valence-band x-ray photoelectron spectroscopic studies of vanadium phosphates and the formation of oxide-free phosphate films on metallic vanadium. J. Vac.Sci. Technol. A, 21: 1133−1138.

Benerjee, S, Chattopadhaya, M.C., 2017. Adsorption characteristics for the removal of a toxic dye, tartrazine from aqueous solutions by a low cost agricultural by-product. Arab. Jl. Chem., 10:1629-1632.

Bestani, B., Benderdouche, N., Benstaali, B., Belhakem, M., Addou, A.; 2008. Methylene blue and iodine adsorption onto an activated desert plant. Bioresour. Technol., 99, 8441–8444.

Journal Pre-proof

Chen, S. H., Zhang, J., Zhang, C. L., Yue, Q.Y., Li, Y., Li, C., 2010. Equilibrium and kinetic studies of methyl orange and methyl violet adsorption on activated carbon derived from Phragmitesaustralis. Desalination, 252:149-156.

Duhan, M., Kaur, R., 2019. Adsorptive removal of methyl orange with polyaniline nanofibers: an unconventional adsorbent for water treatment. Environ. Technol., XX: XXX-XXX. (DOI:10.1080/09593330.2019.1593511)

Filho, N.C., Venancio, E.C., Barriquello, M.F., Hechenleitner., A.A., Pineda., E.A.G.; 2007. Methylene blue adsorption onto modified lignin from sugarcane baggasse. Eclét. Quím., 32. 63–70.

Freundlich,H.Z.,1906. Over the Adsorption in Solution. The Jl. of Phys. Chem.,57: 385-470.

Furukawa, Y., Ueda, F., Hyodo Y., Harada I, Nakajima T, Kawagoe T., 1998 Vibrational-Spectra and Structure of Polyaniline. Macromolecules, 21: 1297−1305.

Gao, X., Jing, X., Li, Y., Zhu, J., Zhang, M.,2018. Synthesis and characterization of phosphorized polyaniline doped with phytic acid and its anticorrosion properties for Mg-Li alloy. J. Macromolar Sci. Part A., 55: 24-35.

Ghosh, D., Bhattacharyya, K.G.; 2002. Adsorption of methylene blue on kaolinite. App. Clay Sci., 20, 295–300.

Gürses, A., Karaca, S., Dogar, C., Bayrak, R., Acikyildiz, M.; 2004. Yalcin, M. Determination of adsorptive properties of clay/water system: methylene blue sorption. J. Colloid Interface Sci., 269, 310–314.

Han, Y., Xu, Z., Gao, C.,2013. Ultrathin graphene nanofiltration membrane for water purification. Adv. Funct. Mater. 23: 3693−3700.

Journal Pre-proof

Iemma, F., Cirillo, G., Spizzirri, U.G., Puoci, F., Parisi, O. I., Picci, N., 2008. Removal of metal ions from aqueous solution by chelating polymeric microspheres bearing phytic acid derivatives. Eur. Polym. J., 44: 1183−1190.

Kannan, N., Meenakshisundaram, M., 2001, Adsorption of congo red on various activated carbons-A comparative study. Water, Air, &Soil Pollution, 138: 289-305.

Karaoglu, M.H., Dogan, M., Alkan, M., 2010. Kinetic analysis of reactive blue 221 adsorption on kaolinite. Desalination, 256:154-165.

Kaur, R, Duhan, M., 2017. Adv. Mater. Wastewater Treatment (Martin Scrivener ,Wiley), 51-100.

Langmuir, I.,1916. The constitution and fundamental properties of solids and liquids Part 1. Solids. Jl of Amer. Chem. Soc., 38: 2221-2295.

Lata, S., Singh, P.K., Samadder, S.R., 2015. Regeneration of adsorbents and recovery of heavy metals: a review. Int. J. Environ. Sci. Teh.,12: 1461−1478.

Muhammad, S., Hussain, S.T.,2012. Surface charge properties of zirconium dioxide. Iranian J. Sci. Tech., 4: 481-486.

Pavan, F.A., Gushikem, Y., Mazzoczto, A.C., Dias, S.L.P., Lima, E.C.; 2007. Statistical design of experiments as a tool for optimizing the batch conditions to methylene blue biosorption on yellow passion fruit and mandarin peels. Dyes Pigm., 72, 256–266.

Rao, G.P., Lu, C., Su, F., 2007. Sorption of divalent metal ions from aqueous solution by carbon nanotubes: A review. Sep. Purif. Tech., 58: 224−231.

Sapurina, I., Stejskal, J., 2008. The mechanism of the oxidative polymerization of aniline and the formation of supramolecular polyaniline structures. Polym. Int.,57: 1295−1325.

Sasikumar, Y., Madhan, K.A., Gasem, Z. M., Ebenso, E.E., 2015. Hybrid nanocomposite from aniline and CeO_2 nanoparticles:Surface protection performanceon mild steel in acidic environmemt, Appl. Surf. Sci., 330: 207–215.

Journal Pre-proof

Sud, D., Mahajan, G., Kaur, M.P.; 2008. Agricultural waste material as potential adsorbent for sequestering heavy metal ions from aqueous solutions - A review. Bioresour. Technol., 99: 6017−6027.

Trchova, M, Sapurina, I, Prokes, J, Stejskal, J., 2003 FTIR spectroscopy of ordered polyaniline films. Synth. Met., 135: 305−306.

Trchova, M., Stejskal J., 2011. Polyaniline: The infrared spectroscopy of conducting polymer nanotubes (IUPAC Technical Report). Pure Appl. Chem., 83: 1803−1817.

Vasca, E., Materazzi, S., Caruso, T., Milano, O.,Fontanella, C., Manfredi, C., 2002. Complex formation between phytic acid and divalent metal ions:a solution equilibria and solid state investigation, Anal Bioanal Chem., 374:173-178.

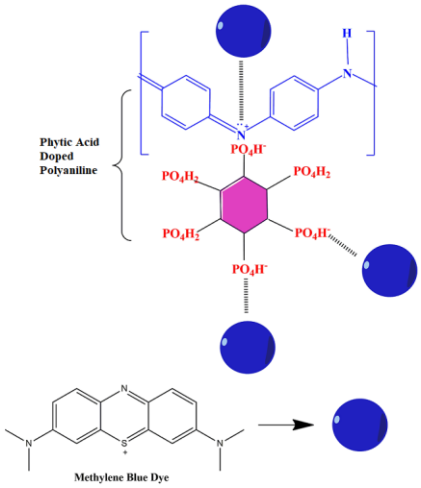
Wang, S., Boyjoo, A., Choueib, A., Zhu, Z.H., 2005. Removal of dyes from aqueous solution using fly ash and red mud. Water Research, 39:129-138.

Journal Pre

Journal Pre-proof

FIGURES CAPTION:

Figure 1: A Schematic Representation of the Interaction between MB dye and Doped Polyaniline



Journal Pre-proof

Journal Pre-proof

Figure 2: FTIR Spectra of Phytic Acid Doped Polyaniline Nanofibers.

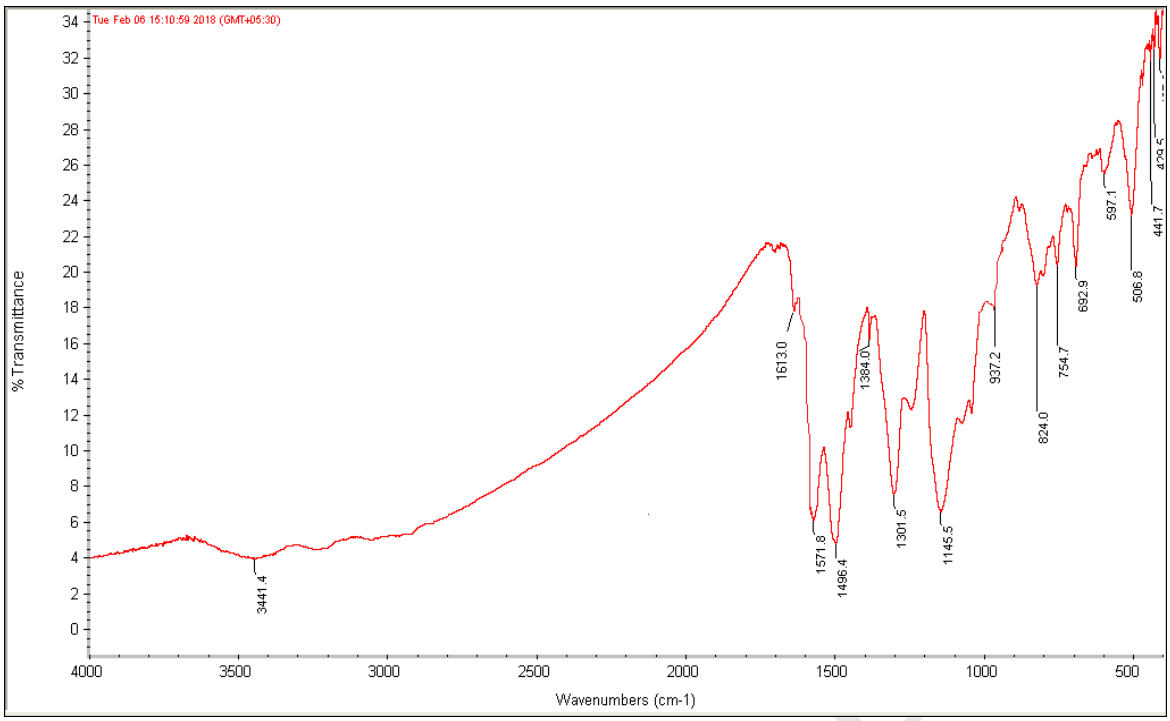
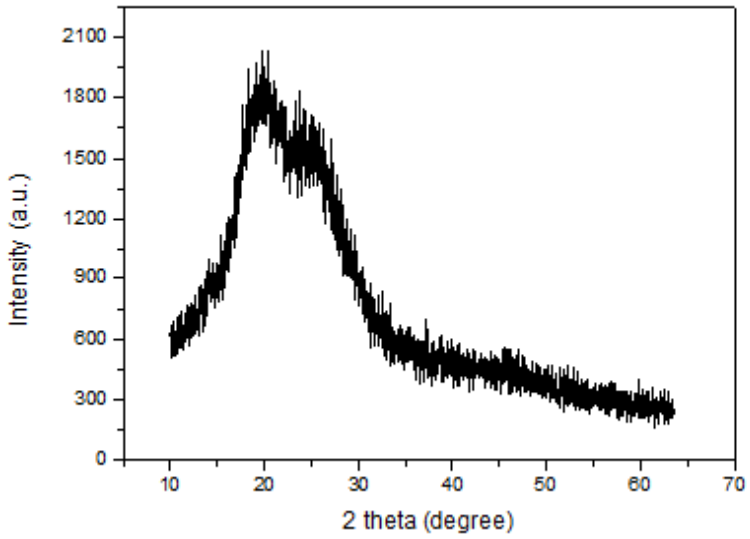


Figure 3: XRD pattern of Phytic Acid Doped Polyaniline Nanofibers.



Journal Pre-proof

Figure 4: (a) SEM image and (b) HRTEM image of phytic acid doped polyaniline nanofibers.

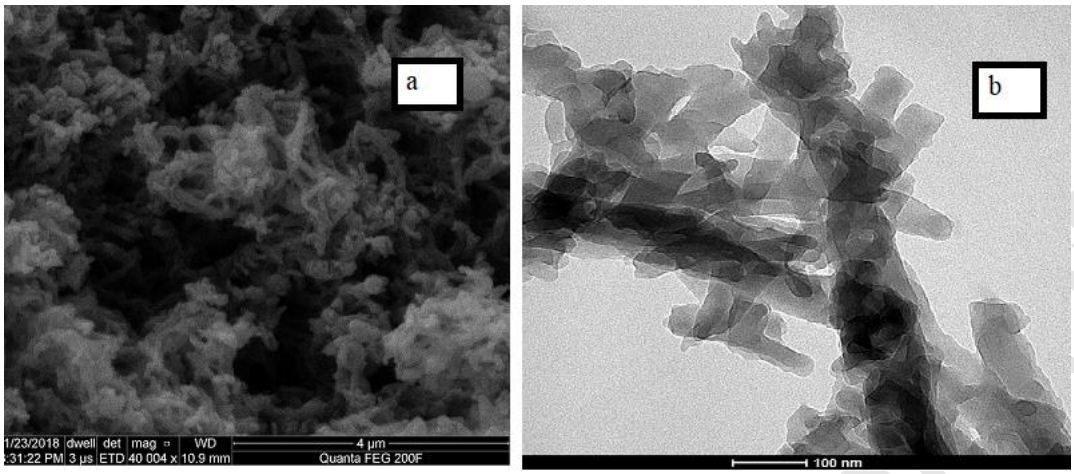
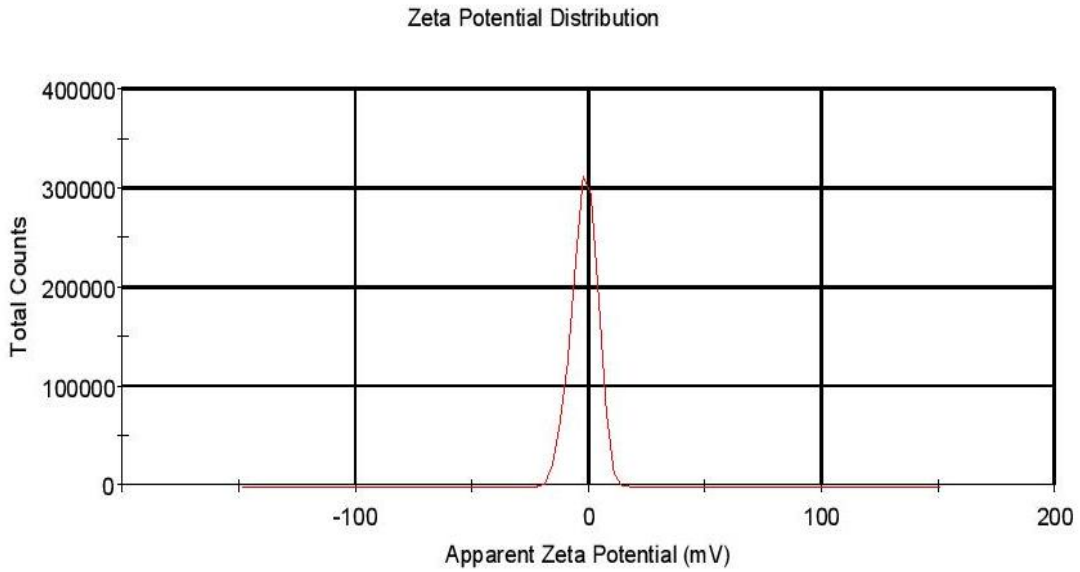


Figure 5: ZETA Potential measurement of Phytic Acid Doped Polyaniline Nanofibers.



Journal Pre-proof

Figure 6: XPS Analysis of Phytic Acid Doped Polyaniline.

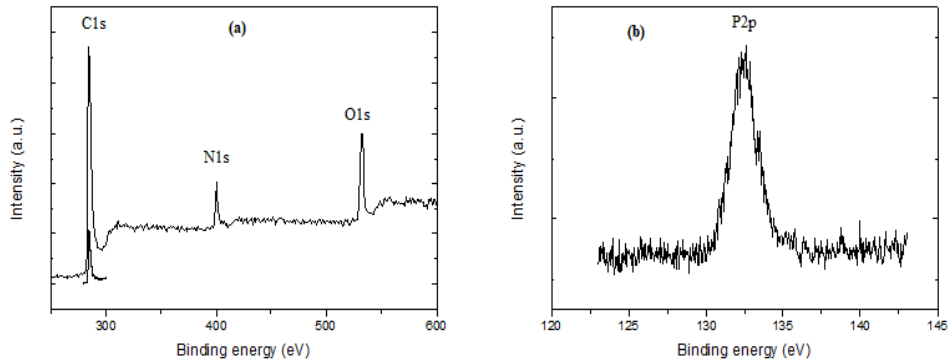
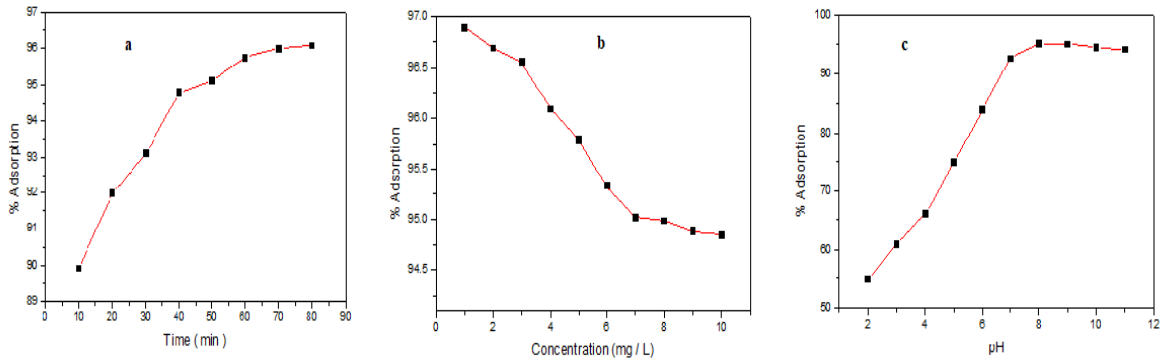


Figure 7: Effects of Adsorption Parameters on Removal of Methylene Blue dye: (a) Contact Time (b) Initial Dye Concentration and (c) Solution pH.



Journal Pre-proof

Figure 8: Variation of methylene blue adsorption with temperature.

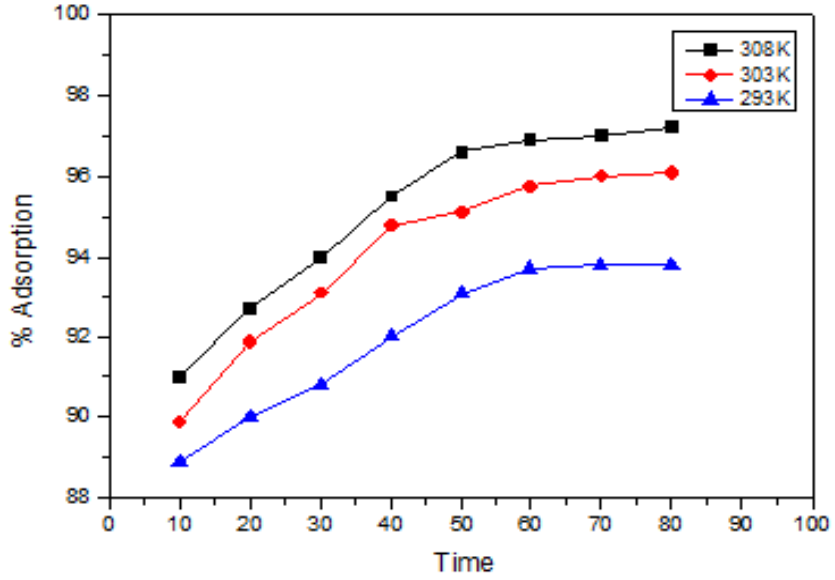
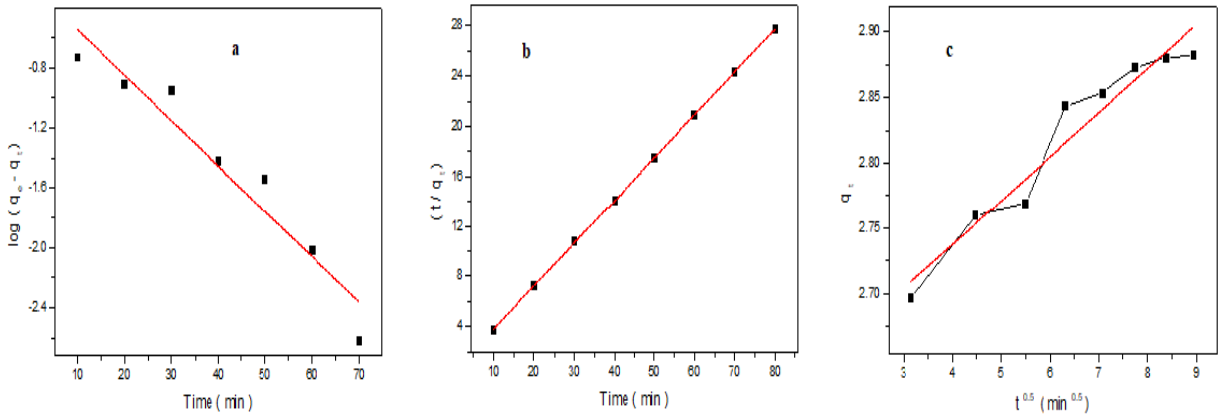


Figure 9: Kinetic Studies: (a) Pseudo-first-order Model; (b) Pseudo-second-order Model and (c) Intra-particle Diffusion Model.



Journal Pre-proof

Figure 10: Isotherm Studies: (a) Langmuir Model (b) Freundlich Model.

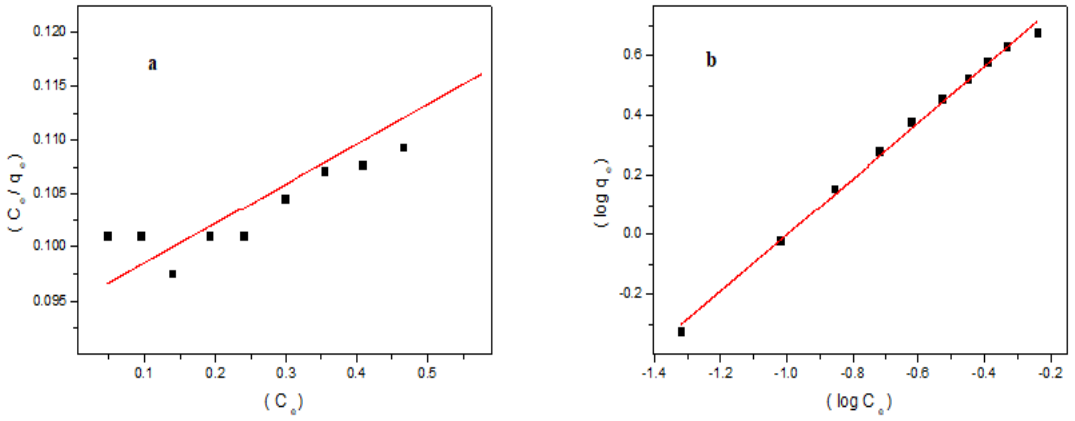
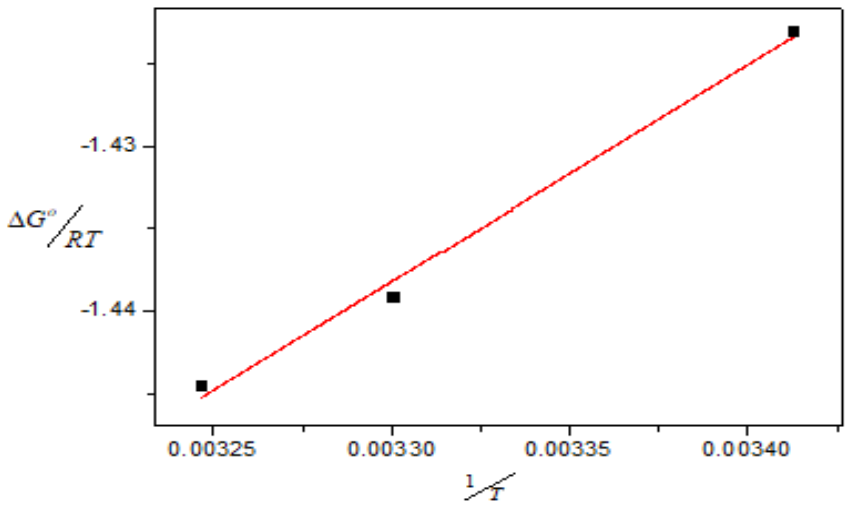


Figure 11: Variation of $\Delta G^{o}/RT$ with 1/T.



Journal Pre-proof

TABLES

Table 1. BET analysis data of the MB adsorption onto Polyaniline nanofibers base (PNB).

Adsorbent	**Surface Area (m^2/g)**	**Pore Volume (cm^3/g)**	**Average Pore Diameter (nm)**
Phytic Acid Doped Polyaniline Nanofibers	33	9.6154×10^{-2}	11.7
Conventional Polyaniline	14	2.638×10^{-3}	34.4

Table 2. Rate Constants Values for Adsorption of Methylene Blue Dye on Phytic Acid Doped PANI Nanofibers for Different Kinetic Models.

Kinetic Models	**Parameters**		**R^2 value**
Pseudo-first-order	$k_1(min^{-1})$	0.048	0.92
	q_e,exp(mg g^{-1})	2.88	
	q_e,cal(mg g^{-1})	0.33	
Pseudo-second-order	k_2(g mg^{-1} min^{-1})	0.047	0.99
	q_e,cal(mg g^{-1})	2.92	
Intraparticle Diffusion	k_i(mg $min^{0.5}$ g^{-1})	0.033	0.93
	C	2.603	

Table 3. Different Adsorption Isotherm Parameters of Methylene Blue onto Phytic Acid Doped PANI Nanofibers.

Isotherms	**Isotherm constant**		**R^2 value**
Langmuir	q_m	43.4	0.70
	b	0.23	
	R_L	0.005	
Freundlich	K_f	10.02	0.99
	n_h	0.99	

Journal Pre-proof

Adsorbent	Adsorption Capacity (mg g^{-1})	Reference
Polyanilline (PANI) film coated Sawdust	11.2	Ansari et al., 2011
Sugarcane Bagasse	34.2	Filho et al., 2007
Yellow Passion Fruit	16	Pavan et al., 2007
Mandarin Peel	15.20	Pavan et al., 2007
Activated Desert Plant	23	Bestani et al., 2008
Clay	6.30	Gurses et al., 2004
Raw Kaolin	13.99	Ghosh and Bhattacharyya., 2002
Bentonite	33	Al-Bastaki and Banat., 2004
Phytic Acid Doped Polyaniline	43	Present work

Table 4: Comparison of Different Adsorbents Used for the Removal of Methylene Blue Dye.

Table 5: Thermodynamic Parameters of Methylene Blue Dye onto Doped Polyaniline.

Temp(K)	ΔG° (kJ/mol)	ΔH° (kJ/mol)	ΔS° (kJ/mol) t
293	-3.466	1.1	0.015
303	-3.625		
308	-3.699		

Journal Pre-proof

Plasma bubble evolution related electron beam parameter estimation in laser wakefield acceleration

M. Yadav[1], S. C. Sharma[1], D.N. Gupta[2]

[1] *Department of Applied Physics, Delhi Technological University (DTU)*

Shahbad Daulatpur, Bawana Road, Delhi 110042, India.

[2]*Department of Physics and Astrophysics, University of Delhi, Delhi-110007, India*

E-mail: raomonika7@gmail.com, suresh321sharma@gmail.com, dngupta@physics.du.ac.in

ABSTRACT

Laser-wakefield acceleration provides a compact electron acceleration scheme for producing a multi-GeV electron beam, utilizing the high longitudinal electric field gradient generated by high intensity, ultrashort laser pulse. The blow-out regime of the laser wakefield acceleration (LWFA) is one of the most recent and reassuring mechanisms for generating quasi-mono-energetic electron beams. In this work, we propose simulation-based studies for plasma bubble evolution and corresponding electron beam quality parameters. The primary focus of this work is to find out the dependency of the electron beam energy and beam-quality on the shape and size of the bubble. The evolution of bubble with time and correlation of bubble length (longitudinal and transverse radius) with the intensity of laser pulse has been revealed in this study. The results show that the longitudinal bubble length grows with time. It has also been confirmed that the shape of the bubble cannot be predicted using fixed shape models as spherical or elliptical. Moreover, the simulation-based findings predict that as the bubble traverses in plasmas, it evolves from spherical shape to the highly elliptical shape. Moreover, as it approaches the dephasing length, the eccentricity decreases further.

INTRODUCTION

Laser Wakefield Acceleration has been widely used as a potential acceleration mechanism due to its ability to yield superior quality mono-energetic electron beams in GeV energy range. The progress in laser technology all over the world has led to the development in the area of LWFA. The Chirped Pulse Amplification (CPA) technique has made Petawatt lasers possible. The PW laser systems can generate very high-intensity lasers (10^{20} to 10^{22} W/cm^2), which further ease the realization of the highly non-linear regime- bubble regime. In this non-linear regime, when a

laser pulse is transmitted in the under dense plasma, the electrons in plasma are completely evacuated due to the ponderomotive force exerted by the laser pulse, creating an electron free region behind it, called blow out or bubble regime[1-3]. The electrons in the background plasma may get trapped at the back of the bubble and get accelerated to high energies due to longitudinal space charge field in the bubble. So, self-injection phenomenon is very advantageous in order to avoid various experimental challenges faced in inducing a charge in the bubble using various techniques.It has already been reported by various researchers that various non-linear processes such as pulse-compression and self-focusing is accountable for the amplification of the laser pulse intensity. The intensity amplification further leads to the lengthening of the bubble, which is responsible for the injection of electrons at the back of the bubble. Hence, understanding the bubble lengthening process is highly desirable to enhance the understanding and control of the self-injection process [4]. It is crucial to control the self-injection process to enhance the electron energy and beam quality in LWFA. The theory for electromagnetic fields in the bubble and its geometry was proposed earlier by few researchers [5]. The bubble is assumed to be spherical. Sadighi-Bonabi *et al.* [6] assumed the bubble shape to be elliptical and found the elliptical cavity model more consistent in explaining the monoenergetic electron trajectory. Benedetti *et al.* [7] proposed a study of the self-injection process, relating the self-injection mechanism and wake properties like wake velocity, wake amplitude, etc. In 2015, Li *et al.* presented the solution for the electromagnetic fields in the non-linear bubble wake and bubble shape, assuming the bubble shape to be elliptical. The effect of residual electrons on the shape of the bubble is also presented in the paper.

SIMULATION RESULTS

In this section,we discuss the evolution of the bubble wake structure while it propagates in plasmas. In order to observe the shape and size of the bubble, 2D-PIC simulations have been performed on plasma simulation VORPAL code. A laser pulse of pulse duration τ = 20 fs, laser spot size r_0 = 12 μm, laser wavelength λ = 1 μm, and power in Petawatt regime is considered. The simulation length of 1.2 mm has been considered. The plasma density considered in the simulations is 10^{19} cm^{-3}. A window of size 100 × 65 μm, moving with the velocity of light has been considered.

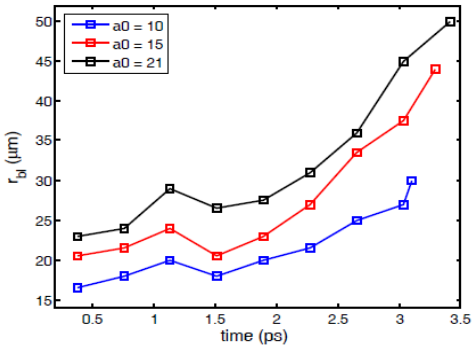


FIG. 1. Variation of longitudinal bubble radius (r_{bl} in micron) with time (in picosecond) for different values of normalized laser strength parameter a_0=10, 15, 21. Other laser-plasma parameters considered are: laser pulse duration τ = 20 fs, laser spot size r_0 = 12 μm, laser wavelengthλ= 1 μm, and plasma density of 10^{19} cm^{-3}.

As the process of self-injection of electrons in bubble wake is closely related to the evolution of bubble. Thus it is crucial to understand the dynamic process of bubble lengthening. Bubble shape is basically characterized by longitudinal bubble radius (r_{bl}). The bubble evolution has been studied for three different values of a_0 (i.e., a_0=10, 15, 21 for the corresponding laser intensities 1.37 ×10^{20} , 3.1×10^{20}, 6.4 ×10^{20} W/cm^2, respectively). Initially, the longitudinal bubble radius increases with time and starts to fall (as shown in Fig.1). The lengthening of bubble is due to the intensity amplification of the pulse in plasmas. To explore the bubble shape during propagation, both longitudinal and transverse radii have been considered. Initially spherical bubble shape has been found to transform in to longitudinally elliptical shape. The eccentricity of the elliptical bubble is found to be varying with time. This is due to the change in residual electrons in bubble depending on laser intensity variation in plasma.

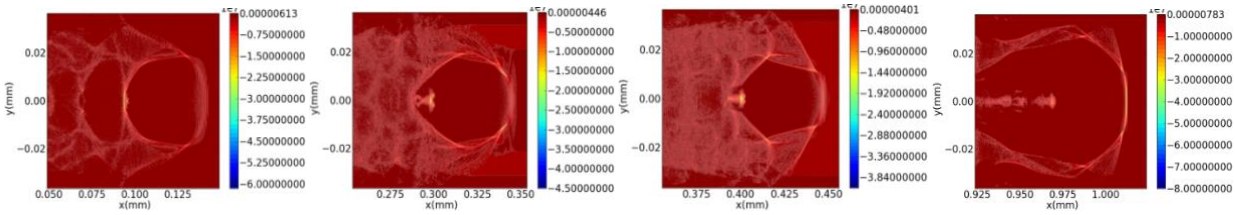


FIG 2: Electron density distribution plots extracted from PIC simulation at different simulation time (i) t= 0.5 ps, (ii) t = 1.2 ps, (iii) t = 1.5 ps, and (iv) t = 3.4 ps. Other parameters are the same as taking in Fig.1.

The density plots shown in Fig. 2 shows that the self-injection of electrons at the bubble back is primarily affected by the bubble evolution. It can be observed from Figs. 1 and 2 that no substantial injection is visible initially. This is because the bubble has not started to expand yet. The injection becomes noticeable as the longitudinal radius starts to grow. At t = 1.5 ps, the longitudinal bubble radius falls and the corresponding injected charge also gets reduced. This is due to the de-trapping of electrons from the bubble. A significant amount of charge is injected into the wake only after the bubble radius has started to increase significantly (for example, after the propagation time 1.5

ps). The simulation results also show that the self-injection of electrons is accompanied by bubble expansion. The beam quality depends on self-injection process, which further depends on bubble evolution. So controlling the bubble shape and its evolution can be of great importance to control the beam quality and energy. Hence simulations have been performed to control the beam quality by reducing energy spread using plasma density parameter, which will be presented in our future work.

CONCLUSION

In this paper, PIC simulations have been performed to analyse the bubble evolution process in laser wakefield acceleration. Previous assumptions of spherical and ellipsoidal geometry of the bubble throughout the propagation length is found deficient to explain various non-linear phenomena such as self-injection. Hence this concept of constant bubble shape has been re-analyzed. The quantitative analysis of the change in bubble shape during its propagation in plasma has also been provided. It has been shown that the evolution of bubble is majorly determined by laser evolution. The self-injection dynamics and its relation with the bubble evolution have been discussed. It has been shown that bubble expansion facilitates electron injection and contraction suppresses it. The beam energy and the beam quality is shown to be dependent on electron injection, hence, it is necessary to understand the mechanism of bubble evolution.

ACKNOWLEDGMENTS

This study was financially supported by the University Grant Commission (UGC), Delhi, Govt. of India for JRF fellowship to Monika Yadav.

REFERENCES

1. J. Faure, Y. Glinec, A. Pukhov, S. Kiselev, S. Gordienko, E. Lefebvre, J. P. Rousseau, F. Burgy, V. Malka, Nature **431** 541(2004).
2. X. Zeng, K. Zhou, Y. Zou, Q. Zu, et al., Opt. Lett. **42,** 2014 (2017).
3. S. Kalmykov, S. A. Yi, V. Khudik, G. Shvets, Phys. Rev. Lett. **103,** 135004 (2009).
4. A. Sävert, S. P. D. Mangles, et al., Phys. Rev. Lett. **115,** 055002 (2015).
5. I. Kostyukov, A. Pukhov, S. Kiselev, Phys. Plasmas **11,** 5256 (2004).
6. R. Sadighi-Bonabi, SH. Rahmatollahpur, Phys. Plasmas **17** 033105 (2010).
7. C. Benedetti, C. B. Schroeder, E. Esarey, F. Rossi, W. P. Leemans, Phys. Plasmas **20** 103108 (2013).

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 15 (2019) pp. 3435-3441
© Research India Publications. http://www.ripublication.com

Prediction of Compressive Strength of Cement Mortar in Normal and Aggressive Environment Using Artificial Neural Network

Kavita Verma[a], Shobha Ram[b], Alok Verma[c]

[a]*Research Student, Gautam Buddha University, Greater Noida, Uttar Pradesh 201308, India.*
[b]*Assistant Professor, Gautam Buddha University, Greater Noida, Uttar Pradesh 201308, India.*
[c]*Professor, Delhi Technological University, Delhi – 110042, India.*

Abstract:

The influence of acidic aggressive chemicals on the behavior of building materials such as cement mortar and cement concrete are a topic of significance interest, as it affects durability of structures. Development of predictive models for the assessment of various properties and characteristics of a construction material is important for deciding its future use in appropriate applications. Various approaches have been used to model the conditions of decay and degradation of cement-based materials in aggressive environments. In this paper, Artificial Neural Networks (ANNs) have been used to predict the compressive strength of a cement mortar exposed to normal and aggressive environments. The results have been found to be encouraging.

Keywords: Cement mortar, Artificial neural network, Aggressive Environment

INTRODUCTION

Cement based materials such as cement mortar and cement concrete have been regarded as the most prominent and promising material of construction in modern times. Durability related issues for such materials become very important in aggressive environments. The influence of such environments may affect various durability related properties of such materials (Al- Amoudi 2002; Tamimi et al 2003; Brown et al 2004; Barbara 2008; Abdelmseeh 2008; Irassar et al. 2009; Aye et al 2012; Zeng et al 2018,). Apart from such issues various properties in the hardening stage are to be assessed so that appropriate decisions about the quality of mixes, to be used, might be taken. In such situations also, some predicting system is of a great importance.

Lakhdari (2017) has considered the durability aspects of 1:3 cement mortar samples in water and aggressive environments of Na2SO4 and MgSO4. The samples were water cured for 28 days and then these were put to aggressive environments. The study lasted for a duration of 3 months. In the environments of Na2SO4, a degradation of compressive strength from 38.5 MPa to 29 MPa has been reported. In the case of other environments, the degradation has been reported to be from 38.5 MPa to 27 MPa. In the initial curing in water for a duration of 28 days, strength increased as usual. In a separate study of 360 days by Huang Ji Zhuanga et al., (2017), the effect of 5% NaCl has been monitored for a duration of one year. In the case of water curing of mortar samples, compressive strength improved but in the aggressive environment of NaCl there was a conflicting pattern of increase and decrease of compressive strength. The reason for the increase of compressive strength at a particular time seems to be filling up of internal pores by ettringite. In a study by Demir et al., (2018), the cement mortar samples were put to aggressive environment of Na2SO4 from the start of the study. Water binder ratio was varied from 0.5 to 0.625. The study lasted for a period of 365 days. Values of compressive strength were reported for different ages from 7 days to 365 days for different mixes as per water binder ratio. Compressive strength in water curing condition increased from 44.6 MPa to 54.85 MPa within a time span of one year for 0.5 % water binder ratio. Kallel et.al., (2016) has reported a study conducted in Na2SO4 and HCl for a period of one year. Decrease of compressive strength has been seen to be more in the case of HCl. In the study reported by Lee (2009) a comparative study of development of compressive strength in water and MgSO4 has been made. Water cement ration of 0.45 was taken for the cement mortar samples. Compressive strength improved from 46.8 MPa to 59.1 MPa in a time span of 360 days. In the environment of MgSO4 it changed from 40.9 MPa to 38.2 MPa. In a study conducted by Sarıdemir (2009) the amount of superplasticizer was changed from 0.5 to 1.3%. It was found that compressive strength increased in the case of use of superplasticizer in cement mortar mixes. In another study by Tamiz (2017) it has been found that compressive strength deterioration occurred faster in the case of MgSO4 in comparison to Na_2SO_4 environment in a time span of 450 days.

The above discussion underlines the importance of prediction of compressive strength of cement mortar mixes in various environmental conditions, seeing to a long span of time which is needed to appreciate the effect of a particular environment on it. Normally a finite time span is needed to appreciate strength development in cement mortar mixes in a particular environment. It would be important to have sound modelling systems to predict properties of cement mortar in such cases to reduce the time requirement to assess the effect of a particular aggressive environment on cement mortar.

EXPERIMENTAL PROGRAMME

Durability performances of cement mortar mixes, with or without superplasticizers, were determined in 1N and 2N H_2SO_4 environment. Properties of cement and superplasticizer used in this sample preparation are given in Table-1 and Table-2 respectively. The fine aggregate used was Yamuna sand.

Mortar cubes samples of size 70.7 mm were prepared as per Indian standard procedure. Three dosage of superplasticizer 0.5%, 1%, 1.5% by weight of cement were taken.

Table 1. Properties of Cement

Properties	values
Cement Type	OPC 43 Grade
Specific gravity	3.15
Autoclave expansion (%)	0.10
Fineness (m^2/kg)	293
Initial Setting time (minute)	140
Final setting time	265
Consistency	31.5

Two mixes of cement: sand proportions (1:4 and 1:6) were prepared. The specimens were placed inside the curing tank in normal water for 28 days. After 28 days, some specimens were put in normal water and some cubes were stored in acidic water made by H_2SO_4 addition (1Normal & 2 Normal concentration). In both aggressive environments, the H_2SO_4 solution was replaced from time to time. Visual examination of the samples was performed at regular intervals and all significant moderation such as color change, efflorescence, change in texture, cracking, expansion, etc, were recorded. The compressive strength of the mortar specimens was measured after 28 days, 90 days, 180 days and 360 days of exposure in water and in both aggressive environments to investigate the effect of the sulfate attack on the strength loss of the samples.

Table 2. Properties of Superplasticizer

Type /chemical origin	Polycarboxylates
Specific gravity	1.03
Color	Light yellow
Form	Liquid
Chloride content	Nil
Alkali content	Typically, less than 1.5g Na_2O equivalent /liter of admixture
P^H	Minimum 6

In this paper, an effort has been made to predict compressive strength of cement mortar in various environmental conditions. The requirement of the development of cement-based mixes having more strength has resulted into the increased use of superplasticizers in concrete mixes. That is why prediction of compressive strength of cement mortar mixes, in which superplasticizer was used, has also been made.

Use of Artificial Neural Network (ANN) for prediction of compressive strength

Use of ANN has been made by various researchers to predict various properties of cement-based materials (Roca et al 2013; Eskandari 2016). Artificial neural network (ANN) is a computational or mathematical non-linear statistical data modeling tool that is inspired by structural and functional neuron aspects of the human brain [Behnood et al., 2018]. It consists of many artificial neurons interconnected where each of them gives a single output induced from all inputs (Deng 2018; Topcu 2008). The neural network is a massively parallel distributed processor that has a tendency for storing experience knowledge and making it available for use. ANN collects the data by input layer of neurons, sometimes referred to as nodes or processing units. After a neural network has been created, it needs to be configured and then trained. Layers between the input and output layers are called hidden layers and may contain a large number of hidden processing units. The input layer neurons receive information from the outside environment and transmit them to the neurons of the hidden layer without performing any calculation. Finally, the output layer neurons produce the network predictions to the outside world (Naderpour et al 2018; Zhou et al 2016; Wang et al 2018; Yadollahi et al 2016; Deshpande et al 2014).

Feed-forward networks are the most popular models and widely used in many practical applications. Feed-forward neural networks are also known by many other names, such as multilayer perceptrons (MLP) which is a modification of standard linear perceptron. Feedforward networks can be used for any kind of input to output mapping. The artificial neurons consist of a series of layers and all the neurons in each layer have connections to all the neurons in the next layer. The first layer has a connection from the network input. Each subsequent layer has a connection from the previous layer. The final layer produces the network's output. It can distinguish data that is not linearly separable. A feedforward network with one hidden layer and enough neurons in the hidden layers, can fit any finite input-output mapping problem. Generally, network consists of one input layer, hidden layers (one or two) and one output layer of neurons. Each neuron in the network behaves in the same way.

RESULTS AND DISCUSSION

In the present study, feed-forward back propagation artificial neural network model is used. ANN has been adopted in mortar mix design due to complex relation between mortar behavior due to addition of superplasticizers and mix design proportion for determining the compressive strength in normal and aggressive environment condition. The modelling has been

carried out with the data available from the literature. Table 3 provides experimental and predicted values of compressive strength (F_c) for various types of cement mortar mixes in water and aggressive environmental conditions. Table 4 compares the differences between these values in percentage terms.

Table 3: Experimental and predicted values of compressive strengths

Mix. No.	Independent variables			F_C, MPa in water		F_C, MPa in H_2SO_4 (1N)		F_C, MPa in H_2SO_4 (2N)	
	Cement: sand	SP, %	Age (days)	Experimental	Predicted	Experimental	Predicted	Experimental	Predicted
1.	1:4	0	28	22.47	21.99	22.47	22.36	22.47	22.15
2.	1:4	0.5	28	27.88	26.51	27.88	27.50	27.88	28.66
3.	1:4	1	28	30.66	29.37	30.66	29.60	30.66	30.79
4.	1:4	1.5	28	29.6	30.48	29.60	29.17	29.6	30.48
5.	1:4	0	90	26.61	26.87	26.30	26.48	26.1	23.14
6.	1:4	0.5	90	33.33	31.85	32.90	33.42	32.6	30.91
7.	1:4	1	90	36.7	34.7	36.30	36.60	36.2	33.77
8.	1:4	1.5	90	35.2	35.54	34.90	36.41	34.8	33.53
9.	1:4	0	180	29.06	31.46	17.30	17.57	17.1	17.85
10.	1:4	0.5	180	36.2	36.58	21.70	21.14	21.5	23.09
11.	1:4	1	180	39.9	39.32	24.20	23.19	24.1	23.73
12.	1:4	1.5	180	37.9	39.87	22.90	23.52	22.9	21.63
13.	1:4	0	360	30.22	25.73	11.90	13.52	11.7	10.83
14.	1:4	0.5	360	37.8	32.39	15.10	14.76	14.6	14.98
15.	1:4	1	360	41.7	36.91	17.20	14.75	17	17.36
16.	1:4	1.5	360	39.7	38.78	16.00	13.68	15.7	16.25
17.	1:6	0	28	18.16	17.5	18.16	18.01	18.16	19.05
18.	1:6	0.5	28	18.52	17.94	18.52	18.37	18.52	20.19
19.	1:6	1	28	20.48	19.33	20.48	19.17	20.48	20.75
20.	1:6	1.5	28	21.77	22.74	21.77	20.77	21.77	21.91
21.	1:6	0	90	21.2	20.51	21	21.45	20.7	19.00
22.	1:6	0.5	90	21.7	20.8	21.5	22.34	21.3	21.75
23.	1:6	1	90	24.2	22.18	24	23.59	23.7	23.48
24.	1:6	1.5	90	25.6	25.85	25.5	25.65	25.2	25.50
25.	1:6	0	180	23.1	23.94	13.2	13.86	13.4	12.17
26.	1:6	0.5	180	23.7	23.97	14.3	14.72	13.9	13.50
27.	1:6	1	180	26.3	25.16	16	16.13	15.6	14.37
28.	1:6	1.5	180	27.9	28.87	17.2	18.47	16.6	15.50
29.	1:6	0	360	23.9	24.01	9.1	10.37	8.4	9.40
30.	1:6	0.5	360	24.4	24.64	9.4	10.29	8.6	10.05
31.	1:6	1	360	27.4	25.96	11.1	10.25	10	10.55
32.	1:6	1.5	360	28.8	29.61	11.8	10.33	10.8	10.82

Table 4: Comparison between Experimental and Predicted values

Mix No.	Independent variables			Variation between predicted and Experimental strength in water, %	Variation between predicted and Experimental strength in H_2SO_4 (1N), %	Variation between predicted and Experimental strength in H_2SO_4 (2N), %
	Mix	SP, %	Age (days)			
1.	1:4	0	28	-2.15	-0.50	-1.41
2.	1:4	0.5	28	-4.92	-1.36	2.78
3.	1:4	1	28	-4.22	-3.46	0.43
4.	1:4	1.5	28	2.98	-1.46	2.96
5.	1:4	0	90	1.00	0.68	-11.35
6.	1:4	0.5	90	-4.43	1.58	-5.20
7.	1:4	1	90	-5.44	0.81	-6.71
8.	1:4	1.5	90	0.96	4.33	-3.64
9.	1:4	0	180	8.26	1.55	4.39
10.	1:4	0.5	180	1.06	-2.59	7.39
11.	1:4	1	180	-1.46	-4.17	-1.54
12.	1:4	1.5	180	5.21	2.71	-5.56
13.	1:4	0	360	-14.86	13.62	-7.43
14.	1:4	0.5	360	-14.32	-2.25	2.62
15.	1:4	1	360	-11.48	-14.22	2.11
16.	1:4	1.5	360	-2.32	-14.47	3.49
17.	1:6	0	28	-3.66	-0.84	4.90
18.	1:6	0.5	28	-3.11	-0.83	9.03
19.	1:6	1	28	-5.60	-6.41	1.30
20.	1:6	1.5	28	4.45	-4.59	0.64
21.	1:6	0	90	-3.27	2.12	-8.23
22.	1:6	0.5	90	-4.16	3.93	2.12
23.	1:6	1	90	-8.36	-1.71	-0.91
24.	1:6	1.5	90	0.99	0.58	1.18
25.	1:6	0	180	3.62	5.00	-9.18
26.	1:6	0.5	180	1.12	2.97	-2.86
27.	1:6	1	180	-4.33	0.80	-7.87
28.	1:6	1.5	180	3.49	7.40	-6.63
29.	1:6	0	360	0.47	13.93	11.90
30.	1:6	0.5	360	1.00	9.45	16.83
31.	1:6	1	360	-5.25	-7.64	5.51
32.	1:6	1.5	360	2.81	-12.43	0.21

It is appreciated that the use of ANN for prediction of compressive strength values of cement mortar samples works in a satisfactory manner in different environments. In the case of water curing the minimum and maximum differences of experimental and predicted values are found to be 0.4% and 14.86% respectively. On an average scale the difference is about 4.4% which seems to be within acceptable limits. A greater value of difference in this study may be lack of knowledge of all the parameters in the past studies made by researchers. Standard deviation of 3.61% was found in this case. For the 1N aggressive environment of H2SO4, these values are 0.5%, 14.47%, 4.7% and 4.54% respectively. These values seem to be in line with those obtained in the case of water curing. It implies that the use of ANN can be effectively made in the case of samples put in aggressive environments. In the case of 2N aggressive environment of H2SO4, these values are found to be 0.21%, 16.83%, 4.95% and 3.9% respectively. Compared to the 1N aggressive environment, the minimum value of this case is found to be appreciably lower. This seems to give the impression that it may be more accurate to predict the effect of an aggressive environment it the severity of the environment is more.

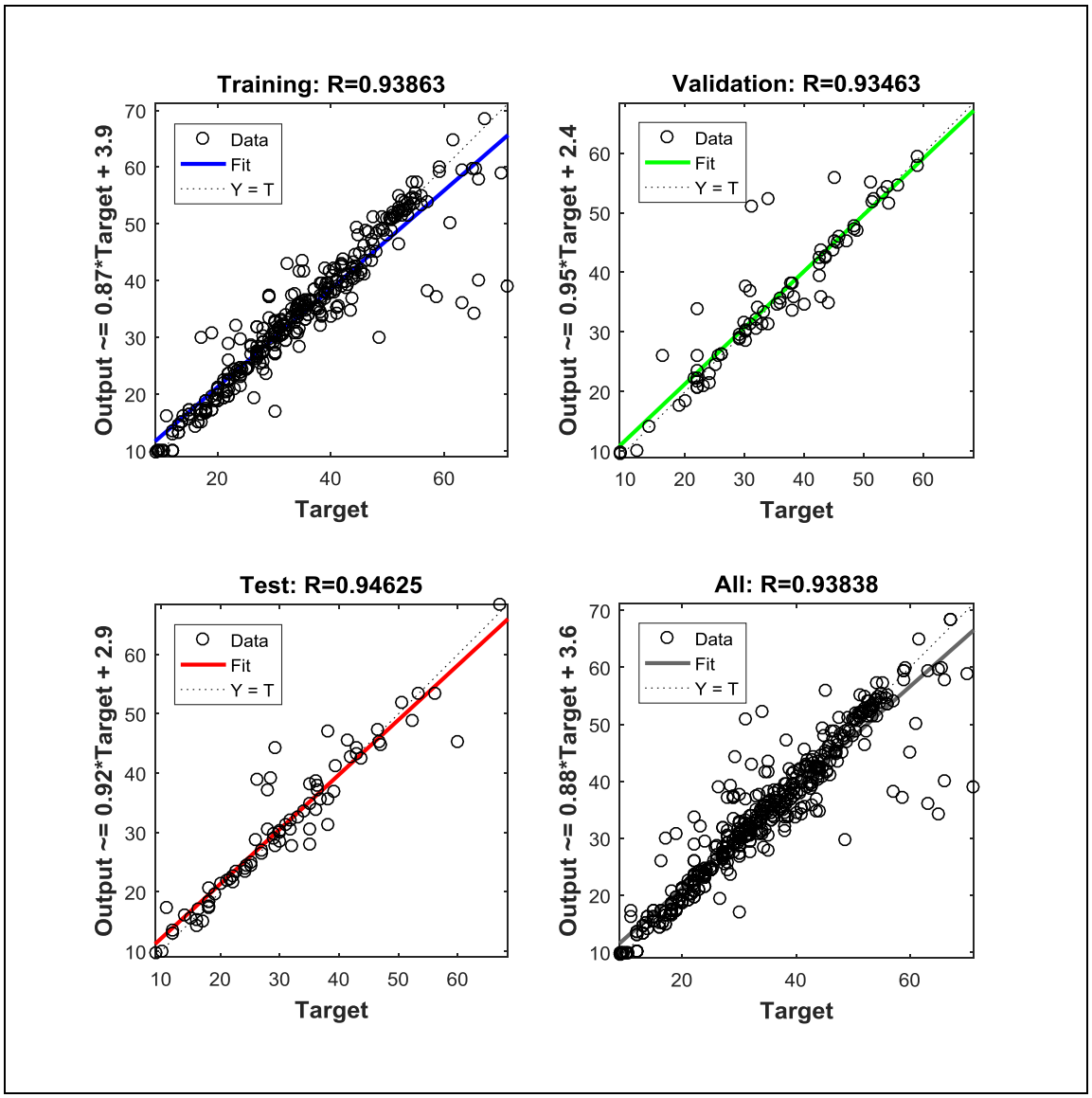


Fig. 1: Regression data for the input and predicted values

The prediction of compressive strength in aggressive environments may be more effective if sufficient parameters of past studies, undertaken by researchers, are known. This study has given sufficiently accurate results even when many parameters of past studies were not exactly known and those could not be verified. The results of regression analysis shown in Fig. 1 gives good results and it validates the use of ANN for the prediction of compressive strength values.

CONCLUSION

Based on the findings of this study it can be concluded that

1. The use of ANN technique for the prediction of compressive strength of cement mortar can be effectively made. The technique has worked well within acceptable limits for two types of aggressive environments also.
2. The prediction of compressive strength in aggressive environments may be more effective if sufficient parameters of past studies, undertaken by researchers, are known.
3. In this study a variation of superplasticizer content was considered. It is seen that variation of predicted values from experimental results change for a corresponding change in superplasticizer content. It shows that the content of superplasticizer also is an important parameter affecting compressive strength and it affects the performance of cement mortar in aggressive environment.
4. It seems that the accuracy of prediction may be more closely aligned with the experimental values if the degree of severity of aggressive environment increases. It is obvious as the effect of more severe environment can be readily assessed. In the case of less severe environments, the effect is difficult to be diagnosed and there are many grey areas in terms of actual effects of various parameters.

REFERENCES

[1] Abdelmseeh V. Assaad, J. Jofriet and Hayward G., 'Sulphate and sulphide corrosion in livestock buildings, Part I: Concrete deterioration', *Biosystems Engineering*, (2008) 372 – 381

[2] Al-Amoudi Omar S. Baghabra, 'Attack on plain and blended cements exposed to aggressive sulfate environments', *Cement & Concrete Composites* 24 (2002) 305–316

[3] Aye T., Oguchi C. T., 'Resistance of plain and blended cement mortars exposed to severe sulfate attacks', *Construction and Building Materials* 25 (2011) 2988 - 2996

[4] Behnood, E.M. Golafshani, Predicting the compressive strength of silica fume concrete using hybrid artificial neural network with multi-objective grey wolves, J. Clean. Prod. 202 (2018) 54–64,

[5] Brown Paul, Hooton R. D. and Boyd Clark, 'Microstructural changes in concretes with sulfate exposure', *Cement & Concrete Composites* 26 (2004) 993–999

[6] F. Deng, Y. He, S. Zhou, Y. Yu, H. Cheng, X. Wu, Compressive strength prediction of recycled concrete based on deep learning, Constr. Build. Mater. 175 (2018) 562–569,

[7] Gad E. A. M., M. R. Mabrouk, and F. H. Mosallamy, 'Rheological properties of different cement pastes made with different admixtures', *Silicate Industries*, vol. 70, no. 3-4, pp. 59–64, 2005.

[8] H. Temiz, M.M. Kose, S. Koksal, Effects of portland composite and composite cements on durability of mortar and permeability of concrete, Construction and Building Materials, Volume 21, Issue 6, 2007, Pages 1170-1176, ISSN 0950-0618,

[9] Hosein Naderpour, Amir Hossein Rafiean, Pouyan Fakharian, Compressive strength prediction of environmentally friendly concrete using artificial neural networks, Journal of Building Engineering, Volume 16, March 2018, Pages 213-219

[10] Huang Ji Zhuang, Hai Yan Zhang, Hao Xu, Resistance of geopolymer mortar to acid and chloride attacks, Procedia Engineering, Volume 210, 2017, Pages 126-131,

[11] I.B. Topçu, M. Sarıdemir, Prediction of compressive strength of concrete containing fly ash using artificial neural networks and fuzzy logic, Comput. Mater. Sci. 41 (2008) 305–311,

[12] İlhami Demir, Selahattin Güzelkücük, Özer Sevim, Effects of sulfate on cement mortar with hybrid pozzolan substitution, Engineering Science and Technology, an International Journal, Volume 21, Issue 3, 2018, Pages 275-283,

[13] Jansen D., J. Neubauer, F. Goetz-Neunhoeffer, R. Haerzschel, W.-D. Hergeth, 'Change in reaction kinetics of a Portland cement caused by a superplasticizer — Calculation of heat flow curves from XRD data', *Cement and Concrete Research*, Volume 42, Issue 2, February 2012, Pages 327-332

[14] Julio Garzón-Roca, Creu Obrer Marco, Jose M. Adam, Compressive strength of masonry made of clay bricks and cement mortar: Estimation based on Neural Networks and Fuzzy Logic, Engineering Structures, Volume 48, 2013, Pages 21-27, ISSN 0141-0296,

[15] Lothenbach Barbara, Thomas Matschei, Göril Möschner, Fred P. Glasser, 'Thermodynamic modelling of the effect of temperature on the hydration and porosity of Portland cement', *Cement and Concrete Research*, 38 (2008) 1–18

[16] Mohammed Fatah Lakhdari, Ali Zaidi, Mohamed Bouhicha, Benharzallah Krobba, Combined Effect of Temperature and Sulfate Attack on the Durability of Repair Mortar Based on Mixture of Dune-Alluvial Sand, Energy Procedia, Volume 139, 2017, Pages 772-777, ISSN 1876-6102,

[17] Mustafa Sarıdemir, Predicting the compressive strength of mortars containing metakaolin by artificial neural networks and fuzzy logic,Advances in Engineering Software,Volume 40, Issue 9,2009,Pages 920-927,ISSN 0965-9978

[18] Neela Deshpande, Shreenivas Londhe, Sushma Kulkarni, Modeling compressive strength of recycled aggregate concrete by Artificial Neural Network, Model Tree and Non-linear Regression, International Journal of Sustainable Built Environment, Volume 3, Issue 2, December 2014, Pages 187-198, ISSN 2212-6090

[19] Qiang Zhou, Fenglai Wang, Fei Zhu, Estimation of compressive strength of hollow concrete masonry prisms using artificial neural networks and adaptive neuro-fuzzy inference systems, Construction and Building Materials, Volume 125, 30 October 2016, Pages 417-426

[20] Safaa M. El Gamal, Heyam M. Bin Salman, 'Effect of addition of Sikament-R superplasticizer on the hydration characteristics of portland cement pastes', *HBRC Journal*, Volume 8, Issue 2, August 2012, Pages 75-80

[21] Safaa M.A. El-Gamal, Fawzia M. Al-Nowaiser, Asmaa O. Al-Baity, 'Effect of superplasticizers on the hydration kinetics and mechanical properties of Portland cement pastes', *Journal of Advanced Research*, Volume 3, Issue 2, April 2012, Pages 119-124

[22] Seung-Tae Lee, Influence of recycled fine aggregates on the resistance of mortars to magnesium sulfate attack,Waste Management,Volume 29, Issue 8,2009,Pages 2385-2391,

[23] Taheni Kallel, Abderrazek Kallel, Basma Samet, Durability of mortars made with sand washing waste, Construction and Building Materials, Volume 122, 2016, Pages 728-735, ISSN 0950-0618,

[24] Tamimi A. K. and Sonebi M., 'Assessment of Self-compacting concrete immersed in acidic solutions', *Journal of Materials in Civil Engineering*, ASCE, Vol. 15, No. 4, 2003, 354-357

[25] Yadollahi, E. Nazemi, A. Zolfaghari, A.M. Ajorloo, Application of artificial neural network for predicting the optimal mixture of radiation shielding concrete, Progress in Nuclear Energy, Volume 89, May 2016, Pages 69-77, ISSN 0149-1970

[26] Zeng, Q., Wang, C., Luo, Y., Yu, C., Huang, Q. and Luo, C. 2018. Effect of temperatures on TSA in cement mortars under electrical field, Construction and Building Materials. 162:88-95.

[27] Zhenjun Wang, Jiayu Wu, Peng Zhao, Nan Dai, Zhiwei Zhai, Tao Ai, Improving cracking resistance of cement mortar by thermo-sensitive poly N-isopropyl acrylamide (PNIPAM) gels, Journal of Cleaner Production, Volume 176, 1 March 2018, Pages 1292-1303

Pattern Analysis and Applications
https://doi.org/10.1007/s10044-019-00847-7

INDUSTRIAL AND COMMERCIAL APPLICATION

Robust object tracking with crow search optimized multi-cue particle filter

Gurjit Singh Walia[1] · Ashish Kumar[2] · Astitwa Saxena[3] · Kapil Sharma[2] · Kuldeep Singh[4]

Received: 7 August 2018 / Accepted: 22 August 2019
© Springer-Verlag London Ltd., part of Springer Nature 2019

Abstract
Particle filter is used extensively for estimation of target nonlinear and non-Gaussian state. However, its performance suffers due to its inherent problem of sample degeneracy and impoverishment. In order to address this, we propose a novel resampling method based upon crow search optimization to overcome low performing particles detected as the outlier. Proposed outlier detection mechanism with transductive reliability achieves faster convergence of the proposed PF tracking framework. In addition, we present an adaptive fusion model to integrate multi-cue extracted for each evaluated particle. Automatic boosting and suppression of particles using the proposed fusion model not only enhance the performance of the resampling method but also achieve optimal state estimation. Performance of the proposed tracker has been evaluated over benchmark video sequences and compared with state-of-the-art solutions. Qualitative and quantitative results reveal that the proposed tracker not only outperforms existing solutions but also efficiently handles various tracking challenges. On average of the outcome, we achieve CLE of 10.99 and F measure of 0.683.

Keywords Particle filter · CSA · Object tracking · Fusion model

1 Introduction

Object tracking is a vital research area in computer vision due to its promising applications in the field of military surveillance, augmented reality, robotics, medical imaging and human–computer interaction. It aims at keeping track of the targeted object in a video sequence. Initially, the target is localized either manually or through a detection algorithm and subsequently, its trajectory is traced over the period of time. Keeping path of target's trajectory in a video sequence is challenging due to its scale and pose variations, fast and abrupt motion, background clutters and full or partial occlusion.

Generally, object tracking methods can be categorized into either stochastic framework or deterministic framework. Stochastic framework includes probabilistic methods in the Bayesian framework for state estimation. Under the stochastic framework, linear and Gaussian state estimation was addressed using Kalman filter [36], extended Kalman filter and unscented Kalman filter [15], whereas nonlinear and non-Gaussian state can be estimated using particle filter (PF) [3], condensation filter [6, 18] and boot strap filter. On the other hand, deterministic framework, viz. mean shift [13], fragment-based tracker [1] and multi-stage tracker [33], employs target search in each frame to maximize the similarity between the target and search space. Generally, these methods used less spatial information of the target and, hence, were vulnerable to occlusion and background clutters [34]. However, stochastic methods reduced target sampling patches during tracking and were able to address the challenges occurred due to similar backgrounds and occlusion.

✉ Kapil Sharma
kapil@ieee.org

Gurjit Singh Walia
gurjit.walia@gmail.com

Ashish Kumar
ashish.gupta14d@gmail.com

Astitwa Saxena
astu9880@gmail.com

Kuldeep Singh
kuldeep.er@gmail.com

[1] Defence Research and Development Organization, Delhi, India

[2] Delhi Technological University, Delhi, India

[3] Netaji Subhas Institute of Technology, Delhi, India

[4] Malaviya National Institute of Technology, Jaipur, India

Published online: 29 August 2019

Under stochastic methods, PF has shown better results than other methods, but its performance is severely degraded due to its inherent problem of particle degeneracy and impoverishment [22]. Particle degeneracy occurs due to updation of particles with insignificant weight. These particles dispense less toward the state estimation, and their usage for subsequent estimation leads to wastage of computational efforts. In order to address this, resampling techniques such as sequential resampling [10], adaptive resampling [35], dynamic resampling [42] and spline resampling [12] were proposed. But, the implication of these methods in PF leads to sample impoverishment, whereas all the particles converge to a single end leading to drastic degradation in tracker's performance. Recently, particle impoverishment is solved using nature-inspired and evolutionary optimization, viz. particle swarm optimization [40], firefly algorithm [14], spider monkey [25], social spider [2], cuckoo search [30], modified galaxy-based search [28], bat algorithm [20]. The exploitation and exploration characteristics of these methods introduce randomness in search space through diversification of particles. The optimal particles generated through these optimizations not only achieved a fast convergence rate but also enhanced the accuracy of PF framework. However, most of PF-based tracking solutions were proposed using only single feature and hence, leading to degradation in performance under dynamic environment.

Multi-cue-based tracking framework has been discussed and reviewed mostly under deterministic methods [31]. Recently, under stochastic methods, multi-cue-based tracking framework was investigated [11, 28, 32]. In these work, multiple complementary cues are integrated to overcome individual cue limitations. In [11], author used color, texture and edge cue for the object representation and integrated them using an adaptive weighting scheme. Walia and Kapoor [32] exploited color, thermal and motion feature for object representation and integrated their scores using PCR-5 rules. In [28], authors calculated color histogram and histogram of orientation gradient (HOG) for each particle. Resampling was performed using modified galaxy-based search algorithm. However, the adaptive integration of multi-cue was not addressed. In addition, resampling methods proposed in the previous work require the processing of all evaluated particles and, hence, require more computational power. Hence, future tracking solutions should consider the adaptive fusion of multi-cue to handle dynamic environmental challenges along with efficient resampling method for real-time applications.

In this paper, we propose a multi-cue object tracking framework based on PF with a novel nature-inspired resampling method. An adaptive fusion model is proposed for automatic suppression and boosting of particles weights. Outlier detection mechanism is also utilized to detect low performing particles to improve the performance of the resampling method. To summarize, the main contributions of our research work are as follows:

- We propose an object tracking framework in which target is represented by complementary multi-cue, viz. color and texture for each particle. Color cue is sensitive to illumination variations, similar background and occlusion, while texture cue can handle these variations effectively. On the other hand, texture cue fails during scale variations, while color cue maintains tracker's efficiency under this challenge.
- We also propose a novel resampling method based on crow search optimization (CSA). CSA maintains diversity in the search space by considering awareness probability. It improves performance by distributing particles in the high-likelihood region to address sample degeneracy. Unimportant particles detected as outlier were resampled to reduce the computational complexity of the proposed tracker.
- An outlier detection mechanism is proposed to identify unimportant particles (having negligible weight). Particle consistency in subsequent frames is exploited for detecting the particles which contribute less toward the state estimation. In addition, this mechanism improves the performance of the proposed resampling technique and ensures faster convergence.
- An adaptive fusion model is proposed for the integration of multi-cues. This model resolves the conflict among the particles and takes the decision on the basis of the fuzzy inference rule. It either boosts or suppresses the particles for their optimal contribution in the state estimation. It creates clear decision boundaries between significant particles and low performing particles. Model is made adaptive through transductive reliability calculated through correlation among the particles.

The rest of this paper is organized as follows. In Sect. 2, we review the literature closely related to our work. Sect. 3 discusses the architecture of the proposed multi-cue object tracking framework. Detailed design of resampling model and fusion model is also presented. Qualitative and quantitative results along with their comparison with existing methods are presented in Sect. 4. Finally, concluding remarks and future directions are drawn in Sect. 5.

2 Related work

Object tracking under PF framework has been extensively investigated due to its potential applications in video surveillance, military and robotics. However, PF inherent problem, viz. sample impoverishment and degeneracy, limits its performance of object tracking. Many solutions have been

proposed to address these problems and reviewed in [22]. In this section, we examine the closely related work of contemporary solutions for object tracking under PF framework.

In [14], the author used firefly algorithm (FA) before resampling to improve particle distribution. FA converged particles rapidly to an optimal solution by changing their positions. Use of optimization generates significant particles which lead to accurate target state estimation. However, the performance of this method degraded when FA trapped in local minimum. Walia and Kapoor [30] presented a resampling method based on improved cuckoo search (ICS) with Levy flight for object tracking under PF framework. ICS abandoned the particles with lesser weight to avoid the local minimum trap. However, optimal selection of step size for Levy flight was not addressed. In [25], authors utilized spider monkey optimization as resampling method in PF-based tracking framework. The particle's position was updated on the basis of experience of local and global leader. Use of optimization improved the quality of particles and converged them in high-likelihood region. Resampling method based on immune genetic algorithm (IGA) for object tracking under PF framework was discussed [17]. IGA used crossover and mutation process to generate new particles in the search region. IGA's regulatory mechanism, viz. promotion and suppression, ensured particles diversity. In [41], author used Pearson correlation coefficient as resampling method under PF framework. Similarly, in [8], authors explored the resampling method based on backtracking search optimization (BSA). But, both of these methods were not evaluated on video sequences. Ahmadi and Salari [2] applied social spider optimization (SSO) as resampling method for tracking framework based on PF. SSO grouped the particles as male and female to increase randomness in search space. If the target was lost, optimization ensured diversity by forcing the redistribution of particles. However, most of the methods under PF framework consider only single cue for particle evaluation. Mostly, single cue is not sufficient to handle dynamic environments. Hence, particles need to be evaluated over multi-cue and subsequently fused for compensating various tracking challenges.

In [40], authors exploited color histogram and HOG for evaluation of each particle and PSO was applied as resampling technique. But, this method was computationally complex due to the multilevel evaluation of each particle with multiple cues. In [28], authors used color histogram and HOG cue with resampling method based on modified galaxy-based search. Adaptive appearance model with multiple parameters was used at each time step to detect and handle occlusion. Due to this, the method was considerably slow. Similarly, Sardari and Moghaddam [27] used color and motion cue for likelihood calculation of particles. The resampling method based on modified galaxy-based search algorithm was proposed to obtain the optimum state of the object. In [32], authors exploited color, thermal and motion cue for each particle evaluation. The particles cues were fused using Dezert–Smarandache theory. Generally, the resampling mechanism for PF considers all the particles to ensure state estimation. This needs high computational power for desired convergence. Methodology to relocate the lost target has not been explored much. Moreover, the adaptive fusion of particles in multi-cue object tracking based on PF is limitedly addressed.

Review of closely related studies indicates that future research in multi-cue object tracking based on PF should consider the adaptive fusion of particles for efficient tracking. In addition, unimportant particles need to be detected not only to reduce computation but also for efficient resampling. The next section will detail the architecture of the proposed tracker.

3 Architecture of the proposed tracker

The proposed tracker is developed under PF-based probabilistic tracking framework with multi-cue evaluation for each and every particle. Extracted color cue (H_C) and texture cue (H_T) for each evaluated particle are integrated using the proposed adaptive multi-cue fusion model. Fusion model is made adaptive through the online calculation of transductive reliability metric for each cue. Particle-level multi-cue fusion is performed considering the boosting of good particles and simultaneously suppressing of the low performing particles. In each frame, low performing particles are detected using the proposed outlier detection mechanism. These particles are relocated to their optimum position using the proposed resampling method based on crow search optimization (CSA) [4]. Architecture of the proposed tracker is depicted in Fig. 1. As shown, at time instant $k = 1$, the target is segmented using Gaussian mixture model (GMM) [29] and N particles are initialized around the detected centroid of the target. These initialized particles are evolved using the multi-component state model to obtain the predicted particles. Each predicted particle is evaluated over two complementary cues, and weights are assigned through comparison of extracted features with the corresponding reference model. Further, these weights are integrated using the multi-cue fusion model to obtain fused weight w_k for each evaluated particles. Based upon their fused weights, particles are classified either as outlier (O) or as inlier (I) using the proposed outlier detection mechanism. Only the outliers O are processed through the proposed resampling method to relocate them in high-likelihood area. Resampling mechanism considers outlier as the initial population for CSA. The diversity of the particles is maintained through the awareness probability. On the other hand, search region is controlled through the flight length to obtain the optimal resampled particles. High convergence rate of CSA achieves

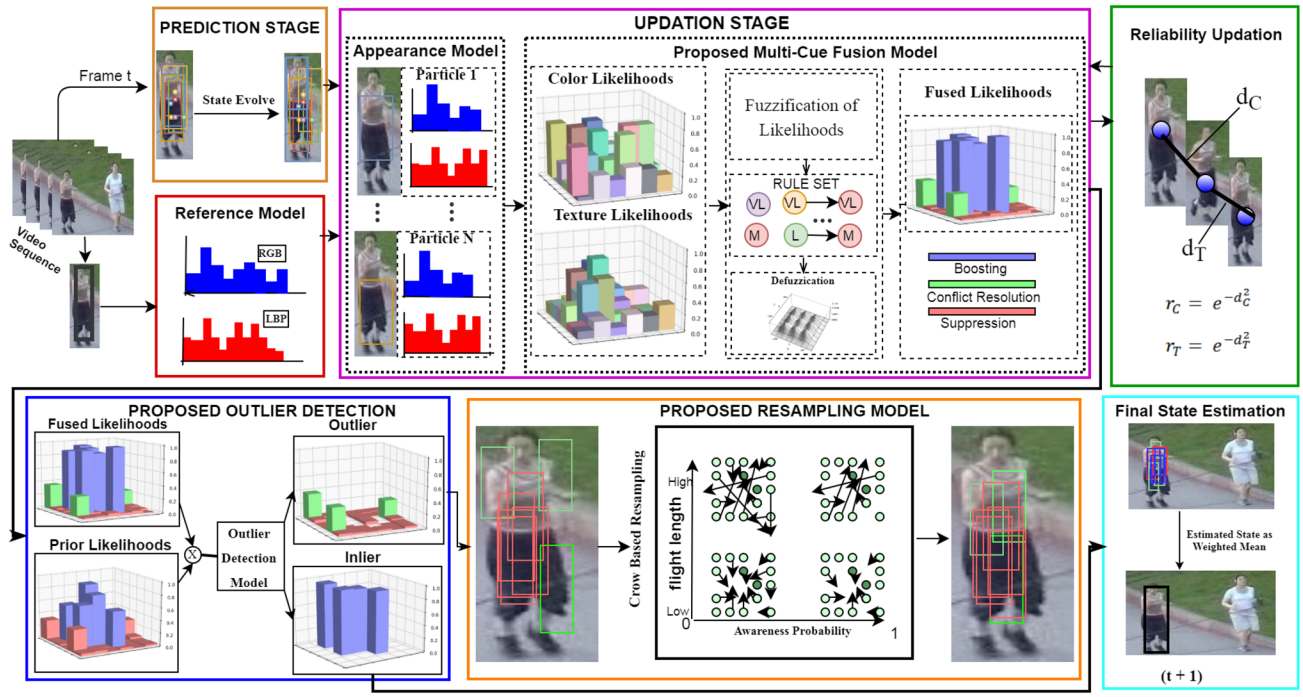


Fig. 1 Overview of the proposed Tracker. N particles are evolve using multi-component state model. Predicted particles are updated through multi-cue evaluation and adaptive fusion of evaluated particles. Detected outlier is resampled to overcome sample degeneracy. Targeted state is determined as weighted mean of N particles

the real-time realization of the proposed tracker. At each time step, transductive reliabilities r_C and r_T associated with the cues are evaluated to discount the particle in concurrence with the cue performance. These reliabilities ensure quick adaptation of the proposed tracker to the dynamic environment. Finally, the target state is determined as the weighted sum of inlier as well as resampled outlier particles. The detailed design of the proposed tracker is follow in turn.

3.1 Multi-cue particle filter

The proposed tracker is designed using PF [16] which is a probabilistic recursive Bayesian framework used for estimation of nonlinear and non-Gaussian state. Particle filter has been exploited for multi-cue tracking framework to precisely localize the target under dynamic environment. Aim is to approximate PDF $p(S_k|w_k)$ of the target state given its prior PDF $p(S_{k-1}|w_{k-1})$, where S_k and S_{k-1} are state vectors represented by N particles and w_k and w_{k-1} are set of measurements for frame k and $k-1$.

In the proposed model, we consider state vector for each particle as $S_k = (X_k, X'_k, Y_k, Y'_k, r_k, \theta_k)$. X_k and Y_k are Cartesian coordinates of the center of bounding box, X'_k and Y'_k are velocities in respective direction, and r_k and θ_k are rotation and scaling components of the particle. Measurement is obtained for each particle using complementary multi-cue, viz. color and texture extraction and their score fusion using proposed multi-cue fusion model. Particles are evolved using the state model as Eq. (1).

$$S_k = f(S_{k-1}, \hat{n}_{k-1}) \tag{1}$$

where $\hat{n}_{k-1}$ is zero-mean white noise and f: $R^{p'} \in R^{q'} \rightarrow R^{p'}$ is a multi-component transition function which includes constant velocity model and random walk model [11].

The PDF of target state S_k, i.e., $p(S_k|w_k)$ is obtained recursively into two stages, viz. prediction and update. The initial PDF of the target $p(S_{k-1}|w_{k-1})$ at frame $k = 1$ is manually obtained through random placement of particles. In the subsequent estimate, prior PDF of the target at frame k is obtained using Eq. (2).

$$p(S_k|w_{k-1}) = \int p(S_k|S_{k-1}) p(S_{k-1}|w_{k-1}) \mathrm{d}S_{k-1} \tag{2}$$

where multi-component state model is used to evolve the particles to obtain predicted particle represented PDF $p(S_k|S_{k-1})$. Further, the predicted particles representing PDF $p(S_k|S_{k-1})$ are updated by assignment of multi-cue measurement and fusion as discussed in Sect. 3.2. The updation of these particles is performed using Bayes rule using Eq. (3).

$$p(S_k|w_k) = \frac{p(w_k|S_k) p(S_k|w_{k-1})}{p(w_k|w_{k-1})} \tag{3}$$

In Eq. (3) normalizing denominator factor $p(w_k|w_{k-1})$ is obtained as:

$$p(w_k|w_{k-1}) = \int p(w_k|S_k)p(S_k|w_{k-1})\mathrm{d}S_k \tag{4}$$

where $p(w_k|S_k)$ is obtained through multi-cue evaluation of each particle and their subsequent fusion using proposed multi-cue fusion model. Also, resampling is performed to overcome particle degeneracy and sample impoverishment as discussed in Sect. 3.3.

Posterior PDF $p(S_k|w_k)$ is represented by N-updated particles and used to determine the state of the target as their weighted mean. In the iterative process of particle filter, these updated particles are used as prior initialization for the next stage estimation. Detail of the proposed tracker updation model is discussed in the next section.

3.2 Proposed updation model

The proposed tracker is realized through online detection of target through GMM [29] where N particles are initialized. Each particle state S_k is evolved through multi-component state model to achieve predicted particles with centroid as $Z_{k,P} \in [X_{k,P}, Y_{k,P})]$. Particles updation consists of mainly: (a) multi-cue evaluation and (b) adaptive fusion of likelihoods. For this, each particle is evaluated over two complementary cues, viz. color and texture. Color cue is invariant to scale variations and occlusion but sensitive to illumination variations and background clutters. On the other hand, texture cue is robust to illumination variations and background clutters. Hence, complementary color and texture cue compensate for each other during the tracking process.

The color cue is calculated using RGB color histogram model. Color histogram (H_{C}^g) for each gth particle with pixel location (a_p, b_p) is determined by Eq. (5).

$$H_{\mathrm{C}}^g = \hat{z}\sum_{p=1}^{N} B(a_p, b_p), C = 1, 2 \ldots M_b \tag{5}$$

where $B(.)$ represents binning function that assigns pixel (a_p, b_p) to one of the M_b histogram bins and normalizing factor $\hat{z}$ is calculated as $\sum_{c=1}^{M_b} H_c = 1$.

Unlike color cue, texture cue is used to extract low-level features from the image. It is determined for each particle by considering rotation invariant local binary pattern (LBP) [24]. LBP of gth particle is calculated from central pixel (a_c, b_c) with radius z and with equally spaced pixels p. The uniform pattern in image region used multi-resolution grayscale rotation for separating target texture information. It is represented as $U(\mathrm{LBP}_{z,p}((a_c, b_c))$ and determined using Eq. (7).

$$\mathrm{LBP}_{(z,p)}(a_c, b_c) = \begin{cases} \sum_{j=0}^{p-1}(T((u_j(a_j, b_j) - u_c(a_c, b_c))), & \text{if} U(\mathrm{LBP}_{z,j}(a_c, b_c)) \leq 2 \\ j+1, & \text{otherwise} \end{cases} \tag{6}$$

$$\begin{aligned} U(\mathrm{LBP}_{(z,p)}((a_c, b_c)) &= |T(u_{j-1}(a_{j-1}, b_{j-1}) - u_c(a_c, b_c)) \\ &- T(u_0(a_0, b_0) - u_c(a_c, b_c))| + \sum_{j=0}^{p-1} |T(u_j(a_j, b_j) \\ &- u_c(a_c, b_c)) - T(u_{j-1}(a_{j-1}, b_{j-1}) - u_c(a_c, b_c))| \end{aligned} \tag{7}$$

Using this LBP operator, normalized histogram is determined for gth particle by mapping the uniformity measure of every pixel and represented as H_T^g.

Further, weights are assigned to each particle through estimation of similarity between predicted particles histogram and corresponding references histograms. This is achieved through Bhattacharyya distance [9]. For this, Bhattacharyya coefficient is determined using Eq. (8).

$$\beta_i^g(H_i^R, H_i^g) = \sum_{r=1}^{N} \sqrt{(H_i^R)^r \times (H_i^g)^r} \tag{8}$$

where H_i^R represents reference histogram, $i \in C, T$ and H_i^g, denotes the histogram of considered cue for gth particle. Using Eq. (8), Bhattacharyya distance is calculated for color and texture cue as illustrated in Eq. (9).

$$D_i^g(H_i^R, H_i^g) = \sqrt{1 - \beta_i^g(H_i^R, H_i^g)} \tag{9}$$

Using this, likelihoods for each particle are determined using Eq. (10).

$$\alpha_i^g \propto e^{\left(-\frac{(D_i^g(H_i^R, H_i^g))^2}{2\sigma_i^2}\right)} \tag{10}$$

where σ_i denotes standard deviation and $i \in (C, T)$.

The color and texture likelihoods so calculated are subjected to the proposed adaptive multi-cue fusion model. The proposed model is designed considering boosting of concordant particles, suppressing of discordant particles and online conflict resolving among the conflicting particles. In order to achieve quick adaptation of tracker, individual likelihoods are discounted with the transductive reliability of the respective cue as determined in the previous frame. For this, particle likelihoods α_i^g are multiplied with the corresponding cue reliability as depicted in Eq. (11).

$$\gamma_i^g = r_{i,k-1} \times \alpha_i^g \tag{11}$$

where $r_{i,k-1}$ is the transductive cue reliability determined in the $(k-1)$ frame using Eqs. (28–29) and $i \in (C, T)$.

The proposed multi-cue fusion model achieves precise decision boundary among high performing particles and low performing particles under complex environment. Study has been reported that fuzzy logic can induce human being inferences in the fused model [7]. It was reported that human sensing system consists of multi-cue framework with online adaptation to dynamic environment [23]. Considering this,

the fuzzy set is defined as: $\hat{S} = \{\text{VS,S,M,L,VL}\}$, representing very small, small, medium, large and very large likelihood value for the particle. Membership value for the evaluated particle is determined as fuzzy set element using Eqs. (12–16):

$$\delta_{\text{VS}}(x) = \frac{1}{1 + e^{-c_{\text{VS}}(x - f_{\text{VS}})}} \tag{12}$$

$$\delta_{\text{S}}(x) = e^{\frac{-(x-d_{\text{S}})^2}{f_{\text{S}}^2}} \tag{13}$$

$$\delta_{\text{M}}(x) = e^{\frac{-(x-d_{\text{M}})^2}{f_{\text{M}}^2}} \tag{14}$$

$$\delta_{\text{L}}(x) = e^{\frac{-(x-d_{\text{L}})^2}{f_{\text{L}}^2}} \tag{15}$$

$$\delta_{\text{VL}}(x) = \frac{1}{1 + e^{-c_{\text{VL}}(x - f_{\text{VL}})}} \tag{16}$$

where $c_{\text{VS}} = -40, f_{\text{VS}} = 0.15, d_{\text{S}} = 0.3, f_{\text{S}} = 0.12, d_{\text{M}} = 0.5, f_{\text{M}} = 0.12, d_{\text{L}} = 0.7, f_{\text{L}} = 0.12, c_{\text{VL}} = 40$ and $f_{\text{VL}} = 0.85$ and $x \in \gamma_i^g, i \in (C, T)$. These linguistic variables values are obtained from the training data. The values are determined so that high performing particles are boosted and low performing particles are suppressed simultaneously.

The functional mapping $\hat{R}_{m,n}$ between the color and the texture likelihood is presented in Table 1, where m and n represent the fuzzy set value assigned to each evaluated particle. This mapping ensures boosting of high performing particles and suppression of low performing particles, creating a precise decision boundary for the proposed outlier detection mechanism.

Further to achieve crisp fused value of particle likelihood, defuzzification using center of gravity (COG) method has been considered for obtaining optimal results. The crisp COG value for a pair of fuzzy set elements m and n is defined as:

$$\text{COG}_{\hat{R}_{m,n}} = \frac{\sum_m \sum_n \delta_{\hat{R}_{m,n}}(x) x}{\sum_m \sum_n \delta_{\hat{R}_{m,n}}(x)} \tag{17}$$

where $\hat{R}_{m,n}$ is determined using Table 1 and $x \in \gamma_i^g$. The weighted mean, Eq. (18), of COG values being determined over pair (m, n) is used for estimation of final fused weight for each evaluated particle for k frame.

$$w_k^g = \frac{\sum_m \sum_n \pi_{m,n} \times \text{COG}_{\hat{R}_{m,n}}}{\sum_m \sum_n \pi_{m,n}(\gamma_{\text{C}}^g, \gamma_T^g)} \tag{18}$$

where $\pi_{m,n}$ is the fuzzy control rule calculated using Eq. (19).

$$\pi_{m,n} = \min(\delta_m(\gamma_{\text{C}}^g), \delta_n(\gamma_T^g)) \tag{19}$$

These evaluated particles are subjected to the resampling method discussed in the next section.

Table 1 Fuzzy mapping rules ($\hat{R}_{m,n}$)

n	m				
	VS	S	M	L	VL
VS	VS	VS	VS	M	L
S	VS	VS	S	M	L
M	VS	S	M	M	L
L	M	M	M	VL	VL
VL	L	L	L	VL	VL

3.3 Proposed resampling method

The proposed multi-cue fusion model creates precise decision boundaries between low performing particles and high-performing particles. Keeping low performing particle for subsequent iteration not only requires high computation but also impairs tracker accuracy. This problem of sample degeneracy has been solved using conventional techniques like adaptive resampling [35], sequential resampling [10], cuckoo search [30]. However, most of the resampling methods consider all the particles for the resampling process which requires high computational requirement. In order to realize the tracker in the real-time scenario, we determine the recurrently low performing particles as outliers which are subjected to the proposed resampling method. We assume outliers are those particles which are either drifted off the target or highly affected by the dynamic environment. Generally, the environment affects the target particles in consequent frames. This has been exploited by assuming that individual particle state estimation as a Markov process and helps us in the design of the proposed outlier detection mechanism. The proposed outlier detection mechanism is designed by considering particles performance in the previous frame along with their current estimation. The particles likelihood for the current frame k is fused with their estimated likelihoods for $k-1$ frame. These likelihoods for each particle are adaptively fused using the proposed multi-cue fusion model. For this, we consider the fuzzy set $\hat{S}$ for the two likelihood values for each particle and membership values ($\delta_m(.)$) are determined using Eqs. (12–16). Further, the final score κ^g for the gth particle for its classification as outlier or inlier is calculated using Eq. (20).

$$\kappa^g = \frac{\sum_m \sum_n \pi_{m,n}(w^g, w_{k-1}^g) \times COG_{\hat{R}_{m,n}}}{\sum_m \sum_n \pi_{m,n}(\gamma_k^g, w_{k-1}^g)} \tag{20}$$

where $\pi_{m,n}$ is the fuzzy control rule calculated using Eq. (21) and used for weighing the COG values.

$$\pi_{m,n} = \min(\delta_m(w_k^g), \delta_n(w_{k-1}^g) \tag{21}$$

Using fused score for each particle, outlier and inlier are classified by comparing their fused score with predefined

threshold value κ_{th}. Optimal value $\kappa_{th} = 0.82$ is used for classifying the particles either as outliers using Eq. (22) or as inliers for all videos using Eq. (23).

$$O_k^{u'} = \{Z_k^g, \forall g \quad \text{where} \quad \kappa^g \leq \kappa_{th}\} \tag{22}$$

$$I_k^v = \{Z_k^g, \forall g \quad \text{where} \quad \kappa^g > \kappa_{th}\} \tag{23}$$

where $|u' \cup v| = N, O_k^{u'} \in [X_k^{u'}, Y_k^{u'}]$, and $I_k^v \in [X_k^v, Y_k^v]$. Precise decision boundary determined through the proposed outlier detection mechanism aids proposed resampling mechanism to eliminate particle degeneracy. Only the particles classified as outliers are resampled using the proposed resampling technique. The resampling technique is based on crow search optimization which diversifies the outliers in the search space. The crow search optimization is a meta-heuristic optimization approach proposed in [4]. Outlier particle set $O_k^{u'}$ at frame k obtained through Eq. (22) is initialized as the crow population, and their position is updated as shown in Eq. (24).

$$O_k^u = \begin{cases} O_k^{u'} + rand \times \text{fl} \times (M_k^{u'} - O_k^{u'}), & r \geq \text{AP} \\ \text{random} \quad \text{redistribution}, & \text{otherwise} \end{cases} \tag{24}$$

where $rand \in [0, 1]$, $O_k^u \in [X_k^u, Y_k^u]$ fl represents the flight length, AP denotes the awareness probability, and $M_k^{u'}$ represents the global best location of each crow. This iterative process is performed for J iterations to achieve the optimal solution. Finally, the optimally localized outliers with updated weights updation are performed using method as discussed in Sect. 3.2 along with inliers are used for determining the target state.

3.4 State estimation and reliability updation

Final centroid of the object (Z_k^F) is estimated by using the updated fused likelihoods w_k^g as weights for inlier particles (w_k^v) and that of resampled outlier states (w_k^u). This is depicted in Eq. (25).

$$Z_k^F = \frac{\sum_v w_k^v \times I_k^v + \sum_u w_k^u \times O_k^u}{\sum_{g=1}^N w_k^g}, \quad |u \cup v| = N \tag{25}$$

For quick adaptation of the proposed tracker against environment challenges, reliability values are calculated for each cue individually. For this, the state is estimated using color and texture likelihoods and depicted in Eqs. (26)–(27).

$$Z_k^C = \frac{\sum_{g=1}^N \gamma_C^g Z_{k,P}^g}{\sum_{g=1}^N (\gamma_C^g)} \tag{26}$$

$$Z_k^T = \frac{\sum_{g=1}^N \gamma_T^g Z_{k,P}^g}{\sum_{g=1}^N (\gamma_T^g)} \tag{27}$$

where $Z_{k,P}^g \in [X_{k,P}^g, Y_{k,P}^g]$ is the predicted particle centroid. The reliability values are obtained through Eqs. (28)–(29) which are further passed on to the next frame estimation.

$$r_{C,k-1} \propto e^{-d_C^2} \tag{28}$$

$$r_{T,k-1} \propto e^{-d_T^2} \tag{29}$$

where d_C and d_T are the Euclidean distance between the final state and the state estimated for color and texture cue, respectively, and are given by Eqs. (30)–(31).

$$d_C = \sigma(Z_k^F, Z_k^C) \tag{30}$$

$$d_T = \sigma(Z_k^F, Z_k^T) \tag{31}$$

Algorithm 1: Proposed Tracker

```
Function Multi-cue Tracker():
    for each frame k do
        if k=1 then
            Intialise N particles on the detected target
        else
            S_k = f(S_{k-1}, n̂_{k-1})
            Obtain α_i^g using Eq. (10)
            γ_i^g ← r_{i,k-1} × α_i^g, i ∈ (C, T)
            Calculate w_k^g using Eq. (18)
            Obtain κ^g using Eq. (20)
            for g in Z_k do
                if κ^g > κ_th then
                    I_k^v = Z_k^g
                else
                    O_k^{u'} = Z_k^g
                end
            end
            for j from 1 to J do
                for c from 1 to u' do
                    rand ← U(0,1)
                    if rand ≥ AP then
                        O_k^u ← O_k^{u'} + rand × fl × (M_k^u − O_k^{u'})
                    else
                        O_k^u ← O_k^{u'} + U(0,1)
                    end
                    Update M_k^{u'}
                end
            end
            Obtain updated w_k^g using Eq. (18)
            Calculate Z_k^C, Z_k^T, Z_k^F using Eqs. (25-27)
            d_C=σ(Z_k^F, Z_k^C), d_T=σ(Z_k^F, Z_k^T)
            Obtain updated r_{C,k-1}, r_{T,k-1} using Eqs. (28-29)
        end
    end
    return Z_1^F, Z_2^F, ..., Z_k^F
```

Pseudocode of the proposed tracker is depicted in Algorithm 1. Updated inliers (I_k^v) and outliers (O_k^u) are used as

prior for next state estimation. These particles are further evolved using the state model Eq. (1). Predicted particles are updated using the multi-cue framework as discussed in Sect. 3.1. In addition, transductive reliabilities estimated at kth frame are used for discounting the particles weights at $k-1$ frame. This iterative process will continue until the detected target is present in the scene. Once the target is lost, the automatic detection mechanism will redetect the target and the particles are re-initialized in the iterative process. Details of experimental validation of the proposed tracker along with the comparison with state-of-the-art are presented in the next section.

4 Experimental validation

In this section, we demonstrate experimental validation of the proposed method and compare the results with nine other state-of-the-art trackers, viz. ASLA [19], CT [39], MIL [5], IVT [26], FRAG [1], WMIL [38], PF [16], PF-PSO [40] and PF-BSA [8] . In order to gauge robustness and reliability of the proposed method, we have evaluated our tracker over 16 benchmarked video sequences [37], encompassing various tracking challenges such as illumination variations (IV), scale variations (SV), full occlusion (FOC) or partial occlusion (POC), background clutters (BC), object's deformation (DEF) and rotation (RO). Table 2 tabulates the details of the considered video sequences.

Qualitative and quantitative analysis of results is performed using standard performance metrics. The proposed algorithm is implemented in Python2 and executed on a 2.4 GHz quad-core processor with 6 GB RAM. Proposed tracker is initialized with $N = 49$ particles on pre-detected target using GMM [29] segmentation technique. Probabilistic nature of PF is handled by iterating the algorithm ten times and taking the mean value of the results. We have taken publicly available trackers codes [1, 5, 19, 26, 38, 39] from object tracking benchmark [37]. PF, PF-PSO and PF-BSA are implemented on Python2 to compare their performance with the proposed method. In the next section, the qualitative analysis of the proposed method is demonstrated.

4.1 Qualitative analysis

4.1.1 Illumination and scale variations

Sample trackers output for *CarScale* frames under illumination and scale variations are illustrated in Fig. 2a. From frames #190 to #240, under prominent variation in the target's scale, the trackers, viz. ASLA, CT, MIL, WMIL and IVT totally deflected from the target. Due to their appearance model, these trackers fail to handle the large-scale variations of the target scale, but our tracker can efficiently track the target. Proposed multi-cue fusion model suppresses the low performing particles during the large-scale variations and, hence, maintains tracker's efficiency. In Fig. 2b, sample frames for sequence *Dog* are shown. At #126 frame, when the targeted object encounters scale variations, PF-BSA, IVT, CT and PF drifted away, while ours successfully tracked the marked object. This is mainly attributed to multi-component state model where scaling factor is included in the state vector. The representative frames for the sequence *Human8* are illustrated in Fig. 2c, and the target has a significant change in illumination from frame #100 to #124. Only the proposed tracker keeps track of the target. This is because LBP feature descriptor handles illumination variation to which color histogram is susceptible. There is a sudden change in illumination at frames #41, #132 and #229 for sequence *Singer2* as depicted in Fig. 2d. Sample frame for *Boy* is shown in Fig. 2e. When there is variation in scale at frames #484 and #575, trackers, viz. ASLA, WMIL, CT, MIL and PF, drift away from the target, while the proposed tracker tracks successfully.

Discriminative trackers CT and MIL are not able to handle the drastic change in illumination. These trackers considered Haar-like features for object representation which are prone to illumination variations. However, the proposed tracker is robust to illumination variations as it considers complementary cues, viz. color and texture. When RGB color histogram performance degrades due to illumination variations, the scale invariant LBP texture histogram will compensate. In addition, the particles which are affected by the illumination and scale variations are suppressed by the proposed multi-cue fusion model and, hence, enhance the tracking accuracy.

4.1.2 Full or partial occlusion

The sample output for video sequences considered under this challenge is presented in Fig. 3. The *Football* sequence has several full and partial occlusions as shown in Fig. 3a. WMIL, PF-PSO and PF-BSA fail to track the target when it undergoes full occlusion at frame #153, while ours tracker performs comparatively better. Figure 3b depicts partial occlusions of the target at frame #190. CT, MIL and FRAG have shown unsatisfied performance under partial occlusion. On the other hand, our tracker with adaptive multi-cue fusion model boosts the high performing particles during occlusion and, hence, prevents drift. At frame #70, target has full occlusion by the pillar as shown in Fig. 3c. The trackers, viz. ASLA, MIL, WMIL, CT, FRAG, IVT and PF, are not sufficiently able to handle this challenge as they are not able to alleviate from drift. While ours perform effectively under this challenge. Critical frames for *Subway* sequence are shown in Fig. 3d. At frame #98, ASLA and WMIL lost the target when it is partially occluded by the other person. Ours

Table 2 Details of video sequences for experiments

Video sequences	Resolution	Frames	Challenge
CarScale	640 × 272	252	SV
Dog	352 × 240	127	SV, DEF, RO
Crossing	360 × 240	120	BC
Couple	320 × 240	140	BC, DEF, RO
Human8	320 × 240	128	SV, IV
Singer2	624 × 352	366	IV, SV, DEF, RO
Crowds	600 × 480	347	FOC, POC,
Jogging2	352 × 288	307	FOC, POC
Subway	352 × 288	175	POC, FOC
Walking	768 × 576	412	BC
Mountainbike	640 × 360	228	BC
Football	624 × 352	362	FOC, POC, RO
Jogging1	352 × 288	307	FOC, POC
Walking2	384 × 288	500	BC
Bolt2	480 × 270	293	DEF, RO
Boy	640 × 480	602	IV,SV

along with MIL and WMIL has shown relatively good performance in comparison with the other trackers. In Fig. 3e, at frame #84, when target recovered from full occlusion, CT, FRAG, MIL and IVT lost the target. Ours keep track of the target as multi-cues, viz. color and texture, are exploited for object appearance model.

The proposed tracker has better performance under full or partial occlusion which is mainly attributed to the outlier detection mechanism and the proposed resampling technique. The particles affected due to occlusion are identified by the outlier detection mechanism. The proposed resampling technique diversified these particles in the search space in order to improve their placement. This helps in maintaining the tracker's efficiency under full or partial occlusion.

4.1.3 Background clutters

We have considered challenging video sequences under this challenge, and certain frames are presented in Fig. 4. In *Walking* video sequence at frame #125, PF drifts away

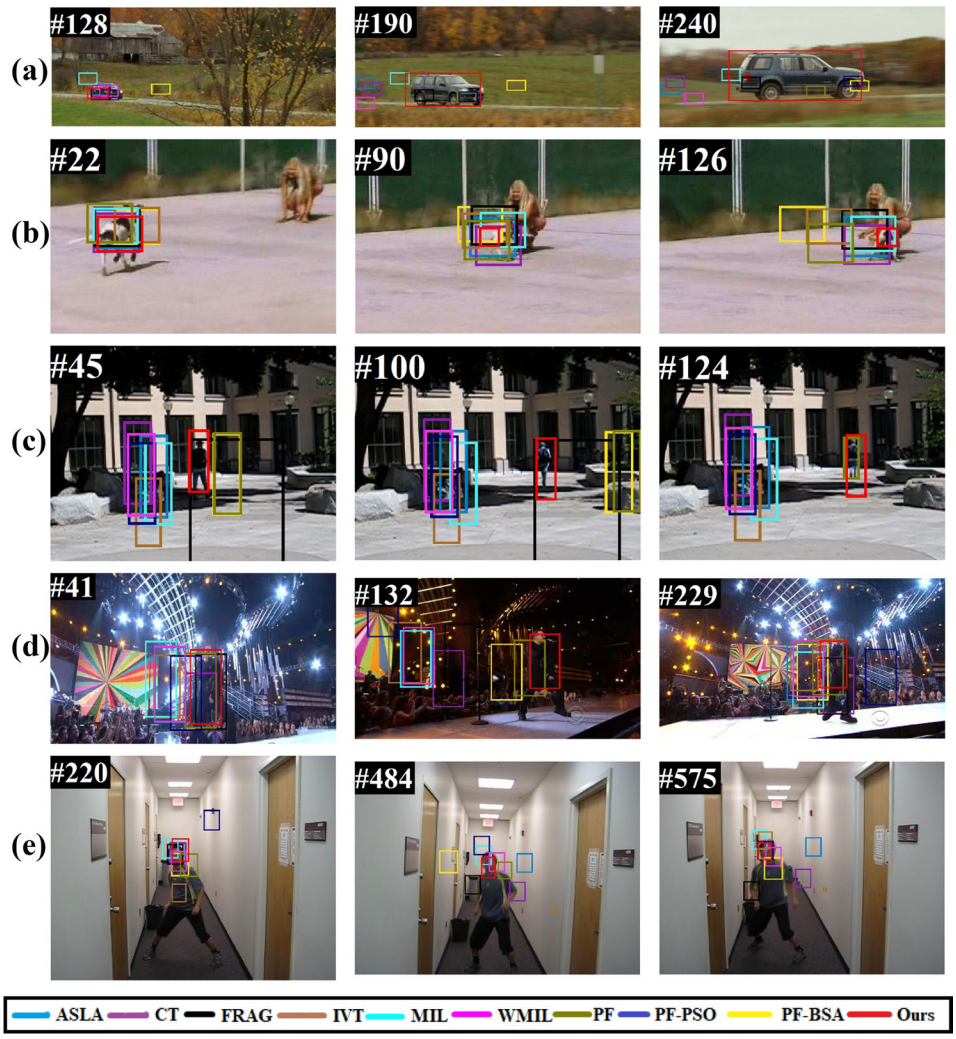


Fig. 2 Critical frames with illumination and scale variations video sequences: **a** *CarScale*, **b** *Dog*, **c** *Human8*, **d** *Singer2* and **e** *Boy*

from the target as shown in Fig. 4a. PF-BSA and ASLA also drift away from the target as they have no mechanism to relocate the target but ours, MIL and FRAG successfully track the target. In Fig. 4b, critical frames for *MountainBike* sequences are shown. MIL, WMIL, CT, FRAG, PF and PF-BSA are inefficient to handle the similar background. However, ours, ASLA and MIL are able to track the target. In our tracker, the outlier detection mechanism identified the unimportant particles due to the similar background and, hence, prevents false updation of particles in our tracker. *Crossing* critical frames are presented in Fig. 4c. At frame #94, after the car passes in front of the target, ASLA, CT, FRAG, PF-BSA and PF deflected away from the target. On the other hand, our tracker has shown substantially good performance in the sequence as proposed adaptive multi-cue fusion supports boosting of good particles during the background clutters. The frames for sequence *Couple* are shown in Fig. 4d. After frame # 90, when the target passes by similar color car in the background, then only ours with FRAG can be able to track the target efficiently. Figure 4e depicts sample frames for *Walking2* sequence. At frame #224, CT, MIL, PF, PF-BSA and PF-PSO lost the target, while ours and ASLA track the target efficiently. The transductive reliability values of the color and texture prevent the drift of target during background clutters, and the re-detection strategy also supports in enhancing the tracking accuracy.

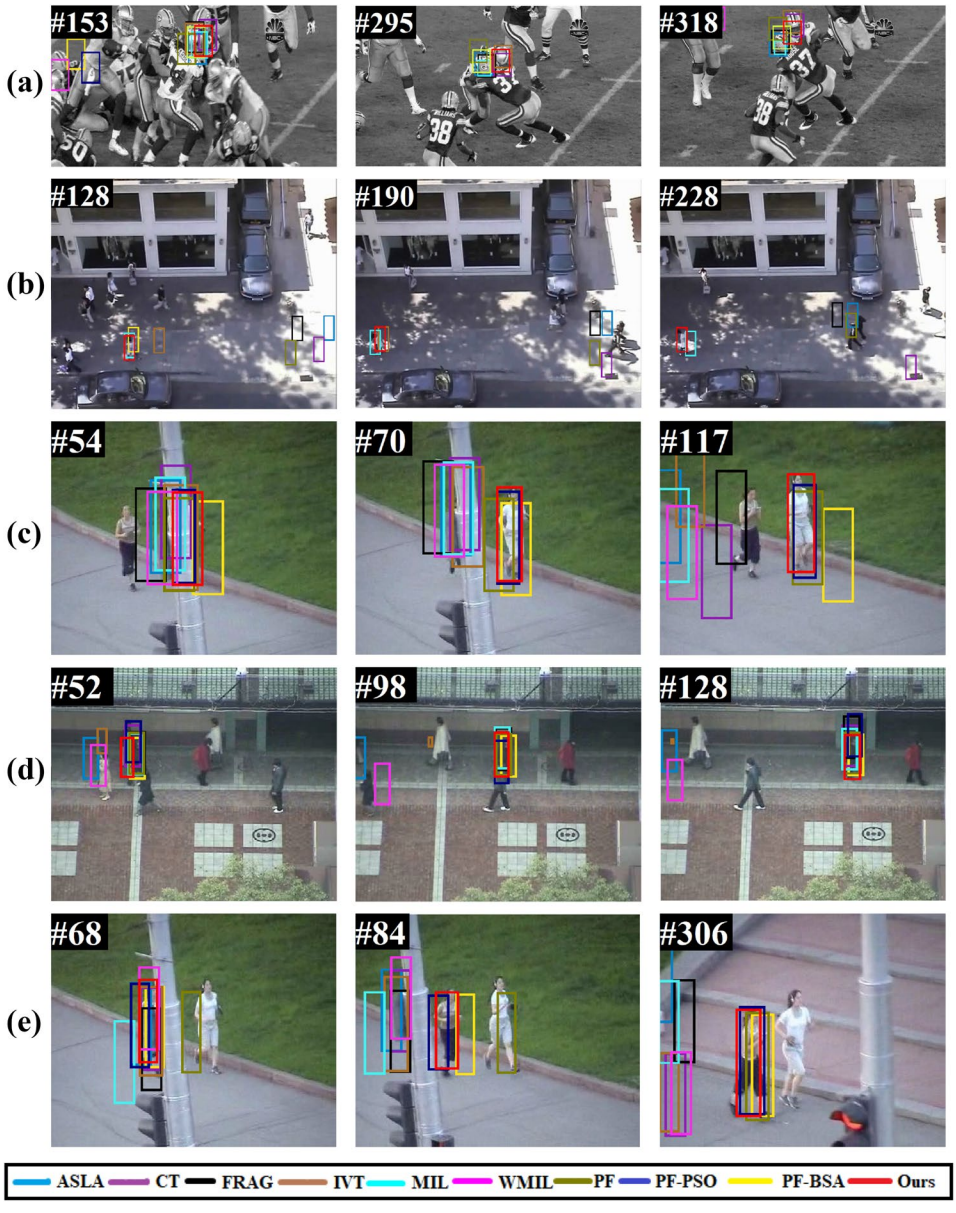


Fig. 3 Critical frames with full or partial occlusion for video sequences: **a** *Football*, **b** *Crowds*, **c** *Jogging2*, **d** *Subway* and **e** *Jogging1*

4.1.4 Object's deformation and rotation

The sample frames for the challenging video sequences under this challenge are presented in Fig. 5. In *Singer2* video sequence at frames #89 and #171, the target deformed due to its sudden pose variations and subsequent rotation. MIL, CT, WMIL, ASLA, CT and PF-BSA have lost the target. However, ours has shown better results. In Fig. 5b, critical frames for *Football* sequences are shown. Non-rigid deformation occurs in the target due to its in-plane rotation at frames #140. PF, PF-PSO and PF-BSA are unable to handle this deformation, while our tracker manages to track the target. The sample frames for *Dog* sequence are shown in Fig. 5c. At frames #12 and #46, the target has deformation due to its fast motion and rotation. MIL, FRAG, PF and PF-BSA are slightly drifted from the target, while ours, ASLA and CT perform efficiently. *Couple* whose sample frames are presented in Fig. 4d. At frame #100, ASLA, CT and PF-BSA fail to keep track of the target as deformation occurs due to sudden camera movement. Ours, FRAG and IVT keep track of the target. However, at frame #108, only our tracker is able to locate the target. It is mainly due to the rotational component which is incorporated in the state vector of the multi-component state model. In addition, the low performing particles due to object's deformation are suppressed by the multi-cue fusion model in order to cater this challenge. Frames for sequence *Bolt2* are depicted in Fig. 4e. At frame #190, when target suffered from deformation due to in-plane rotation, ours, CT and PF-PSO have shown relatively better performance in comparison with other trackers.

Failure cases: Experimental results validate the robustness of the proposed method over other state-of-the-art trackers. However, there are certain frames in video sequences in which the proposed tracker deflected from the target. As depicted in Fig. 6a, b, our tracker is not able to locate the target when there is target's deformation due to its abrupt

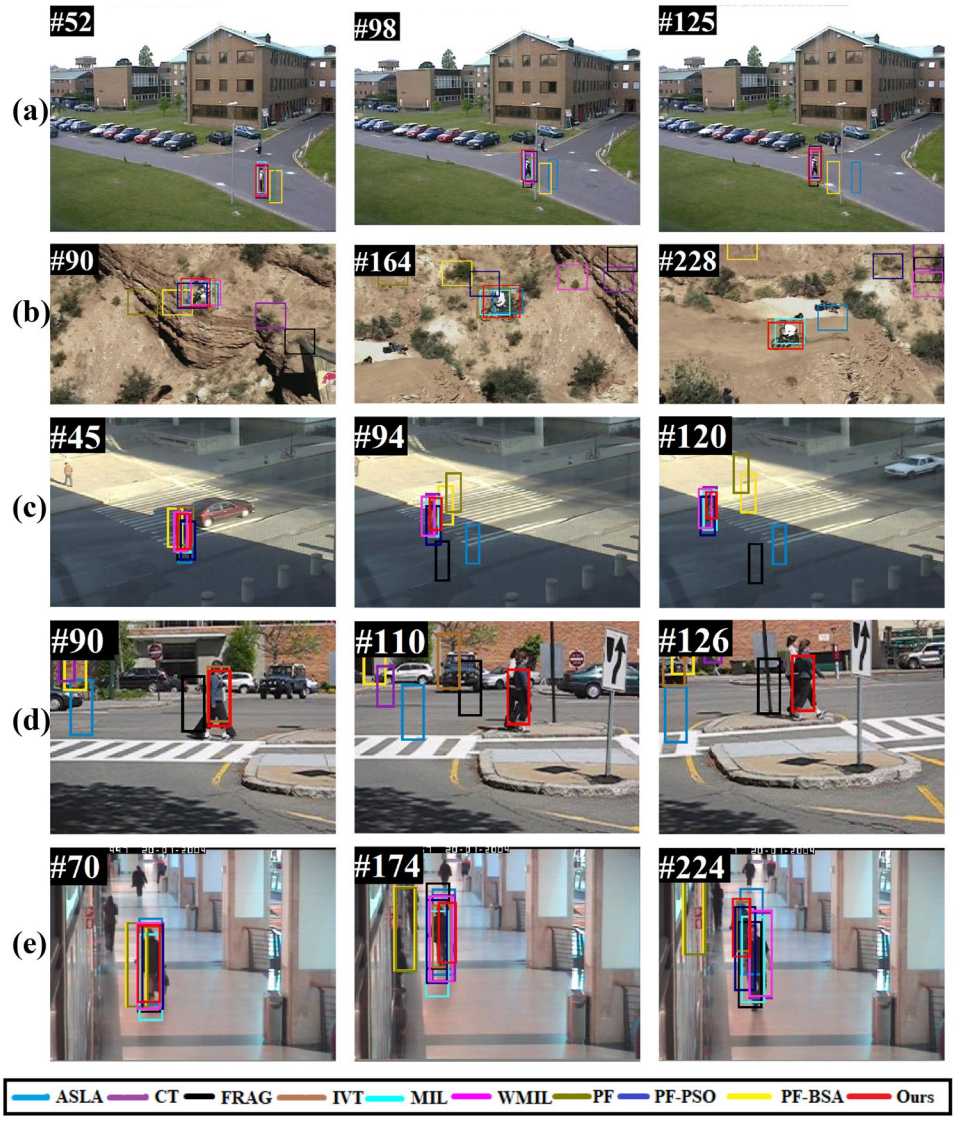


Fig. 4 Critical frames with background clutters for video sequences: **a** *Walking*, **b** *MountainBike*, **c** *Crossing*, **d** *Couple* and **e** *Walking2*

motion. The reason may be due to target's rotation, particles scattered in the search space, and their contribution toward state estimation is reduced. Figure 6c shows failure of our tracker in sequence *Mountainbike* when heavy background clutters are accompanied by in-plane and out-of-plane rotation. Our tracker lost the target for *Walking* sequence when object deformation occurs due to large-scale variations and is represented in Fig. 6d. In this situation, scaling component of the multi-component state model failed to handle the target's scale variations. In sum, the proposed tracker was able to handle scaling and rotational variations in certain videos but failed to render when it is accompanied by fast motion and motion blur.

4.2 Quantitative analysis

Quantitative performance of the proposed tracker is analyzed by considering multiple evaluation metrics which includes center location error(CLE), *F* Measure [21] and plots as precision vs. Location error threshold and success rate vs. overlap threshold [37]. CLE is calculated as the distance between tracker bounding box and the center of the ground truth. *F* Measure is defined as: $\frac{2\times pr\times rl}{pr+rl}$ where pr is the precision and calculated as: $\frac{B_R \cap B_G}{B_R}$and rl is the recall and given as: $\frac{B_R \cap B_G}{B_G}$, where B_R is the tracker bounding box and B_G the ground truth bounding box. Precision plot illustrates the mean precision at various location error thresholds (in pixels). Success plot represents the percent success over various overlap thresholds (in %).

We have achieved an average CLE of 10.99 and average *F* Measure of 0.683 when evaluated on 16 challenging video sequences taken from benchmark datasets [37]. CLE and *F* measure are tabulated in Tables 3 and 4, respectively. Figure 7 represents the percentage overlap of successful frames from a given threshold from the ground truth. Figure 8 illustrates the success rate over overlap threshold. As shown in Fig. 7, for 20px location error threshold (in pixels), we achieve an average precision of 0.90 for illumination and scale variations, 0.98 for full or partial occlusion, 0.91 for background clutters and 0.82 for object's deformation and rotation. As depicted in Fig. 8, at 0.2 overlap threshold (in

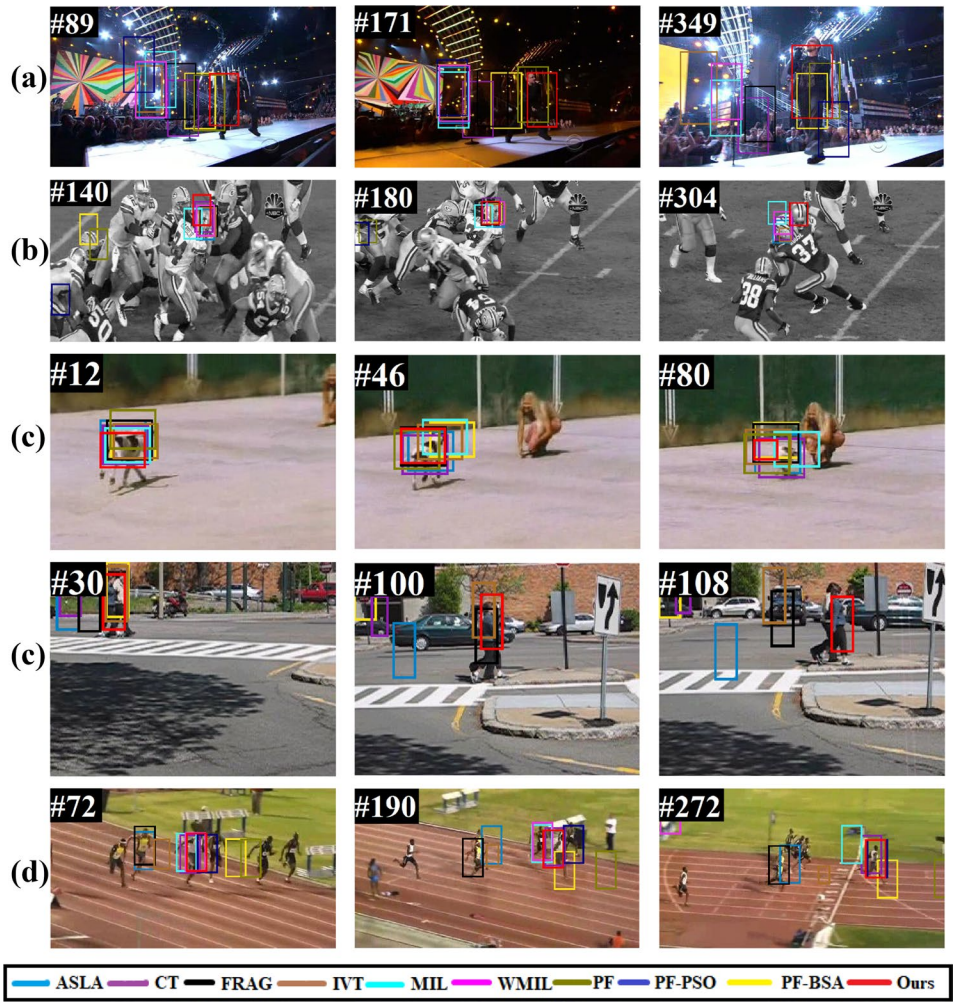


Fig. 5 Critical frames with object's deformation and rotation for video sequences: **a** *Singer2,* **b** *Football*, **c** *Dog*, **d** *Couple* and **e** *Bolt2*

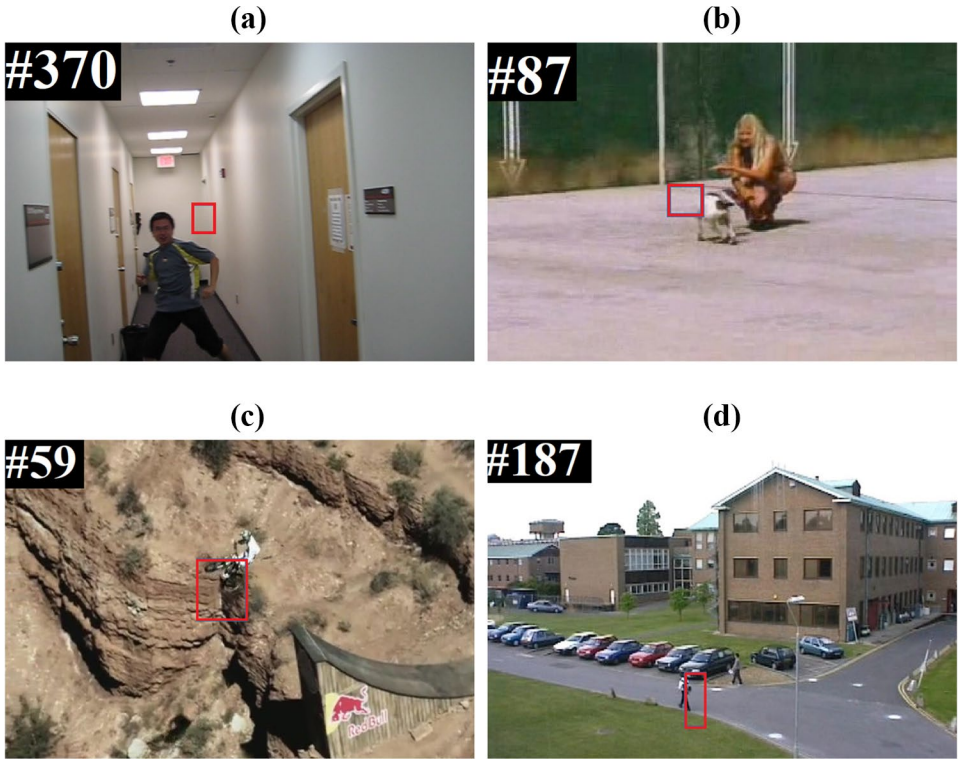


Fig. 6 Failure cases for the proposed tracker for video sequence: **a**, *Boy*, **b** *Dog*, **c** *Mountainbike and* **d** *Walking*

%), our tracker manages average success rate 0.92 for illumination and scale variations as well as for full or partial occlusion, 0.96 for background clutters and 0.90 for object's deformation and rotation. High values of precision and success rate reveal that our tracker can handle various dynamic environment challenges. The proposed tracker exploited complementary multi-cues, viz. color and texture, to handle illumination variations and background clutters. Scaling and rotational component of the multi-component state model handles the object's deformation due to scale variations and

Table 3 Center location error (in pixels)

Video sequence	ASLA	CT	FRAG	IVT	MIL	WMIL	PF	PF-PSO	PF-BSA	Ours (-outlier)	Ours
CarScale	79.80	75.14	31.21	**11.58**	70.97	70.33	135.14	40.15	97.06	17.02	*14.34*
Human8	65.56	82.65	55.65	90.01	62.77	79.26	32.70	71.90	30.12	*15.49*	**5.58**
Singer2	193.91	121.21	88.44	171.09	161.69	160.98	122.00	55.49	56.49	*27.06*	**14.85**
Walking	247.76	8.55	9.54	**2.30**	*4.38*	8.63	81.70	15.72	32.87	12.93	7.35
Mountainbike	21.73	192.44	206.90	**7.21**	*7.78*	107.11	165.75	94.76	85.63	16.81	14.89
crowds	295.40	380.72	–	11.04	261.56	249.34	110.91	*10.49*	146.51	15.43	**9.08**
Jogging2	143.90	116.98	37.42	132.33	134.43	138.36	46.52	20.48	17.73	*9.16*	**8.07**
Crossing	33.41	5.37	38.50	*3.18*	**3.16**	5.12	22.66	12.23	16.19	6.87	4.54
Couple	96.23	79.30	**8.94**	100.50	27.68	39.78	41.69	42.56	43.98	16.98	*15.22*
Dog	*9.21*	**8.33**	–	46.81	13.42	25.13	25.44	13.51	37.26	10.50	9.94
Subway	145.04	11.58	8.50	117.19	6.70	136.73	9.71	11.34	*5.73*	6.07	**4.97**
Football	15.30	15.63	**5.48**	14.86	27.35	14.51	275.40	201.03	234.12	17.37	*10.62*
Boy	43.72	37.86	40.39	92.44	*17.29*	**14.60**	81.44	96.83	68.72	34.39	27.72
Jogging1	99.36	92.95	21.62	83.22	95.90	94.37	44.82	14.75	21.88	*13.16*	**12.38**
Walking2	19.98	63.20	57.74	**4.98**	55.39	59.90	84.10	60.32	50.31	10.48	*7.21*
Bolt2	102.92	13.03	116.80	75.12	28.13	119.07	85.81	19.45	40.12	*12.06*	**9.08**
Average	100.83	81.56	51.94	60.24	61.16	82.70	85.36	48.81	61.55	*13.96*	**10.99**

The first , second and third best results are highlighted in bold, italic and underline, respectively

Table 4 *F* measure

Video sequence	ASLA	CT	FRAG	IVT	MIL	WMIL	PF	PF-PSO	PF-BSA	Ours (-outlier)	Ours
CarScale	0.460	0.432	0.471	0.002	0.172	0.474	0.060	0.300	0.191	*0.666*	**0.702**
Human8	0.130	0.063	0.218	0.049	0.071	0.087	0.379	0.156	0.384	*0.733*	**0.749**
Singer2	0.053	0.138	0.246	0.049	0.048	0.064	0.215	0.384	0.417	*0.718*	**0.751**
Walking	0.069	0.643	0.684	**0.874**	*0.685*	0.659	0.037	0.493	0.172	0.380	0.550
Mountainbike	0.633	0.222	0.149	**0.840**	*0.821*	0.436	0.171	0.428	0.211	0.712	0.728
crowds	0.071	0.004	–	0.075	0.059	0.005	0.056	*0.602*	0.280	0.450	**0.609**
Jogging2	0.133	0.092	0.569	0.158	0.138	0.133	0.254	0.590	0.565	*0.813*	**0.820**
Crossing	0.380	0.779	0.349	0.401	**0.837**	0.660	0.327	0.596	0.447	0.747	*0.787*
Couple	0.056	0.237	**0.692**	0.081	0.100	*0.526*	0.367	0.484	0.138	0.501	0.542
Dog	0.500	0.492	–	0.109	0.481	0.378	0.202	0.487	0.244	**0.600**	*0.587*
Subway	0.194	0.699	*0.757*	0.183	**0.801**	0.189	0.321	0.605	0.572	0.747	*0.758*
Football	0.6312	0.612	**0.814**	0.636	0.321	0.593	0.111	0.127	0.185	0.542	*0.692*
Boy	0.316	0.458	0.431	0.282	*0.500*	**0.559**	0.104	0.366	0.192	0.293	0.398
Jogging1	0.161	0.191	0.618	0.193	0.159	0.138	0.155	0.629	0.377	*0.690*	**0.718**
Walking2	0.464	0.315	0.329	0.686	0.315	0.344	0.131	0.333	0.320	*0.707*	**0.777**
Bolt2	0.024	0.645	0.016	0.022	0.327	0.387	0.036	0.565	0.254	*0.695*	**0.759**
Average	0.267	0.376	0.453	0.290	0.365	0.352	0.183	0.447	0.309	*0.625*	**0.683**

The first, second and third best results are highlighted in bold, italic and underline, respectively

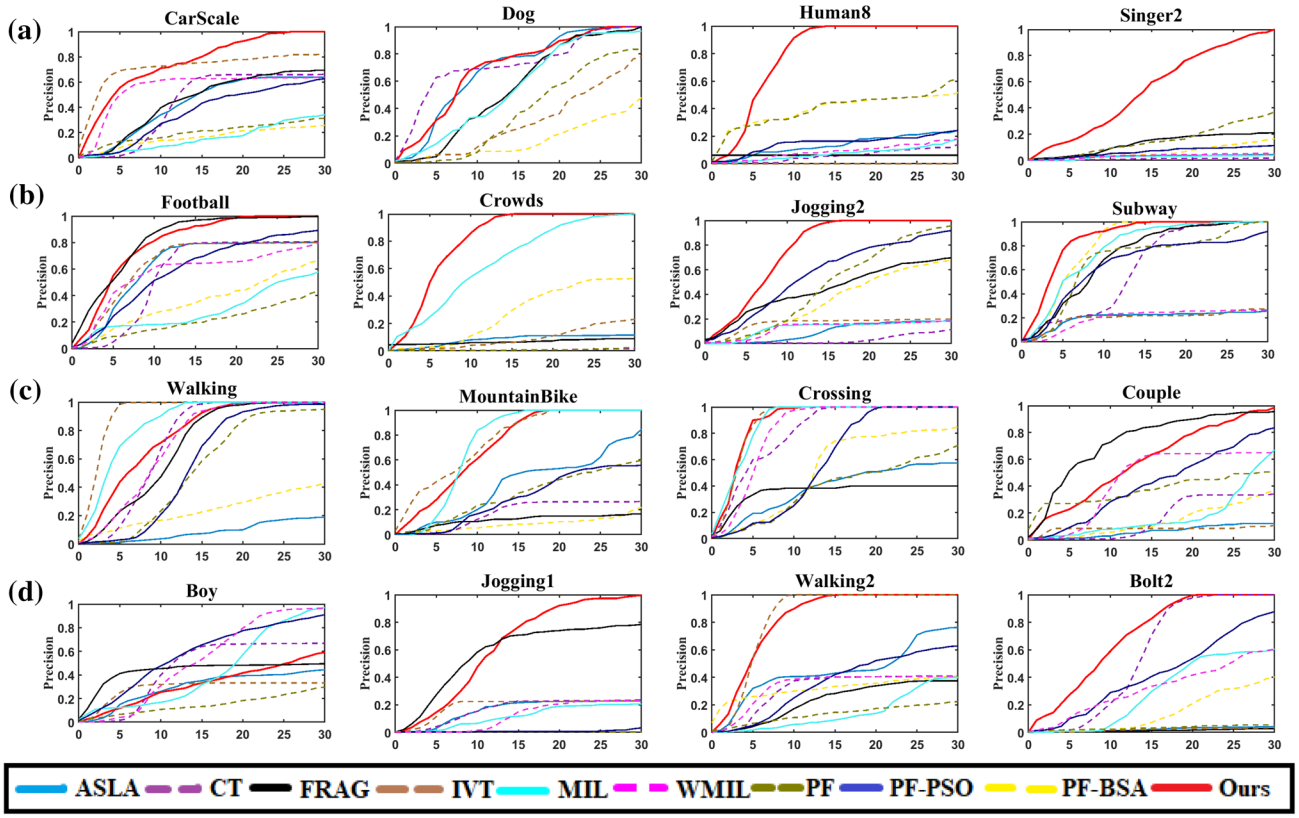


Fig. 7 Precision plot *vs*. CLE threshold (in pixels) under various visual tracking challenges: **a** illumination and scale variations (*CarScale*, *Dog*, *Human8*, *Singer2*, *Boy*), **b** full and partial occlusion (*Football*, *Crowds*, *Jogging-2*, *Subway*, *Jogging1*), **c** background clutters (*Walking*, *MountainBike*, *Crossing*, *Couple*, *Walking2*) and **d** object's deformation and rotation (*Singer2*, *Football*, *Dog*, *Couple*, *Bolt2*)

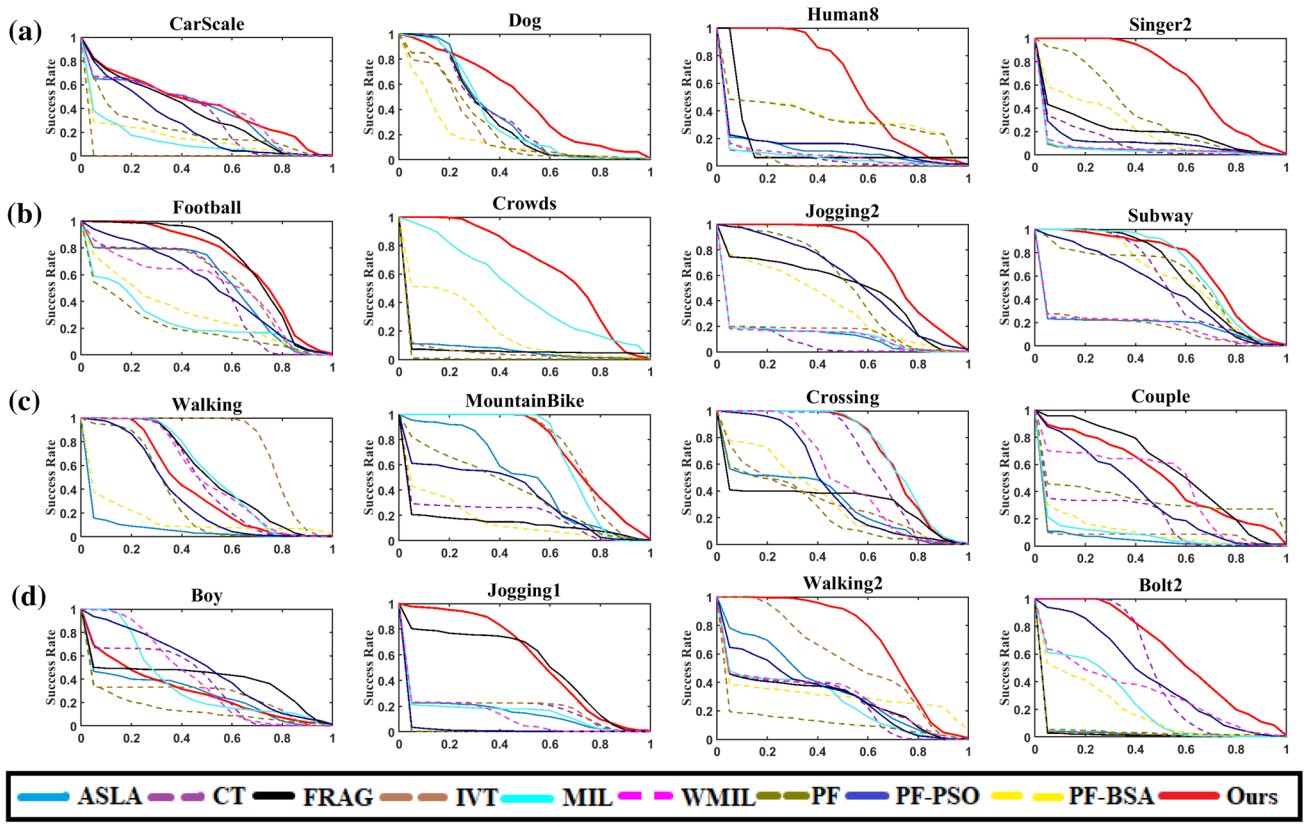


Fig. 8 Success versus percent overlap threshold (%) plot under various visual tracking challenges: **a** illumination and scale variations (*CarScale*, *Dog*, *Human8*, *Singer2*, *Boy*), **b** full and partial occlusion (*Football*, *Crowds*, *Jogging-2*, *Subway*, *Jogging1*), **c** background clutters (*Walking*, *MountainBike*, *Crossing*, *Couple*, *Walking2*) and **d** object's deformation and rotation (*Singer2*, *Football*, *Dog*, *Couple*, *Bolt2*)

rotation. Automatic boosting and suppression of particle by fusion model improve tracker's efficiency during full or partial occlusion. On the other hand, generative tracker such as ASLA [19] is not robust against full or partial occlusion as its appearance model does not consider background information. Discriminative trackers, viz. CT [39] and MIL [5], fail to keep track of target in background clutter and illumination variation. WMIL [38] is considered random Haar-like features for object representation and, hence, not able to tackle sudden illumination and scale variations. IVT [26] performance is limited by a large number of parameters used for developing the appearance model of the target. PF-based trackers such as PF [16], PF-PSO [40] and PF-BSA [8] are not efficient enough to address dynamic challenges, mainly due to single cue and non-optimal solution for sample degeneracy and impoverishment problem. Our tracker has exploited CSA as a resampling technique which ensures the diversity in the search space with two parameters: awareness probability and flight length. The outlier detection mechanism helps in identifying the low performing particles. The complexity of the proposed resampling technique is reduced as it resampled only the low performing particles detected as outlier. Figure 9 depicts the effect of outlier threshold κ_{th} on the final performance of the tracker. It has been inferred that the proposed tracker has obtained best performance when $\kappa_{th} = 0.82$. Table 5 tabulates the re-acquisition mechanism employed by the proposed method. Sequence *Boy* has scale variations accompanied by fast motion and motion blur and, hence, has maximum re-detection in comparison with other videos. On average of outcome, the proposed tracker achieves computational speed of 21 frames/sec when sampling only the outlier. However, the speed has been reduced to 14 frames/sec when all the particles are resampled. The significant improvement in computational efficiency of the proposed tracker reveals that our tracker achieves high accuracy and robustness against challenges in real time. In sum, qualitative and quantitative validation of the proposed tracker on benchmarked video reveals that proposed method can not only efficiently handle various tracking challenges but also envisage applications for real-time tracking of target's in video sequences.

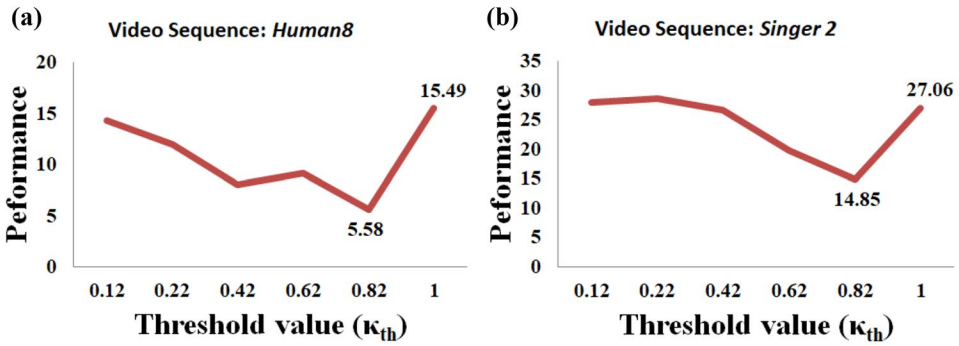


Fig. 9 Effect of threshold values on the proposed tracker for video sequence: **a** *Human8* **and b** *Singer2*

Table 5 Re-acquisition details of the proposed tracker

Video sequence	Total frames	Re-acquisition
CarScale	252	1
Human8	128	0
Singer2	366	0
Walking	412	2
Mountainbike	228	2
crowds	347	1
Jogging2	307	0
Crossing	120	0
Dog	127	2
Subway	175	0
Football	362	1
Boy	602	4
Jogging1	307	0
Walking2	500	0
Bolt2	293	1

5 Conclusion

In this paper, we propose a robust tracker based on multi-cue particle filter framework which comprises of adaptive multi-cue fusion model and a novel resampling method to overcome particle degeneracy. Complementary multi-cues, viz. RGB color and LBP texture, are considered to enhance the tracker accuracy under illumination variation, occlusion, background clutters, object's deformation and rotation. The adaptive multi-cue fusion model either boosts or suppresses the particles in order to ensure clear decision boundary between the unimportant particles and the significant particles. Nevertheless, an outlier detection mechanism is also proposed to classify the unimportant particles from the significant one. The proposed tracker is computationally efficient as only the unimportant particles are resampled by the proposed resampling technique. Quick adaptation of tracker is ensured through online estimation of transductive reliability of each cue. Qualitative and quantitative analysis on 16 challenging video sequences from benchmark infers that the proposed tracker performs favorably against the state-of-the-art trackers.

In future, we look forward to integrating more multi-modal features in the proposed tracking framework. Appearance model can also be made more robust against rotational variations. The inclusion of user-defined source importance in the proposed fusion model is another direction for the future undertaking. We would also like to extend our work to tracking multiple people in video sequences. Online model for guiding particles under uncertain conditions can also be investigated.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflicts of interest.

References

1. Adam A, Rivlin E, Shimshoni I (2006) Robust fragments-based tracking using the integral histogram. In: 2006 IEEE CS conference on computer vision and pattern recognition, vol 1, pp 798–805
2. Ahmadi K, Salari E (2016) Social-spider optimised particle filtering for tracking of targets with discontinuous measurement data. IET Comput Vis 11(3):246–254
3. Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans Signal Process 50(2):174–188
4. Askarzadeh A (2016) A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm. Comput Struct 169:1–12
5. Babenko B, Yang MH, Belongie S (2011) Robust object tracking with online multiple instance learning. IEEE Trans Pattern Anal Mach Intell 33(8):1619–1632
6. Bai L, Wang Y, Fairhurst M (2010) Multiple filters for road detection and tracking. Pattern Anal Appl 13(3):251–262
7. Bai Y, Wang D (2006) Fundamentals of fuzzy logic control fuzzy sets, fuzzy rules and defuzzifications. In: Wang D, Bai Y, Zhuang H (eds) Advanced fuzzy logic technologies in industrial applications. Springer, Berlin, pp 17–36

8. Bhateja A, Walia GS, Kapoor R (2016) Non linear state estimation using PF based on backtracking search optimization. In: International conference on computer, communication and automation, pp 342–347
9. Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability distributions. Bull Calcutta Math Soc 35:99–109
10. Bolić M, Djurić PM, Hong S (2004) Resampling algorithms for particle filters: a computational complexity perspective. EURASIP J Adv Signal Process 2004(15):403686
11. Brasnett P, Mihaylova L, Bull D, Canagarajah N (2007) Sequential Monte Carlo tracking by fusing multiple cues in video sequences. Image Vis Comput 25(8):1217–1227
12. Choe G, Wang T, Liu F, Hyon S, Ha J (2014) Particle filter with spline resampling and global transition model. IET Comput Vis 9(2):184–197
13. Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. IEEE Trans Pattern Anal Mach Intell 25:564–577
14. Gao ML, Li LL, Sun XM, Yin LJ, Li HT, Luo DS (2015) Firefly algorithm based particle filter method for visual tracking. Optik Int J Light Electron Opt 126:1705–1711
15. Gordon N, Ristic B, Arulampalam S (2003) Beyond the Kalman filter: particle filters for tracking applications, vol 3. Artech House, London, pp 1077–2626
16. Gordon NJ, Salmond DJ, Smith AF (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: IEE proceedings F (radar and signal processing), vol 140, pp 107–113. IET
17. Han H, Ding YS, Hao KR, Liang X (2011) An evolutionary particle filter with the immune genetic algorithm for intelligent video target tracking. Comput Math Appl 62:2685–2695
18. Isard M, Blake A (1998) Condensation–conditional density propagation for visual tracking. Int J Comput Vis 29:5–28
19. Jia X, Lu H, Yang MH (2012) Visual tracking via adaptive structural local sparse appearance model. In: IEEE conference on computer vision and pattern recognition, pp 1822–1829
20. Khasnabish N, Detroja KP A (2018) Stochastic resampling based selective particle filter for visual object tracking. In: Indian control conference (ICC) 2018. IEEE, pp 42–47
21. Lazarevic-McManus N, Renno J, Makris D, Jones GA (2008) An object-based comparative methodology for motion detection based on the f measure. Comput Vis Image Underst 111:74–85
22. Li T, Sun S, Sattar TP, Corchado JM (2014) Fight sample degeneracy and impoverishment in particle filters: a review of intelligent approaches. Expert Syst Appl 41(8):3944–3954
23. Murphy RR (1996) Biological and cognitive foundations of intelligent sensor fusion. IEEE Trans Syst Man Cybern Part A Syst Hum 26(1):42–51
24. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
25. Rohilla R, Sikri V, Kapoor R (2016) Spider monkey optimisation assisted PF for robust object tracking. IET Comput Vis 11(3):207–219
26. Ross DA, Lim J, Lin RS, Yang MH (2008) Incremental learning for robust visual tracking. Int J Comput Vis 77(1–3):125–141
27. Sardari F, Moghaddam ME (2016) An object tracking method using modified galaxy-based search algorithm. Swarm Evolut Comput 30:27–38
28. Sardari F, Moghaddam ME (2017) A hybrid occlusion free object tracking method using particle filter and modified galaxy based search meta-heuristic algorithm. Appl Soft Comput 50:280–299
29. Stauffer C, Grimson WEL (1999) Adaptive background mixture models for real-time tracking. In: IEEE CS conference on computer vision and pattern recognition, vol 2, pp 246–252
30. Walia GS, Kapoor R (2014) Intelligent video target tracking using an evolutionary particle filter based upon improved cuckoo search. Expert Syst Appl 41(14):6315–6326
31. Walia GS, Kapoor R (2016) Recent advances on multicue object tracking: a survey. Artif Intell Rev 46:1–39
32. Walia GS, Kapoor R (2016) Robust object tracking based upon adaptive multi-cue integration for video surveillance. Multimed Tools Appl 75(23):15821–15847
33. Walia GS, Raza S, Gupta A, Asthana R, Singh K (2017) A novel approach of multi-stage tracking for precise localization of target in video sequences. Expert Syst Appl 78:208–224
34. Wang D, Lu H, Xiao Z, Yw Chen (2013) Fast and effective color-based object tracking by boosted color distribution. Pattern Anal Appl 16(4):647–661
35. Wang Z, Liu Z, Liu W, Kong Y (2011) Particle filter algorithm based on adaptive resampling strategy. In: 2011 international conference on electronic and mechanical engineering and information technology (EMEIT), vol 6, pp 3138–3141
36. Weng SK, Kuo CM, Tu SK (2006) Video object tracking using adaptive Kalman filter. J Vis Commun Image Represent 17(6):1190–1208
37. Wu Y, Lim J, Yang MH (2013) Online object tracking: a benchmark. In: IEEE conference on computer vision and pattern recognition, pp 2411–2418
38. Zhang K, Song H (2013) Real-time visual tracking via online weighted multiple instance learning. Pattern Recognit 46:397–411
39. Zhang K, Zhang L, Yang MH (2012) Real-time compressive tracking. In: European conference on computer vision. Springer, pp 864–877
40. Zhao J, Li Z (2010) Particle filter based on particle swarm optimization resampling for vision tracking. Expert Syst Appl 37:8910–8914
41. Zhou H, Deng Z, Xia Y, Fu M (2016) A new sampling method in particle filter based on Pearson correlation coefficient. Neurocomputing 216:208–215
42. Zuo J (2013) Dynamic resampling for alleviating sample impoverishment of particle filter. IET Radar Sonar Navig 7:968–977

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

Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Role of Co-Solvent in the Super Critical Esterification process of Bio-Diesel Production through Karanja oil

To cite this article: Chandra Shekhar Singh and Naveen Kumar 2019 *J. Phys.: Conf. Ser.* **1240** 012159

View the article online for updates and enhancements.



IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 14.139.251.100 on 13/08/2019 at 11:31

Role of Co-Solvent in the Super Critical Esterification process of Bio-Diesel Production through Karanja oil

Chandra Shekhar Singh[1], Naveen Kumar[2]

[1,2]CASRAE, Delhi Technological University
rchandrashekhar85@gmail.com

Abstract. Biofuel is necessity for the today's era for developing country like India to fulfil its growing energy demand and concern towards the environment pollution. Bio fuel production involve following major steps such as preparation of feed stock, selection of conversion technology and the up-gradation techniques. Conventional method of bio diesel production is of time consuming and sensitive to the feed stock quality. And due to these limitations in conventional esterification processes, Supercritical Esterification has now become one of the most attractive conversion technologies for bio diesel production due to its flexibility towards the feed stock quality, high conversion rate and fast reaction time. Although, it is costlier process as compare to the conventional processes because of its extreme condition of pressure and temperature requirement. So in the present study, the Super Critical Trans-esterification of Karanja (Pongamia Pinnata) oil with Methanol is investigated. And the effect of co-solvent on the reaction condition is studied. It has been observed from the results that use of co-solvent could be a potential means to reduce the critical condition required in the Super Critical Esterification process for the bio diesel production.

Keywords: Critical Point, Trans-esterification, Solvent

1. Introduction

Biofuel is necessity for the today's era for developing country like India to fulfill its growing energy demand and concern towards the environment pollution. Biodiesel is one of the renewable sources to fill the gap. Biodiesel are known for their low PM (particulate matter) value, low emissions of CO and low Sulfur content[1]. Cetane number for biodiesel are also at higher side in compare to conventional diesel[2]. Vegetable oil like Mustard, Canola, Corn, Soybean, Palm and many more may be used to produce the Biodiesel but their cost and edibility makes them economically unviable. Second generation oil like Karanja(Pongamia Pinnata),Jatropha are now considered as the source for the biodiesel production[3].Biofuel production involve following major steps such as preparation of feed stock, selection of conversion technology and the up-gradation.

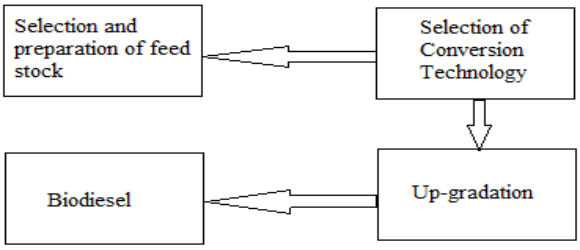


Figure 1 Biodiesel Production Step

Bio diesel from vegetable oil can be converted by different methods like Pyrolysis, Dilution, Micro-emulsification and trans-esterification[4]. In trans-esterification process fat or oil react with alcohol

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

which results in form of esters and glycerides[5][6]. In the process, a triglyceride molecule or fatty acid neutralize the free fatty acids removes the glycerin and creates an alcohol ester[7]. Figure 2 is showing the actual reaction that takes place during the trans-esterification of vegetable oil.

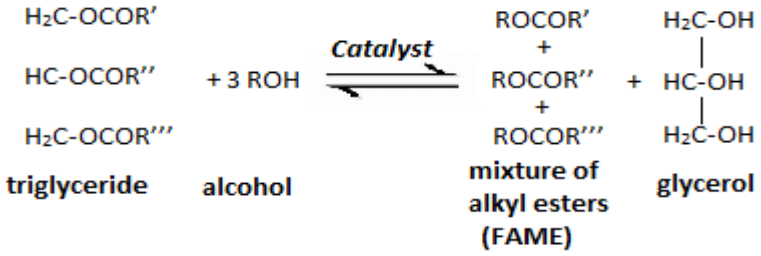


Figure 2 Trans-esterification reaction of vegetable oil[8]

n the trans-esterification reaction larger amount of alcohol is used to shift equilibrium towards the production of more alkyl esters. The reaction may be slow and need to be fastening to increase the yield. In that case catalyst is used to fast the reaction and yields. Trans-esterification process performed in the presence of base acid or enzyme catalyst [6]. Selection of conversion technology depends upon the quality of feed stock such as acid value and water content[9].
In base catalytic trans-esterification, presence of water may cause saponification process that will lead the lower yields of esters. Separation of catalysts is also a problem. Although, Heterogeneous base catalyzed technology may reduce the cost of the catalyst separation. But this method also required to keep the water content in feed stock at very low level because of its sensitivity towards the water concentration[8].Acid catalyst may be an alternative to avoid the separation of the catalyst if the reaction is taking place homogeneously[10].Although, acid catalyzed reaction rate is low. Supercritical method is flexible towards the feed stock quality and free from the separation difficulty of catalysts[5]. Saka and Kusdiana were reported the first time production of biodiesel under supercritical conditions. They were reported biodiesel production with 75-95% yield within 2 minutes at 350°C, 450 bar and M:O molar ratio of 42:1.

Table -1(Super critical Trans-esterification reaction condition of non-edible Vegetable oils[4][15][16][17][18])

Non-Edible oil	**Alcohol**	**Oil to alcohol molar ratio**	**Temperature and Pressure**	**Time (min.)**	**Conversion rate(%)**
Jatropha Curcas	Methanol	01:40	350°C,200 bar	40	> 90
Cottonseedoil	Ethanol	01:41	250°C	8	85
Linseed oil	Ethanol	01:41	250°C	8	>85
Sesame oil	Ethanol	01:40	325°C, 20 Mpa	40	100
Pongamia Pinnata	Ethanol	01:50	400°C, 20 Mpa	30	90

Super critical esterification of vegetable oil is catalyst free weather it is homogeneous or heterogeneous[11]. SAT (supercritical alcohol transestrification) may be defined as the process in which oil and alcohol(solvent) undergo above their critical point of pressure and temperature. Methanol and Ethanol are mostly used and proved alcohols for the super critical esterification

process[12]. Super critical esterification process is very fast and yield is also very high in comparison of the conventional processes. In supercritical condition, the increase of reaction temperature resulted in higher yield irrespective of feed stock quality[13][14]. Additionally, in this method purification steps have also become easier as no catalysts are generally used. The trans-esterification reaction condition of different non-edible vegetable oils under supercritical condition tried by various researchers is shown in Table 1.

Despite of these advantages, Supercritical trans- esterification are costly as it requires high pressure and temperature which in turn increases the equipment cost and energy consumption[19].The use of co-solvent could be a potential means to reduce the critical condition required in the Super Critical Trans-esterification process for the bio diesel production as well as lowering the cost of production. Co-solvent such as Carbon dioxide, propane, hexane and calcium oxide may be used to lower the extreme condition required[16]. So in the present study, the Super Critical Trans-esterification of Karanja (Pongamia Pinnata) oil with Methanol was investigated. And the effects of co-solvent on the reaction condition were studied.[13]

2. Supercritical Esterification of Karanja Oil for producing Biodiesel

Supercritical method of trans-esterification are required very extreme condition of temperature and pressure [20][21][22][23].These extreme conditions makes the choice of equipment material very limited and so increases the initial cost of set-up. Despite of these, it has the advantages of being fast and tolerant towards the feed stock quality. It also facilitate the simultaneous trans-esterification and esterification of triglyceride and FFA respectively. Temperature and pressure as reported, affecting the distribution of biodiesel, glycerol, methanol and triglycerides. Kusdiana and Saka[24] were the first one to produce biodiesel through supercritical method and to explain the reaction mechanism in super critical condition. In supercritical method of trans-esterification process alcohol to oil ratio, temperature, pressure play very important role in the yielding of the biodiesel. It has been reported by researcher that in most of the supercritical methods of bio diesel production the reaction took place around 300°C to 400°C range of temperature and pressure range was around 200 bar[25][24]. Alcohol to oil ratio for various vegetable oil for optimum level was in the range of 40:1 to 42:1 [26][3][15]. Few reports also had shown that solubility of methanol was increased by 2%-3% on 10 °C increase of temperature[27]. But the above mentioned conditions are very severe and would cost very high in case of actual production. So, it is now a need to reduce the conditions required for the reaction and the use of co-solvent may be the one of the option to minimize the harsh condition required. Different types of co-solvents have been proposed such as carbon dioxide, propane, n-butane, n-heptane, n-hexane and many more has to be explored. It is reported that in biodiesel production through supercritical esterification with CO2 as co-solvent increases the yielding of methyl ester in case of canola oil bio diesel production[28]. A researcher is also reported that 0.1 molar ratio of propane reduced the extreme condition from 350 °C to 280 °C to obtain full conversion.

In the present paper, a lab scale supercritical trans-esterification of Karanja oil was conducted to determine the effect of co-solvent on the conditions required and yielding of bio diesel.

Method

The super critical trans-esterification was conducted in a tube of stainless steel[29]. The volume of steel vessel was 86 ml volume. The tubular reactor was made by stainless steel of grade SS 316 which is sufficiently hard and had strength enough to sustain the high temperature and pressure during the reaction. The schematic diagram of the experimental setup is given in figure 3.The experiment's optimum condition was predefined for the Karanja oil with methanol in terms of molar ratio. During the experiment on set-up , when the reactor temperature reaches around the range of 300 to 400 °C then the pressure was noted around the 200 bar. The reactor was put into the heating bath to rise and maintained its temperature and time duration was recorded. Then reactor was cooled into the cold water bath to stop the further reactions. After 30 minutes for settle down, the fatty acid alkyl esters and glycerol were separated and analyzed.

The experiment was conducted in two phases, in first case non catalytic super critical-esterification was conducted without any co-solvent and result was investigated. In the second phase the non-catalytic super critical trans-esterification reaction or experiment were conducted with CO_2 as co-solvent and the result were compared with the previous one.

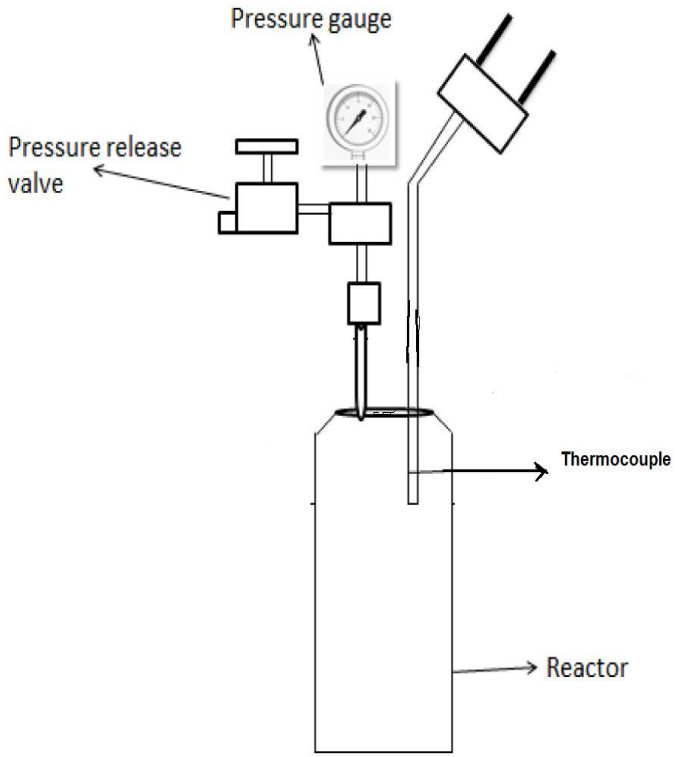


Figure 3 Experiment Setup

for the study, locally purchased Karanja oil was used for supercritical esterification. The Methanol was fed in pure form and filled into the reactor vessel with the help of injection pump. Volume flow rate was calculated during the filling to maintain the methanol oil ratio. Important summary of reaction conditions are given in table 2.

Table 2 (Experimental Condition for experiment)

Type of Condition	**Experimental Range**
Feed stock	Karanja Oil
Reaction time	5- 30 min.
Oil to alcohol molar ratio	1: 41
Temperature	200-400 °C
Pressure	100-200 bar

2.2 Result and Discussion

The molar ratio of methanol, ethanol to oil plays very important role for the conversion of Esters. In the present study, when solvent was not used then for more than 90% conversion into methyl ester, the reaction temperature required was 350°C. The molar ratio was kept as 41:1 as suggested by various researchers for optimum yield. Due to high molar ratio, as temperature increases the pressure was also increasing in vessels. It was found that around 200 bars and within 12 minutes of time optimum yield in supercritical esterification was obtained. The conversion percentage with respect to the temperature when molar ratio was 41:1, can be observed in figure 4. For the conversion of 80%,the temperature required was 350 °C and the pressure was around 200 bars. From the above data, it is evident for more than 80% conversion a very severe condition is required which in turn increases the cost. While the introduction of CO2 as co-solvent was reducing the harsh condition required as compare to without it. Co-solvent was reducing the critical point of methanol oil mixture. Figure 5 shows that with CO2 as co-solvent, high yield of biodiesel can obtain at milder conditions.

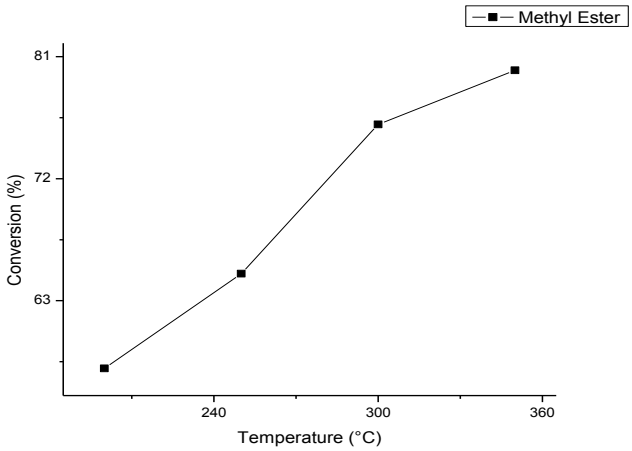


Figure-4 Conversion percentage into Methyl Ester as function of Temperature without co-solvent

As it can be observed from the figure 5 that with the use of the CO_2 as co-solvent significant 67% yielding was obtained at the milder temperature around 200 °C as compare to the condition without co-solvent for the corresponding yield. More than 90 % conversion was found to be around the temperature 300 °C and conversion was also increased as compare to without Co-solvent.

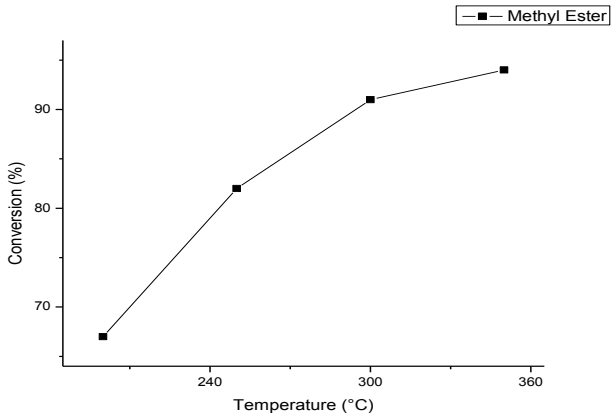


Figure 5 Conversion into Methyl Ester as function of Temperature with co-solvent

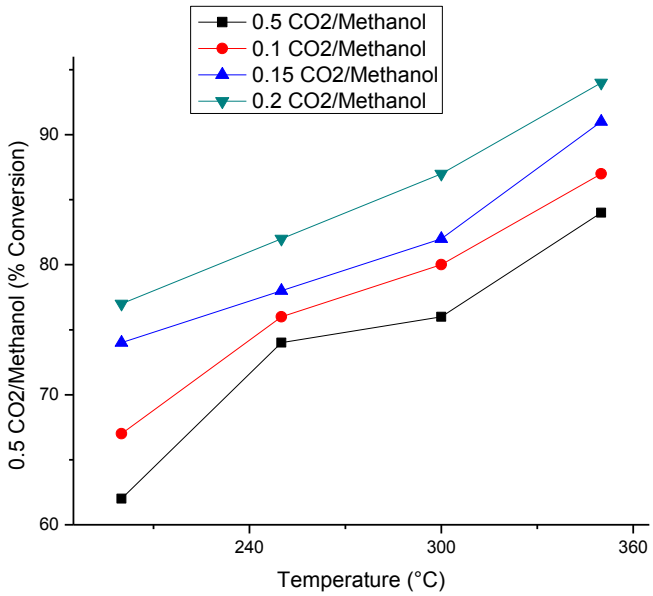


Figure-6 Effect of co-solvent to Methanol molar ratio on yielding.

From the figure 5 it can be seen that further increasing the temperature above 300°C there is no significant yielding in terms of conversion. After setting these parameters, again experiment was conducted with different molar ratio of CO2 to Methanol to investigate the effects on reaction. From the figure 6 it can be seen that when molar ratio is increased from 0.05 to 0.20, the conversion is starting from the 77% and obtaining conversion around 94% when the temperature reaches to the level of 350°C. From the figure, it can notice that temperature has positive effect on the yielding.

3. Conclusion

Present investigation is supporting the results of various researchers for the use of co-Solvent may be the one option to reduce the extreme conditions required. It can be observed from the results use of CO2 as co-solvent decreasing the extreme reaction conditions required and improving the reaction quality. Solubility of methanol in Karanja oil is increasing at low temperature with the use of co-solvent. Co-solvent is also decreasing the critical point of methanol so it resulting in milder condition required for the whole reaction process. Use of Co-solvent is enhancing the solubility between methanol and triglycerides at mild conditions and increases the biodiesel production. In the present study, a significant conversion around 70 to 80 % had been found at mild conditions as compare to the super critical biodiesel production without the solvent use. It was also observed that by increasing the molar ratio of CO2 to methanol, have the positive effect on the conversion rate.

References:

[1] O. Farobie and Y. Matsumura, “State of the art of biodiesel production under supercritical conditions,” Prog. Energy Combust. Sci., vol. 63, pp. 173–203, 2017.
[2] Ö. Can, “Combustion characteristics, performance and exhaust emissions of a diesel engine fueled with a waste cooking oil biodiesel mixture,” Energy Convers. Manag., vol. 87, pp. 676–686, 2014.
[3] R. L. Patel and C. D. Sankhavara, “Biodiesel production from Karanja oil and its use in diesel engine: A review,” Renew. Sustain. Energy Rev., vol. 71, no. December, pp. 464–474, 2017.
[4] M. M. K. Bhuiya, M. G. Rasul, M. M. K. Khan, N. Ashwath, and A. K. Azad, “Prospects of 2nd generation biodiesel as a sustainable fuel - Part: 1 selection of feedstocks, oil extraction techniques and conversion technologies,” Renew. Sustain. Energy Rev., vol. 55, pp. 1109–

1128, 2016.
[5] V. I. Anikeev and E. Y. Yakovleva, “Biodiesel synthesis from vegetable oils with supercritical methanol,” J. Supercrit. Fluids, vol. 77, pp. 100–102, 2013.
[6] B. Singh, N. Kumar, and H. Muk, “A study on the performance and emission of a diesel engine fueled with Jatropha biodiesel oil and its blends,” Energy, vol. 37, no. 1, pp. 616–622, 2012.
[7] A. Demirbas, “Biodiesel production via non-catalytic SCF method and biodiesel fuel characteristics,” Energy Convers. Manag., vol. 47, no. 15–16, pp. 2271–2282, 2006.
[8] S. Semwal, A. K. Arora, R. P. Badoni, and D. K. Tuli, “Biodiesel production using heterogeneous catalysts,” Bioresour. Technol., vol. 102, no. 3, pp. 2151–2161, 2011.
[9] N. America, “Methyl Esterification of Free Fatty Acids of Rapeseed Oil as Treated in Supercritical Methanol,” vol. 34, no. 3, pp. 383–387, 2001.
[10] S. Bensaid, D. Hoang, P. Bellantoni, and G. Saracco, “Supercritical fluid technology in biodiesel production: Pilot plant design and operation,” Green Process. Synth., vol. 2, no. 5, pp. 397–406, 2013.
[11] A. Demirbas, “Biodiesel fuels from vegetable oils via catalytic and non-catalytic supercritical alcohol transesterifications and other methods : a survey,” vol. 44, pp. 2093–2109, 2003.
[12] M. Balat and H. Balat, “Progress in biodiesel processing,” Appl. Energy, vol. 87, no. 6, pp. 1815–1835, 2010.
[13] A. Demirbas, “Biodiesel production from vegetable oils via catalytic and non-catalytic supercritical methanol transesterification methods,” Prog. Energy Combust. Sci., vol. 31, no. 5–6, pp. 466–487, 2005.
[14] A. Demirbas, “Comparison of transesterification methods for production of biodiesel from vegetable oils and fats,” vol. 49, pp. 125–130, 2008.
[15] V. Rathore and G. Madras, “Synthesis of biodiesel from edible and non-edible oils in supercritical alcohols and enzymatic synthesis in supercritical carbon dioxide,” Fuel, vol. 86, no. 17–18, pp. 2650–2659, 2007.
[16] V. B. Borugadda and V. V Goud, “Biodiesel production from renewable feedstocks : Status and opportunities,” Renew. Sustain. Energy Rev., vol. 16, no. 7, pp. 4763–4784, 2012.
[17] A. Galia, O. Scialdone, and E. Tortorici, “Transesterification of rapeseed oil over acid resins promoted by supercritical carbon dioxide,” J. Supercrit. Fluids, vol. 56, no. 2, pp. 186–193, 2011.
[18] M. M. Gui, K. T. Lee, and S. Bhatia, “Supercritical ethanol technology for the production of biodiesel: Process optimization studies,” J. Supercrit. Fluids, vol. 49, no. 2, pp. 286–292, 2009.
[19] M. Y. Koh, T. Idaty, and M. Ghazi, “A review of biodiesel production from Jatropha curcas L . oil,” Renew. Sustain. Energy Rev., vol. 15, no. 5, pp. 2240–2251, 2011.
[20] M. Tobar and G. A. Núñez, “Supercritical transesterification of microalgae triglycerides for biodiesel production: Effect of alcohol type and co-solvent,” J. Supercrit. Fluids, vol. 137, no. November 2017, pp. 50–56, 2018.
[21] P. S. Montcho, C. Konfo T. R., P. Agbangnan D. C., A. Sidouhounde, and D. Sohounhloue C. K., “Comparative Study of Transesterification Processes for Biodiesel Production (A Review),” Elixir Appl. Chem. , vol. 120, no. July, pp. 51235–51242, 2018.
[22] M. González Prieto, M. Fortunatti Montoya, P. E. Hegel, and S. Pereda, “Supercritical reactors for the production of advanced bio-fuels: A review,” J. Supercrit. Fluids, vol. 134, pp. 106–113, 2018.
[23] O. Aboelazayem, M. Gadalla, and B. Saha, “Biodiesel production from waste cooking oil via supercritical methanol: Optimisation and reactor simulation,” Renew. Energy, vol. 124, pp. 144–154, 2018.
[24] S. Saka and D. Kusdiana, “Biodiesel fuel from rapeseed oil as prepared in supercritical methanol,” vol. 80, pp. 225–231, 2001.
[25] D. Wen, H. Jiang, and K. Zhang, “Supercritical fluids technology for clean biofuel production,” Prog. Nat. Sci., vol. 19, no. 3, pp. 273–284, 2009.

[26] A. Abbaszaadeh, B. Ghobadian, M. R. Omidkhah, and G. Najafi, "Current biodiesel production technologies : A comparative review," Energy Convers. Manag., vol. 63, pp. 138–148, 2012.
[27] M. N. Varma and G. Madras, "Synthesis of biodiesel from castor oil and linseed oil in supercritical fluids," Ind. Eng. Chem. Res., vol. 46, no. 1, pp. 1–6, 2007.
[28] H. Imahara, E. Minami, S. Hari, and S. Saka, "Thermal stability of biodiesel in supercritical methanol," vol. 87, pp. 1–6, 2008.
[29] K. T. Tan, M. M. Gui, K. T. Lee, and A. R. Mohamed, "An optimized study of methanol and ethanol in supercritical alcohol technology for biodiesel production," J. Supercrit. Fluids, vol. 53, no. 1–3, pp. 82–87, 2010.

AIP Conference Proceedings

Stabilization of plane polarized Alfven waves by anomalous Doppler resonance

Cite as: AIP Conference Proceedings **2136**, 060015 (2019); https://doi.org/10.1063/1.5120961
Published Online: 09 August 2019

Rajesh Gupta, Suresh C. Sharma, and Ruby Gupta


View Online

Export Citation

ARTICLES YOU MAY BE INTERESTED IN

Excitation of kinetic reverse shear Alfven eigenmode by energetic ions in a tokamak plasma
AIP Conference Proceedings **2136**, 060016 (2019); https://doi.org/10.1063/1.5120962

Non-equilibrium study of Coulomb glass at small disorders using Kawasaki dynamics
AIP Conference Proceedings **2136**, 020002 (2019); https://doi.org/10.1063/1.5120907

Self-focusing of Hermite-Gaussian laser beam in plasma in relativistic and ponderomotive regime
AIP Conference Proceedings **2136**, 060017 (2019); https://doi.org/10.1063/1.5120963



AIP Publishing

© 2019 Author(s).

Stabilization of Plane Polarized Alfven Waves by Anomalous Doppler Resonance

Rajesh Gupta[1,a)], Suresh C. Sharma[1,b)] And Ruby Gupta[2,c)]
[1]*Department of Applied Physics, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi-110 042, India.*
[2]*Department of Physics, Swami Shraddhanand College, University of Delhi, Alipur, Delhi-110036, India.*

[a)]rudratheshakti@rediffmail.com
[b)]sureshsharma321@gmail.com
[c)]rubyguptassn@gmail.com

Abstract: This work presents non-linear decay of plane polarized Alfven wave via anomalous Doppler resonance with an electron beam in a magnetized plasma. First order perturbation theory has been employed to derive the expressions for dispersion relation of Alfven waves and their growth rate. Numerical calculations have also been carried out for growth rate and unstable mode frequencies for typical plasma parameters. There exists two modes of wave propagation, one having frequency $< \omega_{ci}$ corresponds to Alfven wave, and the other has frequency approaching ω_{ce} which corresponds to electron cyclotron wave. The waves interact with the electron beam via normal and anomalous Doppler resonance. The normal resonance interaction increases the frequency of plane polarized Alfven wave but shows no growth or decay in amplitude. However, in anomalous resonance interaction the frequency remains unchanged and the wave stabilizes. The variation of frequency and growth rate with magnetic field and number density of plasma electrons is also discussed in the present paper. The phase velocity and the unstable frequency of the wave decrease with an increase in the number density of plasma electrons. The results of our work are applied to explain some of the experimental observations.

1. Introduction

In 1942 Hannes Alfven proposed the existence of an Electromagnetic-Hydrodynamic wave which would carry energy from the photosphere to heat up the corona and the solar wind in nature. After the Argus Nuclear Test, Berthold Harris and Hope detected Alfven waves in the ionosphere in 1958. Alfven waves are the slowest and have the lowest frequencies of any known electromagnetic wave. Such waves require the presence of magnetic field. The Alfven wave is the dominant low frequency transverse mode of magnetized plasma. The Alfven wave propagates along the magnetic field and displays a continuous spectrum. This wave may play an important role in energy transport, in driving field-aligned currents, in particle acceleration and heating, and in explaining inverted-V structure in magnetosphere – ionosphere coupling [1].

The interaction of electron beam or ion beam with plasma waves in magnetized plasma can excite many instabilities, which are helpful for understanding the stability confinement and heating of the plasma [2, 3]. In plasma, with an electron beam introduced parallel to the magnetic field, electrostatic ion cyclotron wave can easily be destabilised [4]. Gupta *et al* [5] have discussed analytically the whistler wave interaction with an electron beam propagating through magnetized plasma for oblique as well as parallel propagation. Shevchenko et al [6] have studied low frequency Alfven wave excitation, saturation and subsequent nonlinear evolution by an ion beam in a plasma. They also studied how the steepened wave profiles grew out of low frequency Alfven fluctuations. Gupta *et al* [7] have shown that growth rate of the instability increases with the beam density via Cerenkov interaction and fast cyclotron interaction when surface plasma waves are excited. Yin *et al* [8] have investigated, the generation of kinetic Alfven waves by ion-ion streaming along the magnetic field at relatively low Alfven speed and low beta plasma. They have found electron parallel heating increases both the wave growth rates and the range of unstable modes of the instability. Shukla and Stenflo [9] have shown that Alfven and magnetosonic wave can be non linearly excited by a high frequency external pump wave. They have illustrated that finite amplitude Alfven waves can either interact with the background plasma or with themselves, giving rise to a number of nonlinear phenomenons such as wave amplitude modulation or filamentation, density profile modification as well as self organization in vortical structure. Voitenko *et al* [10] have studied non linear excitation of kinetic Alfven wave and whistler wave by electron driven Langmuir waves. They have shown that the resonant pairs of Kinetic Alfven wave and whistler wave are nonlinearly coupled to the pump Langmuir waves and their amplitude undergo exponential growth from the thermal level. Tripathi *et al* [11] have experimentally demonstrated the excitation of travelling shear Alfven Wave in a laboratory magnetoplasma. They found that the ambient plasma density and electron temperature were significantly enhanced by the beam.

International Conference on Photonics, Metamaterials & Plasmonics
AIP Conf. Proc. 2136, 060015-1–060015-6; https://doi.org/10.1063/1.5120961
Published by AIP Publishing. 978-0-7354-1879-0/$30.00

Chen *et al* [2] have investigated kinetic Alfven wave driven by a fast electron beam in finite beta plasma. Their results show that the kinetic resonant interaction of beam electron is the driving source for kinetic Alfven Wave instability. Shukla and Sharma [12] have studied non linear decay of a beam driven plasma mode into an Alfven wave plus a circularly polarized electromagnetic wave near the electron plasma frequency.

In the present paper, we study the excitation of Alfven waves by an electron beam injected parallel to the magnetic field in magnetized plasma. In section 2, we study instability analysis and calculate the growth rate of Alfven waves for slow and fast cyclotron interactions with electron beam. Results and discussion are given in section 3 and finally the conclusion is given in section 4.

2. Instability Analysis

Consider a plasma immersed in a d.c magnetic field $\mathbf{B_s} \parallel \hat{\mathbf{z}}$, with electron number density n_{eo}, ion number density n_{io}, electron mass m_e, ion mass m_{io}, electron charge -e and ion charge e. An electromagnetic Alfven wave propagates through it, with electric field

$$\mathbf{E} = \mathrm{A}e^{-i(\omega t - k_z z)}.$$

The magnetic field of the wave $\mathbf{B} = c\mathbf{k}_z \times \mathbf{E}/\omega$.

An electron beam with density n_{bo} flows through the plasma along the z-direction, parallel to magnetic field.
The equation of motion, governing the drift velocity of plasma electrons, plasma ions and beam electrons is

$$m\left[\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\,\mathbf{v}\right] = -e\,\mathbf{E} - \frac{e}{c}\mathbf{v}\times(\mathbf{B_s} + \mathbf{B}). \tag{1}$$

In equilibrium (i.e. in the absence of the wave), $\mathbf{v_e} = 0$ and $\mathbf{v_i} = 0$. When wave is present, **E** and **B** of the wave are treated as small or perturbed quantities and result in perturbed velocities $\mathbf{v_{1e}}$, $\mathbf{v_{1o}}$ and $\mathbf{v_{1be}}$ for plasma electrons, plasma ions and beam electrons, respectively.
On linearizing Eq. (1), we obtain

$$\frac{\partial \mathbf{v_1}}{\partial t} = -\frac{e\,\mathbf{E}}{m} - \mathbf{v_1}\times\hat{\mathbf{z}}\,\omega_{ce}, \tag{2}$$

where $\omega_{ce} = \dfrac{e\mathbf{B_s}}{m_e c}$.

The number density of electrons and ions have been calculated by using equation of continuity

$$\frac{\partial n}{\partial t} + \nabla.\left(n\mathbf{v}\right) = 0. \tag{3}$$

Current densities of different components have been calculated by formula

$$\mathbf{J} = ne\mathbf{v}. \tag{4}$$

The wave equation used for obtaining dispersion relation is

$$\nabla^2\mathbf{E} - \nabla\left(\nabla.\mathbf{E}\right) + \frac{\omega^2}{c^2}\mathbf{E} = -\frac{4\pi i\omega}{c^2}\mathbf{J}. \tag{5}$$

Now we consider a plane polarized Alfven wave propagating along magnetic field i.e. $\mathbf{E} = E_x\,\hat{\mathbf{x}}$ and $\mathbf{k} = k_z\,\hat{\mathbf{z}}$. Writing x, y, z-components of Eq. (2), we obtain the perturbed plasma electron velocities as

$$v_{1ex} = -\frac{e(-i\omega)E_x}{m_e\left(\omega_{ce}^2 - \omega^2\right)}, \tag{6}$$

$$v_{1ey} = -\frac{e\,E_x\omega_{ce}}{m_e\left(\omega_{ce}^2 - \omega^2\right)}, \tag{7}$$

and $v_{1ez} = 0$. (8)

Substituting Eqs. (6), (7) and (8) in Eq. (4), we get the perturbed plasma electron current densities as

$$J_{1ex} = n_{eo}\frac{e^2}{m_e}\frac{(-i\omega)E_x}{\left(\omega_{ce}^2 - \omega^2\right)}, \tag{9}$$

and $$J_{1ey} = n_{eo}\frac{e^2}{m_e}\frac{\omega_{ce}E_x}{\left(\omega_{ce}^2 - \omega^2\right)}. \tag{10}$$

The perturbed plasma ion current densities are obtained from Eqs. (9) & (10), by replacing e by -e, ω_{ce} by $-\omega_{ce}$, and m_e by m_i.
Similarly, the perturbed beam electron velocities are obtained using Eq. (2) as:

$$v_{1bx} = -\frac{e}{m_e}\frac{\left(-i\bar{\omega}^2\right)E_x}{\omega\left(\omega_{ce}^2-\omega^2\right)}, \tag{11}$$

$$v_{1by} = -\frac{e}{m_e}\frac{\bar{\omega}\omega_{ce}E_x}{\omega\left(\omega_{ce}^2-\omega^2\right)} \quad , \tag{12}$$

and $v_{1ez} = 0$. (13)

where $\bar{\omega} = \omega - k_z v_{bo}$.

The perturbed beam electron current densities obtained using Eq. (4) are

$$J_{1bx} = \frac{e^2}{m_e}\frac{n_{bo}\left(-i\bar{\omega}^2\right)E_x}{\omega\left(\omega_{ce}^2-\omega^2\right)}, \tag{14}$$

and $$J_{1by} = \frac{e^2}{m_e}\frac{n_{bo}\bar{\omega}\omega_{ce}E_x}{\omega\left(\omega_{ce}^2-\omega^2\right)} \tag{15}$$

Substituting Eqs. (9), (10), (14) and (15) in Eq. (5),
we get

$$\left[-k_z^2+\frac{\omega^2}{c^2}+\frac{\omega_{pe}^2\omega^2}{c^2\left(\omega_{ce}^2-\omega^2\right)}+\frac{\omega_{pi}^2\omega^2}{c^2\left(\omega_{ce}^2-\omega^2\right)}+\frac{\omega_{pb}^2\bar{\omega}^2}{c^2\left(\omega_{ce}^2-\omega^2\right)}\right]E_x = 0$$

or

$$\omega^2\left[P+\frac{\omega_{pi}^2}{\left(\omega_{ce}^2-\omega^2\right)}\right]-k_z^2c^2 = \frac{\omega_{pb}^2\bar{\omega}^2}{\left(\omega_{ce}^2-\bar{\omega}^2\right)} \quad , \tag{16}$$

where $P = 1+\frac{\omega_{pe}^2}{\omega^2}$.

For Alfven waves $\omega_{ce} << \omega$, Eq. (16) becomes

$$-P\omega^4+(P\omega_{ci}^2+\omega_{pi}^2+k_z^2c^2)\omega^2 - k_z^2c^2\omega_{ci}^2 = \frac{\omega_{pb}^2\bar{\omega}^2}{\left(\omega_{ce}^2-\bar{\omega}^2\right)}$$

Or

$$\left(\omega^2-\omega_1^2\right)\left(\omega^2-\omega_2^2\right) = \frac{\omega_{pb}^2\bar{\omega}^2}{\left(\omega_{ce}^2-\bar{\omega}^2\right)} \quad , \tag{17}$$

where ω_1 and ω_2 are roots of Eq.(17) given as

$$\omega_1 = \left[\frac{P\omega_{ci}^2+\omega_{pi}^2+k_z^2c^2}{2P} - \frac{\left[\sqrt{\left(P\omega_{ci}^2+\omega_{pi}^2+k_z^2c^2\right)^2-4Pk_z^2c^2\omega_{ci}^2}\right]}{2P}\right]^{1/2} \tag{18}$$

$$\omega_2 = \left[\frac{P\omega_{ci}^2+\omega_{pi}^2+k_z^2c^2}{2P} + \frac{\left[\sqrt{\left(P\omega_{ci}^2+\omega_{pi}^2+k_z^2c^2\right)^2-4Pk_z^2c^2\omega_{ci}^2}\right]}{2P}\right]^{1/2}$$

Here ω_1 corresponds to Alfven wave frequency and ω_2 corresponds to electron cyclotron frequency. In cyclotron interaction $\bar{\omega}^2 = \omega_{ce}^2$; we have slow cyclotron interaction or anomalous Doppler resonance for $\bar{\omega} \approx -\omega_{ce}$, and fast cyclotron interaction or normal resonance for $\bar{\omega} \approx \omega_{ce}$. Considering fast cyclotron interaction, we rewrite Eq. (17) as

$$\left(\omega - \omega_1\right)\left(\bar{\omega} - \omega_{ce}\right) = -\frac{\omega_{pb}^2 \bar{\omega}}{2\left(\omega + \omega_1\right)\left(\omega_1^2 - \omega_2^2\right)} \quad . \tag{19}$$

Assuming perturbed quantities $\omega = \omega_1 + \Delta$ and $\bar{\omega} = \omega_{ce} + \Delta$, we get from Eq. (19)

$$\Delta = \left[-\frac{\omega_{pb}^2 \omega_{ce}}{4\omega_1\left(\omega_1^2 - \omega_2^2\right)} \right]^{1/2} \quad . \tag{20}$$

As $\omega_1 < \omega_2$, there is no imaginary part of Δ, hence there is no growth rate and only the Alfven wave frequency gets changed.

Now, considering slow cyclotron interaction and assuming perturbed quantities $\omega = \omega_1 + \Delta$ and $\bar{\omega} = \omega_{ce}$, we get from Eq. (19)

$$\Delta = \left[\frac{\omega_{pb}^2 \omega_{ce}}{4\omega_1\left(\omega_1^2 - \omega_2^2\right)} \right]^{1/2}$$

As $\omega_1 < \omega_2$, Δ has an imaginary part and there is a growth rate given as

$$\gamma = \left[\frac{\omega_{pb}^2 \omega_{ce}}{4\omega_1\left(\omega_1^2 - \omega_2^2\right)} \right]^{1/2} \quad . \tag{21}$$

3. Results And Discussion

The numerical calculations have been carried out using the typical parameters of plasma: - electron number density $n_{eo}=10^{13}cm^{-3}$ and $10^{14}cm^{-3}$, mass of electron $m_e=9.1\times10^{-28}g$, charge of electron e = 4.8×10^{-10}ergs, magnetic field $\mathbf{B_s}$ = 0-1000 G and beam density $n_{bo}=2.8\times10^{9}cm^{-3}$.

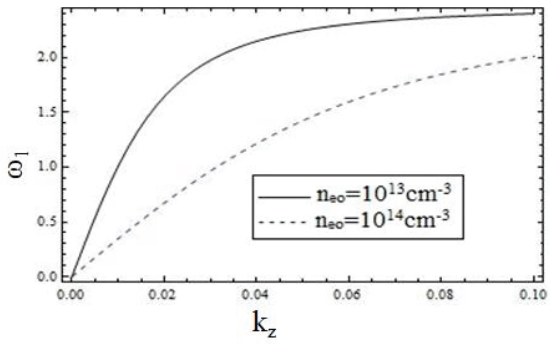


Figure 1. Alfven wave frequency ω_1 as a function of k_z.

Fig. 1 shows the variation of the Alfven wave frequency ω_1 as a function of parallel wave number k_z using Eq. (18) for two values of electron number density $n_{e0}=10^{13}cm^{-3}$ and 10^{14} cm^{-3} via slow cyclotron interaction with electron beam. From Fig. 1, it is clear that the Alfven wave frequency increases with increase in the value of k_z and saturates below the ion cyclotron frequency. The saturation value of wave frequency is also less for higher number density of electrons in plasma. Thus the phase velocity (ω_1/k_z) of Alfven wave decreases with an increase in number density of electrons or ions.

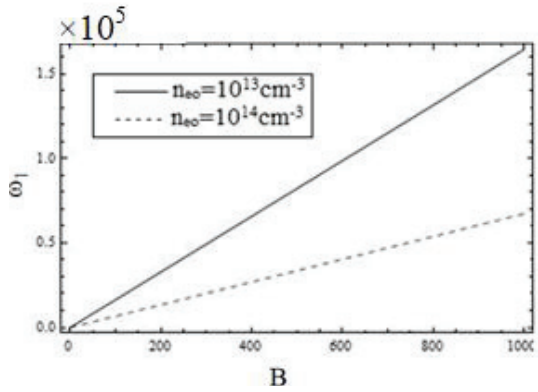


Figure 2. Alfven wave frequency ω_1 as a function of static magnetic field B.

In Fig.2, we have plotted the Alfven wave frequency as the function of magnetic field (in Gauss) by taking k_z =0.02cm^{-1} and for electron number density n_{eo} =10^{13}cm^{-3} and 10^{14}cm^{-3}. The Alfven wave frequency rises with rise in magnetic field. The slope of curve of ω_1 vs. **B** for lesser electron number density is more than that for higher electron density i.e. ω_1 increases more rapidly with magnetic field with decrease in number density of electrons of plasma.

Fig.3 shows the variation of growth rate (γ) for slow cyclotron interaction with wave number k_z (cm^{-1}) for different electron number densities of plasma n_{eo} =10^{13}cm^{-3} and 10^{14}cm^{-3} plotted using Eq. (21). The growth rate for slow cyclotron interaction decreases with increase in parallel wave number. In fast cyclotron interaction, the beam is not able to interact with the plane polarized Alfven wave for the unstable mode and therefore, no excitation or growth is obtained (c.f.Eq.20).

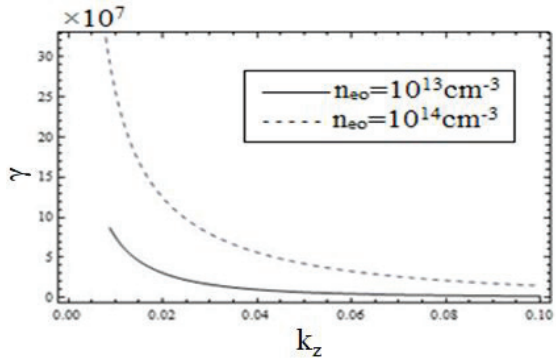


Figure 3. Growth rate γ of the unstable mode as a function of wave number k_z for anomalous Doppler resonance.

Fig.4 shows the variation of growth rate (γ) for slow cyclotron interaction with static magnetic field B_s for different electron plasma densities. The growth rate decreases with increase in value of static magnetic field.

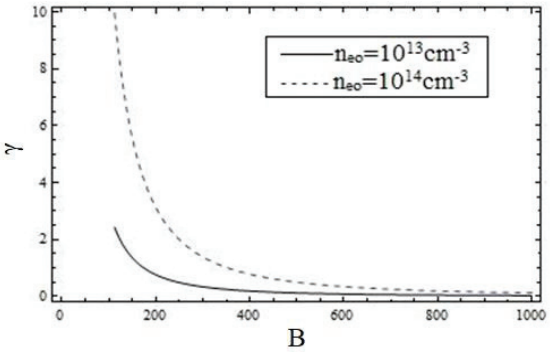


Figure 4. Growth rate γ of unstable mode as a function of static magnetic field B for anomalous Doppler resonance.

4.Conclusion

The plane polarized Alfven wave shows only cyclotron interaction with the beam mode and Cerenkov interaction is absent in the present case. For fast cyclotron interaction, the Alfven wave neither grows nor damps but there is an increment in Alfven wave frequency. However, for slow cyclotron interaction or anomalous Doppler resonance, the wave stabilizes with growth rate directly proportional to the beam density and inversely proportional to the external magnetic field. The wave frequency remains unaltered in the case of slow cyclotron interaction.

REFERENCES

1. J. Shrivastava and M. S. Tiwari, "Alfven wave in dusty magnetosphere", *Indian Journal of Radio & Space Physics*, **33**, 25 (2004).
2. L. Chen , D. J. Wu, G.Q. Zhao, J. F. Tang, and J. Huang, "Excitation of Kinetic Alfven waves by fast electron beams", *The Astrophysical Journal,* 793:13(5pp) (2014).
3. P. K. Shukla and R. P. Sharma, "Alfven wave generation in a beam plasma system", *Physical Review A*, **25**, 2816 (1982).
4. V. Prakash, S. C. Sharma, Vijayshri and R. Gupta "Ion beam driven resonant ion –cyclotron instability in magnetized dusty plasma", *Phys. Plasmas* **21**, 33701 (2014).
5. R. Gupta, V. Prakash, S. C. Sharma and Vijayshri, "Interaction of an electron beam with whistler waves in magnetoplasma", *Laser and particle beams*, **33**, 455 (2015).
6. V. I. Shevchenko, V. L. Galinsky and S. K. Ride, "Excitation of left-hand-polarized nonlinear Alfven waves by an ion beam in a plasmas", *Journal of Geophysical Research*, **107**, A11, 1367 (2002).
7. R. Gupta, S. C. Sharma and V. Prakash, "Excitation of surface plasma waves by a density-modulated electron beam in a magnetized plasma cylinder", *Physics of Plasmas*, **17**, 122105 (2010).
8. L. Yin, D. Winske, W. Daughton, and K. J. Bowers, "Kinetic Alfven waves and electron physics. I. Generation from ion-ion streaming*", Physics of Plasma*, **14**, 62104 (2007).
9. P. K. Shukla and L. Stenflo, "Decay of a magnetic-field-aligned Alfven wave into inertial and kinetic Alfven waves in plasmas"*, Physics of Plasma,* **12**, 84502 (2005).
10. Yu. Voitenko, M. Goossens, O. Sirenko, and A. C.-L. Chian, "Nonlinear excitation of kinetic Alfven waves and whistler waves by electron beam-driven Langmuir waves in the solar corona", *EDP Sciences*, A&A 409, 331, (2003).
11. S.K.P. Tripathi, B. Van Compernolle, W. Gekelman, P. Pribyl, and W. Heidbrink. , "Excitation of shear Alfven waves by a spiraling ion beam in a large magnetoplasma", *Physical Review E,* **91**, 13109 (2015).
12. P. K. Shukla and R. P. Sharma, "Alfven waves generation in a beam plasma system", *Physical Review A,* **25**, 2816 (1982).

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/JLT.2019.2931077, Journal of Lightwave Technology

System-level Performance of Chip-based Brillouin Microwave Photonic Bandpass Filters

Yiwei Xie, Amol Choudhary, Yang Liu, David Marpaung, *Member, IEEE*, Khu Vu, Pan Ma, Duk-Yong Choi, Steve Madden and Benjamin J. Eggleton, *Fellow, IEEE*

***Abstract*—Microwave photonic (MWP) filters using on-chip stimulated Brillouin scattering (SBS) offer great potential for next-generation radio frequency (RF) applications due to the unprecedented high spectral resolution, flexible programmability, and ultra-wideband frequency agility. Although impressive functionalities have been reported, limited link performance, due to the amplification noise induced by the Brillouin gain process, hinders the practical deployment in existing RF applications. In this paper, we present the first comprehensive numerical study and experimental of chip-based Brillouin MWP bandpass filters implemented in a range of link configurations. In the experiments, key RF performance figures of merit, including link gain, noise figure, passband resolution, bandwidth tunability and passband extinction of chip-based Brillouin filters are experimentally investigated and numerically analyzed. This comprehensive study points out a preferable filter scheme with optimized system-level performance. The numerical model developed in this work allows for extrapolations to further improve the performance. The systematic work establishes a feasible design route to achieving high link-performance chip-based Brillouin microwave photonic bandpass filters.**

***Index Terms*—Microwave photonics, link performance, photonic integrated circuits, noise figure, stimulated Brillouin scattering.**

I. Introduction

RADIO frequency (RF) filters are a critical building block in modern electronic RF systems, providing functionalities for selecting RF signals of interest. Integrated microwave photonic (IMWP) filters have recently gained significant interest for RF signal processing [1], due to the broad bandwidth and superior central frequency tunability over tens of gigahertz (GHz), in combination with the compact size, weight, low power consumption (SWaP) and potential low cost offered by photonic integration [2–4].

Numerous IMWP filters have been demonstrated [5–12] by mapping the filter responses of linear optical devices, for example, using ring resonators (RR) [11], or an array of Mach-Zehnder interferometers [12]. Such demonstrations have successfully achieved wideband filter frequency tuning, but these schemes suffer from limited spectral resolution (typically GHz instead of MHz linewidths) and non-programmable filter shape. Recently, novel RF photonic filters using integrated Kerr frequency combs have enabled flexible filter bandwidth tunability [13 – 15]. However, achieving high spectrum resolution requires comb tooth number exceeding 80, which significantly increases the complexity in the spectrum shaping process. Ultrahigh-quality resonators enabled by whispering gallery mode resonances in silica spheres do offer MHz linewidths but come at the expense of incompatibility with

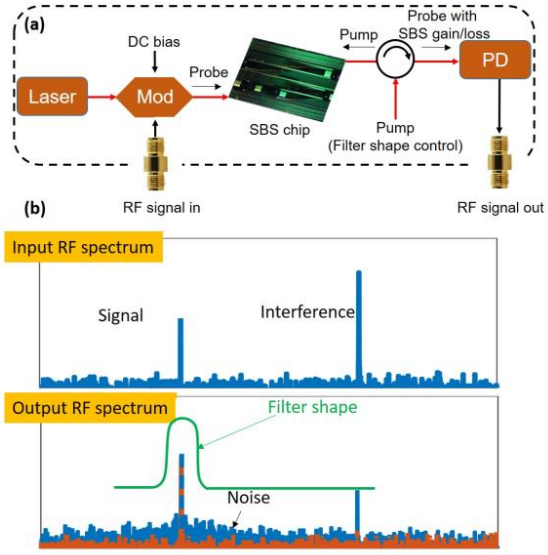


Fig. 1. (a) A general architecture of the on-chip SBS-based MWP filter. (b) The schematic of the input and output RF signal after passing through a SBS-based MWP filter in the frequency domain, blue line: input or output signal, orange dashed line: Ideal signal with sinusoidal waveform. (Mod: modulator)

Manuscript received XXXX; revised XXXX; accepted XXXX, Date of publication XXXX; date of current version XXXX. This work was supported by Lockheed Martin under a University of Sydney contract. (*Corresponding author: Yang Liu*)

Y. Xie, Y. Liu, and B. J. Eggleton are with the Institute of Photonics and Optical Science (IPOS), School of Physics, University of Sydney Nano Institute (Sydney Nano), Camperdown, NSW 2006, Australia (e-mail: yiweixie@zju.edu.au; yang.liu@sydney.edu.au; benjamin.eggleton@sydney.edu.au). A. Choudhary is with the Department of Electrical Engineering, The Indian Institute of Technology, Hauz Khas, New Delhi 110016, India (achoudhary@ee.iitd.ac.in). He was previously with the Institute of Photonics and Optical Science (IPOS), School of Physics, University of Sydney Nano Institute (Sydney Nano), Camperdown, NSW 2006, Australia. D. Marpaung was with the Institute of Photonics and Optical Science (IPOS), School of Physics, University of Sydney, Sydney, NSW 2006, Australia. He is now with the Faculty of Science and Technology, University of Twente, Enschede 7522 NB, The Netherlands (e-mail: d.a.i.marpaung@utwente.nl).
Khu Vu, Pan Ma, Duk-Yong Choi, Steve Madden are with Laser Physics Centre, School of Physics and Engineering, The Australian National University, Canberra ACT, 2601, Australia (e-mail: duk.choi@anu.edu.au; pan.ma@anu.edu.au; duk.choi@anu.edu.au, stephen.madden@anu.edu.au).

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

integration [16, 17].

Stimulated Brillouin scattering (SBS), a nonlinear optical process, provides an efficient approach to achieving precise RF spectral processing, exhibiting a high spectrum resolution in the level of tens of MHz enabled by the inherent narrow linewidth of the SBS gain resonance. The bandwidth of SBS responses can be flexibly adjusted and tailored from several MHz to hundreds of MHz [18 – 24] by simply pump tailoring. SBS-based MWP filters were demonstrated in optical fibers [10, 20], and recently implemented using integrated Brillouin photonic chips [19]. SBS-based filters usually exploit the SBS gain resonance in a MWP link with intensity modulation (IM) [21, 25] or phase modulation (PM) [19]. However, reported realizations of MWP bandpass filters using SBS gain resonance exhibit high in-band noise induced by the spontaneous emission associated with the SBS amplifications [26]. The additional noise leads to a high noise figure (NF) penalty, resulting in reduced link performance below the targeted performance required in many practical RF applications. The noise generated by the spontaneous Brillouin scattering process ultimately will limit the deployment of these devices in many real-world RF systems [27].

An intuitive understanding of a bandpass filter for interference signal is illustrated in Fig. 1, showing the side effects of the increased in-pass noise. Figure 1(a) shows the conceptual architecture of the MWP filter using the SBS chip incorporated in an analog photonic link with an optical pump block for filter shape control and a photodetector (PD). As shown in Fig. 1(b), a signal of interest together with a strong interferer is sent to the filter input with very low background noise. After passing through the MWP bandpass filter, the interferer is strongly suppressed. However, the RF output signal (blue trace) shows additional in-band noise and distortions, compared to the ideal RF signal (orange trace). The total noise power delivered to the RF signal consists of spontaneous emission associated with SBS amplification, thermal noise, shot noise, relative intensity noise (RIN), and amplified spontaneous emission (ASE) noise [27, 28]. Among them, SBS noise induced by Brillouin amplification is the dominant noise source, especially at high pump power [29]; it has an approximately Lorentzian shape centered on the Brillouin central frequency [30]. As we can see from Fig. 1(b), the bandpass filter rejects the interferer, while it introduces noise to the signal thus degrading the signal SNR.

One approach to reducing the SBS induced noise is by using the SBS loss resonance [20, 21]. In this approach, which was demonstrated using optical fibers, multiple broadened SBS loss responses suppress out-of-band signals without affecting the in-band signals, allowing the RF signals in the MWP filter passband to be transmitted through the link free from the additional SBS noise. A record-low NF of 15 dB was achieved, which was dominated by the inherent link performance [21]. However, extending this approach to chip-based MWP filters will incur an increased noise figure penalty, due to the increased optical insertion loss, reduced optical power handling and the requirement for additional optical amplification. The additional complexity involved and the different arrangements for the filter schemes will influence the link performance. Thus, analyzing and comparing the MWP link performance for on-chip SBS-MWP links requires a comprehensive investigation.

In this work, we present the first systematic experimental and numerical study of the overall link performance of chip-based Brillouin MWP bandpass filters implemented using different filter configurations. Furthermore, we analyze the dominant noise contributions to the overall filter performance, providing a concrete basis of filter performance optimization. Our results highlight that the bandpass filter using SBS loss resonance in an intensity modulation link configuration has the most favorable properties; the filter passband of this scheme is free from additional SBS noise, thus suffering the lowest noise compared with other techniques. In section II, we introduce four filter architectures by combining SBS gain or loss resonances with different modulation schemes. In Section III, a numerical model of various MWP filter links is established. In Section IV and V, we present the experimental setup and the experimental results, respectively, with the comparison to numerical simulations. In Section VI, we discuss the possible approaches to improving the link performance of chip-based Brillouin MWP bandpass filter.

II. Principle of Brillouin-based Bandpass Filter

Stimulated Brillouin scattering (SBS) is a nonlinear optical process that produces strong interactions between a counter-propagating strong optical (pump) wave and a weak optical wave (probe) mediated by an optically excited acoustic wave (density wave) in the medium. The pump is inelastically scattered backward by the material refractive index modulation, resulting in a frequency downshift to the Stoke frequency, producing a narrowband gain resonance. Similarly, a loss resonance can be produced at the anti-Stokes frequency. The gain and loss spectrum generated by the SBS process can be used to implement RF bandpass and bandstop filter responses, through the optical-to-electrical conversion via photodetection [31].

We implement the bandpass filter functionality using SBS gain and losses with a MWP link in intensity modulation (IM) or phase modulation (PM) configuration. As shown in Fig. 2(a), the pump is placed at a higher frequency with respect to the probe frequency, and the probe is in the single sideband modulation configuration. The pump is modulated by the electric signal generated by an arbitrary waveform generator (AWG), producing single sideband modulation with suppressed carrier. The pump wave induces SBS gain resonance on the probe sideband. The gain response can be flexibly reconfigured by tuning the power and frequency of the electric signals generated by arbitrary waveform generator (AWG) [21]. It should be noted that the signal within the SBS gain spectrum is overlapped with the spontaneous noise induced by the SBS gain process.

In contrast, the bandpass filter based on SBS loss using IM is shown in Fig. 2(b). In this case, the pumps are placed on the lower frequency side of the probe to induce two SBS loss responses, carving a passband response between two loss

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

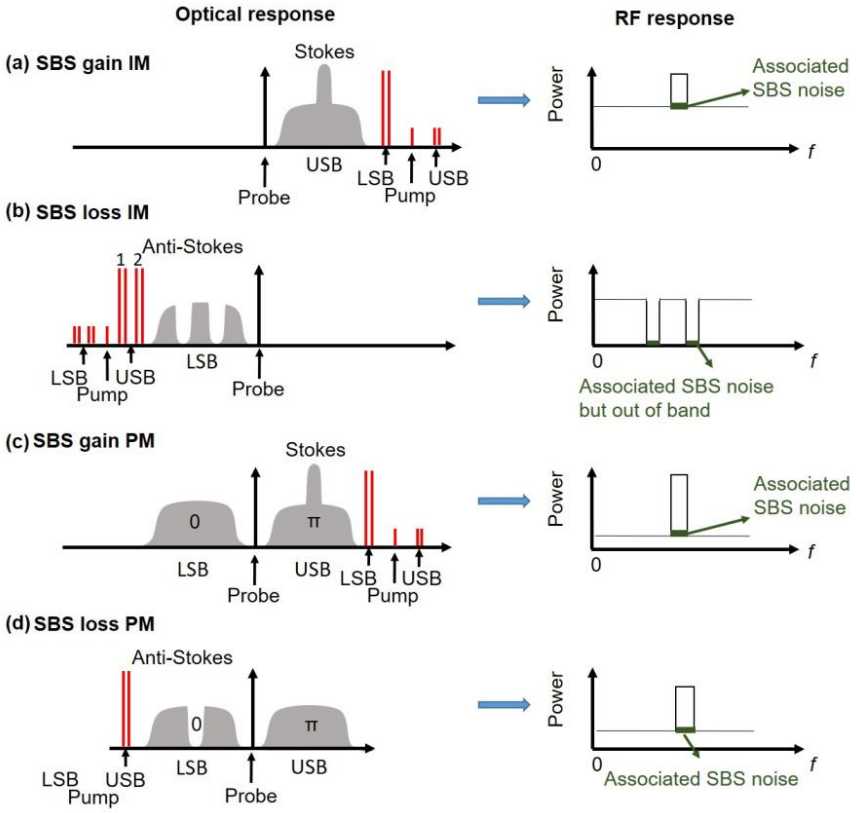


Fig. 2. Schematic of the operation principle of various Brillouin-based bandpass filters. (a) SBS gain resonance processes the upper sideband of a single-sideband optical signal implemented by an intensity modulator (IM). (b) SBS loss resonance processes the lower sideband of a single-sideband optical signal implemented by an IM. (c) SBS gain resonance processes the upper sideband of a dual-sideband optical signal implemented by a phase modulator (PM). (d) SBS loss resonance processes the lower sideband of a dual-sideband optical signal implemented by a PM. It should be noted that the signals produced by the upper sideband and the lower sideband of the phase modulation exhibits a π phase difference which leads to desctructive interference in the RF domain. LSB: lower sideband; USB: upper sideband.

resonances. The passband width can be adjusted by tuning the frequency spacing of the comb lines groups (1 and 2) through the AWG. The advantage of this approach is that the NF within the filter passband is free from the noise associated with the SBS gain process, as the passband is not processed within the SBS bandwidth. An additional coarse optical filter would be needed, in the case when wide stopband suppression is desired.

A MWP link with IM has several advantages including the ability to operate at low-biased configuration to optimize the link performance [32]. In contrast, a MWP link with PM is not able to achieve link performance optimization, due to the lack of carrier-to-sideband ratio tunability. However, PM can generate two sidebands with a phase difference of π, leading to destructive interference at the PD. As shown in Fig. 2(c), once the Brillouin gain response is applied to one of the sidebands, the complete destructive interference condition is no longer met within the SBS bandwidth, resulting in the generation of a bandpass response with a high extinction. Compared to the filter formed using single sideband modulation that allows for one-to-one spectrum mapping of the optical SBS response to the electrical domain, a higher filtering selectivity can be achieved, since the RF signals destructively interfere outside the SBS bandwidth [19, 33] as shown in Fig. 2(c).

Similarly, when the SBS loss response is applied to one of the sidebands as shown in Fig. 2(d), the condition for destructive interference is not met within the SBS bandwidth, resulting in a passband response [33]. Compared with the filter scheme shown in Fig. 2 (d), the PM link combined with SBS gain avoid the noise generated by the SBS gain process, exhibiting a much lower level of noise that is produced by the anti-Stokes process (associated with the loss resonance) [33].

III. The Numerical Model of the Microwave Photonic Link

In this section, a numerical model is developed to describe and analyze the link performance of the MWP bandpass filters in various configurations. The optical field at the output of the IM and PM are given by [34]:

$$E_{m1}(t)=\sqrt{P_{0,in}\alpha_{Mod}}\exp\left[j(\omega_c t+\phi_c)\right]\cdot\sum_{n=-\infty}^{\infty}J_n(m_{RF})\cos\left(\frac{\phi_{dc}-n\pi}{2}\right)\exp\left[jn(\omega_{RF}t+\phi_{RF})\right]\ \text{(IM case)} \tag{1a}$$

$$E_{m2}(t)=\sqrt{P_{0,in}\alpha_{Mod}}\exp\left[j(\omega_c t+\phi_c)\right]\cdot\sum_{n=-\infty}^{\infty}j^n J_n(m_{RF})\exp\left[jn(\omega_{RF}t+\phi_{RF})\right]\ \text{(PM case)} \tag{1b}$$

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

where $J_n(x)$ is the n-order first-kind Bessel function, α_{Mod} is the optical insertion loss of the modulator, $m_{RF} = \pi V_{RF}/V_\pi$ is the optical modulation index, V_π is the frequency-dependent RF half-wave voltage for the modulator, V_{RF} is the voltage of the RF signal. $\sqrt{P_{0,in}}\exp[j(\omega_c t+\phi_c)]$ is the input optical field to the modulator. $P_{o,in}$, ω_c, ϕ_c are the power, frequency, phase of the input signal to the modulator, respectively. ϕ_{RF} is the phase of the RF signal, and ω_{RF} is the angular frequency of the applied signal. For the IM case, the signal is in single sideband modulation (SSB) format, after one of the sidebands is filtered out by a tunable filter. For PM cases, the signal is in double sideband modulation (DSB) format, with a π phase difference. The subsequent integrated Brillouin waveguide produces a gain or a loss resonance on one of the signal sidebands at frequency f_{RF}. Considering a small-signal model and neglecting the high-order sidebands, the output optical field of the photonic chip for IM and PM cases that eventually goes into the PD is approximately expressed as:

$$\begin{aligned} E_{PD1}(G,m_{RF},t) &= \sqrt{\alpha_{chip}\alpha_0 G_{probe}}\,\varsigma \\ &\cdot[\cos(\phi_{dc}/2)J_0(m_{RF}) + A_1(G,m_{RF})\sin(\phi_{dc}/2)\exp[j(\omega_{RF}t)] \\ &- A_{-1}(G,m_{RF})\sin(\phi_{dc}/2)\exp[-j(\omega_{RF}t)]] \text{ (IM case)} \end{aligned} \tag{2a}$$

$$\begin{aligned} E_{PD2}(G,m_{RF},t) &= \sqrt{\alpha_{chip}\alpha_0 G_{probe}}\,\varsigma \\ &\cdot[J_0(m_{RF}) + A_1(G,m_{RF})\exp\left[j(\omega_{RF}t)+i\frac{\pi}{2}\right] \\ &- A_{-1}(G,m_{RF})\exp[-j(\omega_{RF}t)+i\frac{\pi}{2}]] \text{ (PM case)} \end{aligned} \tag{2b}$$

Where, A_{-1}, A_1 are lower and upper sideband amplitudes, respectively, which depend on SBS gain, G and optical modulation index, m_{RF}. The SBS gain, G, depends on the input pump power, $P_{p,in}$. G is the SBS gain in a linear scale, G >1 corresponds to SBS gain, while G <1 corresponds to SBS loss [35]. $\zeta = \sqrt{P_{0,in}\alpha_{mod}}\,exp[j(\omega_c t + \phi_c)]$, α_{chip} is the loss of the chip. a_0 is the loss of optical components in the setup. G_{probe} is the EDFA gain on probe path described in [36].

Link gain: According to Eq. (1) and (2), the RF link gain g_{link} can be written as

$$g_{link} = \left[r_{PD}\left(E_{PD}(G,m_{RF},t)\cdot E_{PD}^{*}(G,m_{RF},t)\right)\right]^2 R \tag{3}$$

where r_{PD} is the responsivity of the photodiode and R is the load impedance. The amplitude and phase notations for four different filter schemes denoted by Eq. 2 are summarized in Table I. For SBS gain using IM, as shown in Fig. 2(b), the probe is in the SSB modulation format, in which the lower

TABLE I
NOTATIONS OF KEY VARIABLES FOR DIFFERENT SCHEMES

Schemes	Lower sideband, A_{-1}	Upper sideband, A_1	Phase, θ	G
SBS gain IM	0	$\sqrt{G}\cdot J_1(m_{RF})$	0	>1
SBS loss IM	$\sqrt{G}\cdot J_{-1}(m_{RF})$	0	0	<1
SBS gain PM	$J_{-1}(m_{RF})$	$\sqrt{G}\cdot J_1(m_{RF})$	Π	>1
SBS loss PM	$\sqrt{G}\cdot J_{-1}(m_{RF})$	$J_1(m_{RF})$	Π	<1

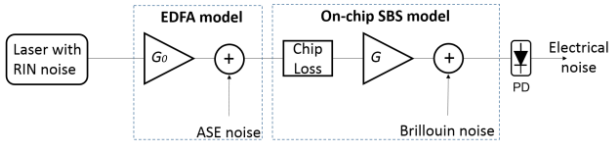


Fig. 3. Block diagram of possible noise contributions in a MWP filter link. Optical noises including the laser RIN noise, spontaneous noise from the EDFA, spontaneous Brillouin noise induced by the SBS process are distributed along the link. At the photodetector, the optical noises are converted to electrical noises via photodetection, along with the shot noise and thermal noises. It should be noted that the EDFA model was not used in the SBS PM link architectures.

sideband of the probe is filtered out by a tunable filter, and SBS gain is applied to the upper sideband. In this case, the sideband amplitudes are presented by $A_{-1} = 0$ and $A_1 = \sqrt{G}\cdot J_1(m_{RF})$, while the phase difference between two sidebands is $\theta = 0$. In contrast, in the SBS loss case using IM, the upper sideband of the probe is filtered out by a tunable filter, and the SBS gain is applied onto the lower sideband, with the relationship of $A_1 = 0$, $A_{-1} = \sqrt{G}\cdot J_{-1}(m_{RF})$, and $\theta = 0$. PM induces a π phase difference between two sidebands, i.e. $\theta = \pi$, while the two sidebands have the same amplitude. The SBS gain G is applied on upper sideband for SBS gain PM case, while on lower sideband for SBS loss PM case.

Noise Figure: Noise figure (NF) is a measurement of signal-to-noise degradation of the RF signal between the output and input. It depends on two variables: the RF link gain g_{link} and the noise power p_N. NF is defined as [28]

$$NF = 10\log_{10}\left(\frac{p_N}{g_{link}kTB}\right) \text{ (dB)} \tag{4}$$

where, k is the Boltzmann constant, T is the room temperature and B is the receiver electrical bandwidth. The noise sources in SBS based-MWP links include thermal noise p_{th}, shot noise p_{shot}, relative-intensity noise (RIN) of the probe, p_{RIN1}, reflected pump RIN p_{RIN2}, amplified spontaneous emission (ASE) noise p_{ASE}, and thermal noise p_{SBS} from SBS process.

The detected electrical power of these noise contributions are given by [31, 37, 38]

$$p_{th} = kTB(g_{link}+1) \text{ (W)} \tag{5}$$

$$p_{shot} = 2qI_D BR_L \text{ (W)} \tag{6}$$

$$p_{RIN1} = RIN_{probe} I_{probe}^2 BR_L \text{ (W)} \tag{7}$$

$$\begin{aligned} p_{pump} &= p_{RIN2} + p_{ASE_pump} \\ &= RIN_{pump} I_{pump}^2 BR_L + \frac{1}{4}\left[2r_{PD}I_D\left(2hvn_{sp1}(G_{pump}-1)B\right)R_L\right]\cdot\chi \text{ (W)} \end{aligned} \tag{8}$$

$$\begin{aligned} p_{ASE,EDFA} &= \frac{1}{4}\left[2r_{PD}^2 P_o P_{ASE} R_L\right] \\ &= \frac{1}{4}\left[2r_{PD}I_D\left(2hvn_{sp}(G_0-1)B\right)R_L\right] \text{ (W)} \end{aligned} \tag{9}$$

$$\begin{aligned} p_{ASE,EDFA+SBS\ gain} &= \frac{1}{4}\left[2r_{PD}^2 P_o (G_{SBS}a_{loss}P_{ASE})R_L\right] \\ &= \frac{1}{4}\left[2r_{PD}I_D\left(G_{SBS}a_{loss}2hvn_{sp2}(G_0-1)B\right)R_L\right] \text{ (W)} \end{aligned} \tag{10}$$

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

$$p_{ASE,EDFA+SBS\ \mathrm{loss}} = \frac{1}{4}\left[2r_{PD}^2 P_o (a_{loss} P_{ASE}) R_L\right] = \frac{1}{4}\left[2r_{PD} I_D \left(a_{loss}\, 2h\nu n_{sp}(G_0 - 1)B\right) R_L\right] \text{ (W)} \quad (11)$$

where I_D is the photocurrent including the probe (I_{probe}) and the reflected pump (I_{pump}), R_L is the load resistance, q is the electron charge, and $\mathrm{RIN_{probe}}$, $\mathrm{RIN_{pump}}$ are the relative intensity noise of probe and pump, respectively, n_{sp} is the spontaneous emission factor of EDFAs. The noise power induced by the pump include pump RIN noise and EDFA ASE noise, resulting from the pump back reflection at the facet of the chip; χ is the reflection ratio. G_{pump} is the EDFA gain on the pump path for the pump amplification. The optical power evolution of the SBS induced thermal noise in the waveguide can be approximated as [26]

$$p_{SBS0}(\Delta, z) = \left(\frac{1}{A_{eff}}\right) h\nu g_0(\Delta)(N+1)G(z)\int_z^L P_P(z')G(z')^{-1}dz' \text{ (W/Hz)} \quad (12)$$

where h is the Planck's constant, ν is the frequency of the light, and N is the thermal equilibrium number of the acoustic phonons, g_0 is the Brillouin gain coefficient varies with respect to the frequency, A_{eff} is the effective area and Brillouin gain coefficient of the waveguide, G is the SBS gain function, which can be described as

$$G(z) = \exp\left\{\int_z^L \left[P_P(z) g_0(\Delta)/A_{eff} - \alpha\right] \mathrm{d}z\right\} \quad (13)$$

$P_P(\mathrm{z})$ is the pump power evolution along the waveguide, which is

$$P_P(z) = P_{P,in} e^{-\alpha z} \quad (14)$$

under pump nondepletion case. $P_{p,in}$ is the input pump power, $\mathrm{g_0}$ is the Brillouin gain coefficient, given by

$$g_0(\Delta) = \left[2\pi n^7 p_{12}^2 \gamma / \left(c\lambda_p^2 \rho_0 \upsilon_a \Delta\nu_B\right)\right]\left[\Gamma^2 / \left(\Delta^2 + \Gamma^2\right)\right] = g_0(0)\left[\Gamma^2 / \left(\Delta^2 + \Gamma^2\right)\right] \text{ (m/W)} \quad (15)$$

where Γ^{-1} is the acoustic phonon lifetime resulting in a spontaneous Brillouin scattering linewidth ($\Delta\nu_B = \Gamma/\pi = 30$ MHz in Chalcogenide waveguides), ρ_0 is the average material density, n is the refractive index, λ_p is the wavelength of the pump, υ_a is the acoustic velocity, p_{12} is the longitudinal elastic-optic coefficient, c is the vacuum speed of light, γ is the integrated average of the unit vectors in the directions of the pump and Stokes optical fields. $\Delta = \omega_p - \omega_s - \omega_a$ is the frequency detuning from SBS gain line centre. ω_p, ω_s, ω_a are the angular frequency of the pump, stokes and acoustic waves, respectively. The total SBS-induced thermal noise is the integration of the p_{SBS1} over the gain profile. Photo-detection is a square-law process, so that the photo-current is composed of several beat signals between the signal and noise optical fields. The mixing product of signal-spontaneous (sig-sp) beat noises is given as

$$p_{SBS1} = \frac{1}{4} 2r_{PD}^2 P_{probe} p_{SBS0} R_L = \frac{1}{2} r_{PD} I_{probe} p_{SBS0} B_1 R_L \text{ (W)} \quad (16)$$

where, B_1 is the SBS bandwidth within which SBS noise is induced. The factor 1/4 is due to the impedance matching in the photodetector. According to Ref. [39, 40], SBS loss noise, p_{SBS2}, is almost directly proportional to the pump power, and will not be amplified during the SBS process. The power of SBS loss noise, p_{SBS2} is comparable to that of SBS gain noise, p_{SBS1} only when the SBS gain is small.

Figure 3 shows the general block diagram of noise contribution in different MWP filter links. The noise for different schemes can be expressed by

$$p_N = \begin{cases} p_{th} + p_{RIN1} + p_{pump} + p_{ASE,EDFA+SBS\ \mathrm{gain}} + p_{SBS1} + p_{shot} & \text{SBS gain IM} \\ p_{th} + p_{RIN1} + p_{pump} + p_{ASE,EDFA+SBS\ \mathrm{loss}} + p_{shot} & \text{SBS loss IM} \\ p_{th} + p_{RIN1} + p_{pump} + p_{SBS1} + p_{shot} & \text{SBS gain PM} \\ p_{th} + p_{RIN1} + p_{pump} + p_{SBS2} + p_{shot} & \text{SBS loss PM} \end{cases} \quad (17)$$

The thermal noise is due to thermally induced, random movements of charge carriers in conductors. Shot noise is generated at the PD due to the random arrival of photons which generate random fluctuations in the detected photocurrent. RIN arises from random emissions of spontaneous and stimulated photons in the gain medium of the laser. Optical amplifiers contribute additional ASE noise, which will be amplified by the SBS gain in SBS gain IM case only [41]. At the same time, ASE noise will undergo optical loss a_{loss} caused by the SBS chip. To reduce the effect of the ASE noise in IM link, the amplifier is placed before the photonic chip, because an appreciable seed power to the EDFA is necessary to reduce the ASE noise [28]. For the SBS PM cases, the insertion loss of the phase modulator is much lower than the intensity modulator in a low-biased configuration, so that there is no need for an extra EDFA. In the SBS gain cases, the SBS induced thermal noise contributes to the majority of the noise in the link, especially in the high pump power situation [26]. While for the SBS loss PM, the anti-Stokes noise is much lower than the Stokes noise [33]. Ideally, the SBS loss IM case does not suffer any noise from the SBS process because the signal of interest in not within the SBS bandwidth.

IV. Experimental Setup

The experimental setup in Fig. 4 shows the chip-based Brillouin MWP filter schemes. A high-power tunable laser with 30 dBm output power is used as the probe laser. A variable optical attenuator (VOA1) is connected to the laser output for power control. The pump is amplified, before being modulated by a dual parallel Mach-Zehnder modulator (DPMZM) to implement single sideband modulation with suppressed carrier. An AWG is used to generate electric comb lines, which modulates the pump laser to form broadband SBS pump, allowing for programmable filter shape control. Figure 4(a) shows the schematic of the IM case. The probe is modulated by a MZM (EOSpace 20 GHz) with an RF half-wave voltage of 4.4 V and an optical insertion loss of 5 dB. The modulator is operated in the low-biased configuration to reduce the noise power [42]. The RF input is generated by a vector network analyzer (VNA) with amplification provided by a low noise amplifier (LNA), which is inserted before the modulator to improve the link [21]. The tunable fiber Bragg grating (FBG) after the modulator is used to filter out one of the sidebands, and the insertion loss of the FBG is compensated by increasing the laser power. An erbium doped optical fiber amplifier

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

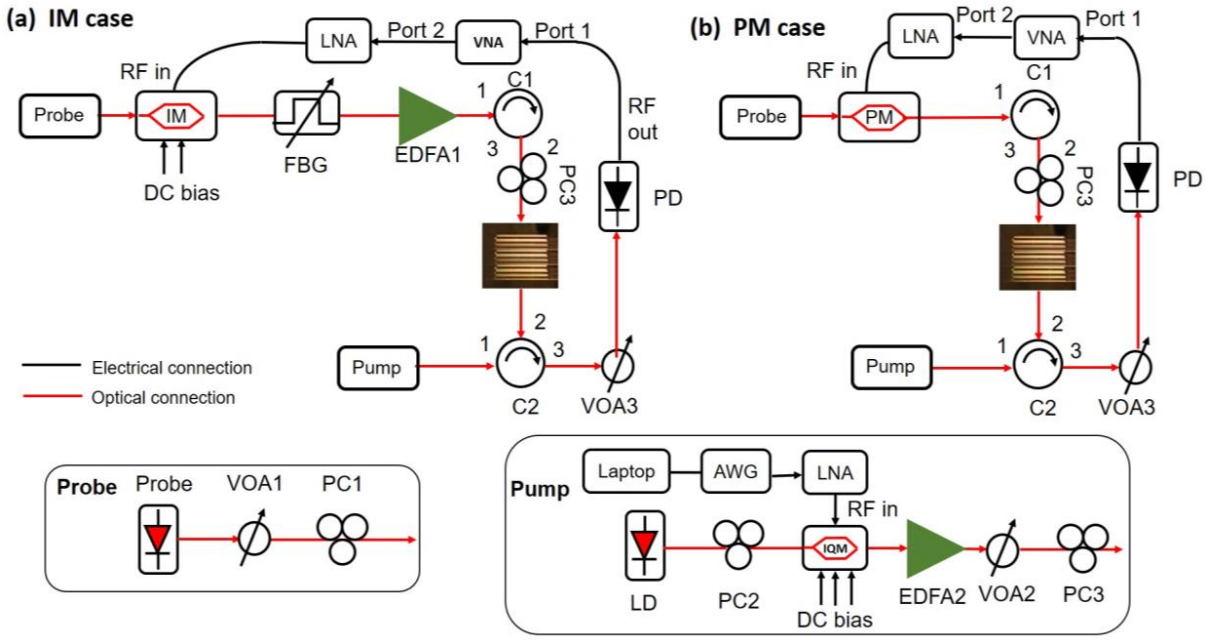


Fig. 4. Schematic of bandpass filter experiment setup. The probe path includes the high power laser source, variable optical attenuator (VOA1 and VOA3), polarization controller (PC1), intensity modulator (IM), DC bias input, fiber Bragg grating filter (FBG), erbium-doped fiber amplifier (EDFA1), RF input from vector network analyzer (VNA), low noise amplifier (LNA), photodetector (PD), and circulator (C1). The bottom side of the figure shows the pump of the filter setup including laser diode (LD), polarization controller (PC2, PC3), in-phase and quadrature Mach Zehnder modulator (IQM), arbitrary waveform generator (AWG), low noise amplifier (LNA), high resolution optical spectrum analyzer (OSA), and circulator (C2).

(EDFA) is placed before the chip to offset the optical insertion loss of the photonic chip [28].

An integrated chalcogenide photonic circuit (incorporating a 23.7-cm waveguide) is used as the SBS medium [43]. S-shaped spiral with a bend radius of 100 µm is used for the waveguide to minimize the area. The output of this waveguide is then connected to circulator (C2) before it is linked to a VOA3. This optical signal is detected by a PD (InGaAs APIC 20 GHz with 0.65 A/W responsivity and 50 Ohm termination). The RF output from the PD is measured by the VNA, producing the S21 response of the MWP link. For PM setup as shown in Fig. 4(b), the IM is replaced by a PM in Fig. 4(a).

V. Experimental Results Analysis

First, we compare the filter performance with a single pump for SBS gain IM, SBS gain PM, and SBS loss PM techniques, and two pumps with 300 MHz frequency spacing for SBS loss IM method (this effectively creates a 300 MHz passband response). For the comparison purpose, the pump and probe are aligned so that the central frequency of the bandpass filter is at 10 GHz.

A. Filter response comparison

The measured filter responses using SBS gain IM, SBS loss IM, SBS gain PM, and SBS loss PM techniques with pump power of 24 dBm are plotted in Fig. 5 for an RF signal power of -30 dBm. The 3-dB bandwidths of SBS gain IM and PM are around ~30 MHz, as shown in Fig. 5(a). From Fig. 5 (b), the SBS loss IM has a 3-dB bandwidth of 300 MHz, which depends on the frequency spacing between the two groups of the pumps. The passband bandwidth of SBS loss PM is around 150 MHz, as shown in Fig. 5(c). This is because the 3-dB bandwidth increases when the pump power increases for SBS loss cases, while it decreases for SBS gain cases [44]. At a high pump power (~24 dBm), the 3-dB bandwidth for SBS gain is smaller than that for SBS loss cases. As shown in Fig. 5(b), with the same pump power, the extinction ratio (ER) for SBS loss IM

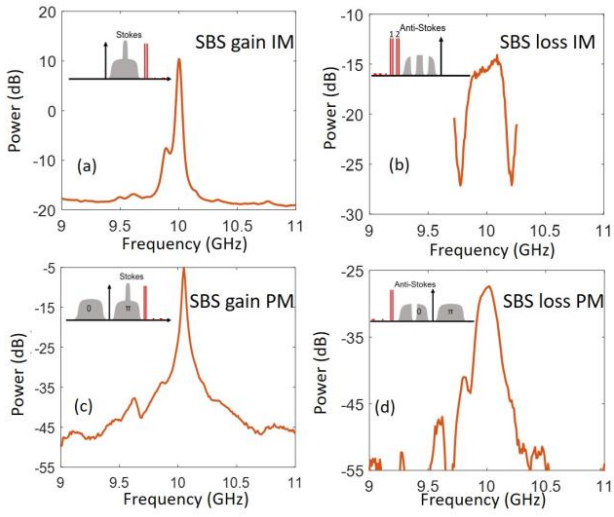


Fig. 5. Bandpass filter shape using (a) SBS gain IM, (b) SBS loss IM (Bandwidth = 300 MHz), (c) SBS gain PM, and (d) SBS loss PM at 24 dBm on chip pump power. Insets show the schematic of the sideband (grey) and the SBS pump (red).

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

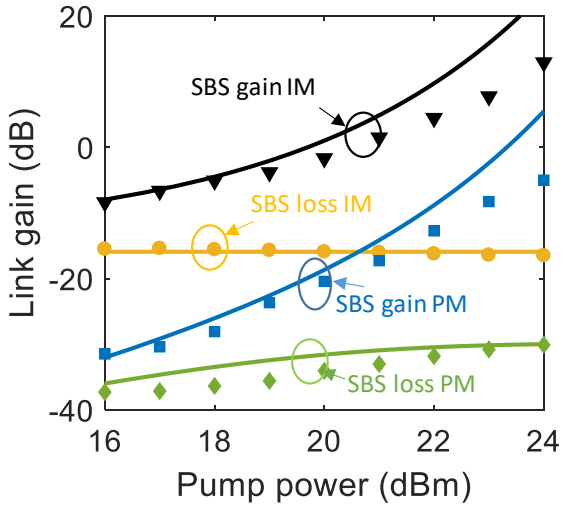


Fig. 6. Theoretical (solid line) and experimental (dots) link gain of passband filter using SBS gain IM, SBS loss IM, SBS gain PM, and SBS loss PM with respect to different pump power.

approach is much lower than other three methods (~15 dB), this is because the pump power is split to two frequencies. More detail of the bandwidth and ER parameters will be described in the next section.

B. *Link gain*

Figure 6 shows the measured link gain g_{link}, as a function of pump power, together with the calculated link gain using Eq. (3) with experimental parameters. The parameters used in the calculation for different components of the link is given in Table 2. The theoretical and experimental results show similar trends. However, when the pump power increases to 22 dBm,

TABLE II
PARAMETERS USED IN NUMERICAL SIMULATIONS

Device	Parameters
Laser	Laser power input to the IM: 23 dBm Laser power input to the PM: 28 dBm Measured laser RIN: RIN = -155 dB/Hz
IM	Insertion loss: α_{mod} = 6.5 dB IM output power: P_{in} = -7 dBm Vπ DC = 8 V Vπ RF = 5 V Bias ϕ_{dc} = 0.05
PM	Insertion loss: α_{mod} = 6.5 dB
RF source	RF output power: P_{RF} = -30 dBm LNA gain = 26 dB
EDFA	EDFA maximum output power: P_{out} = 22 dBm
SBS Chip	Waveguide loss: 0.5 dB/cm Chip coupling loss: 8 dB Chip length: L = 23.7 cm Measured Brillouin gain coefficient: g_0/A_{eff} = 423/mW Frequency shift: 7.6 GHz Brillouin linewidth: 30 MHz
PD	PD responsivity: r_{PD} = 0.7 A/W Load resistant: R = 50 Ω Current: $I_D \approx$ 0.002 A when no pump power
Component Loss	3 dB (including the loss of circulator, coupler, and attenuator)

the RF link gain exhibit small reduction due to the possible pump power depletion effect. In the experiment, the optical pump power was limited within 28 dBm (on chip 24 dBm), to avoid the increase in optical loss due to the photo-darkening effect [45]. For the SBS gain IM case, we tuned the on-chip SBS pump range from 16 to 24 dB, results in an increase in the passband link gain from -10 dB up to 17 dB. SBS gain using IM method has the highest link gain compared with other three methods because the signal of interest is amplified through the SBS process. As for the SBS loss IM with 300 MHz passband bandwidth, the link gain in the passband remains unchanged around -16 dB. This constant link gain is attributed to the fact that the passband is not processed, since the SBS loss response only suppresses the out-of-band frequency components. Therefore, the level of the link gain is determined by the RF photonic link. For PM cases, the link gain is lower than that using IM cases, due to the partial destructive cancellation induced by the properties of phase modulation. It should be noted that we consider the non-depleted pump condition in the numerical modeling depicted in Section III.

C. *Noise figure*

Figure 7 shows the measured and simulated noise power and NF with respect to different pump powers for four schemes. We measured the noise power of the link using an electrical spectrum analyzer. In the NF measurement, we selected the central frequency of the SBS bandwidth to measure the noise figure; all the noise powers were measured within 1 Hz bandwidth resolution. The comparison of measured and simulated noise power (Eq. 17) is shown in Fig. 7(a), the noise power of SBS gain cases is much larger than SBS loss cases due to the additional SBS noise and EDFA ASE noise amplified by SBS gain process. The discrepancy between the simulated results and the experimental results is due to the pump depletion and the increased insertion loss at high pump powers.

As shown in Fig. 7(b), the measured NF of SBS gain IM is around 32 dB, higher than that of SBS loss IM (28 dB) at 24 dBm pump power. As described in Eq. (4), NF depends on the link gain and the noise power of the link. Although the noise in the gain case is 40 dB larger than that in the loss case at 24 dBm pump power, due to SBS gain induced in-band SBS noise and amplified EDFA noise, but the link gain of SBS gain IM method is higher than SBS loss IM method, because the signal of interest is amplified through the SBS gain process, resulting in a noise figure slightly higher than that of SBS loss IM. The NF using PM schemes is in general larger than 40 dB, due to the low link gain. The lack of flexible tuning of carrier-to-sideband ratio prevents the link performance being optimized by the low-biasing operation. We also note that, the NF grows gradually as the pump power increased for SBS gain IM and SBS loss PM cases, in which the link gain increases at a similar rate as the noise power. The NF of SBS gain PM decreases from 50 dB to 40 dB when the pump power increases from 16 dBm to 23 dBm, and slightly increases to 42 dB when the pump power further increases to 24 dBm; this is because when the pump power is small, the increase rate of the link gain is higher than that of the noise as shown in Fig. 6 and 7(a), resulting in the decrease of the NF. When the pump power reaches around 23 dBm, the total noise is dominated by the SBS noise, which increases exponentially, thus increasing the NF, this

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

phenomenon is more obvious when the pump power is larger than 24 dBm. The calculated NF for four cases using Eq. (4) shown in Fig. 7(b) show a similar trend as the measured results. The discrepancy between the simulated results and the experimental results is due to the pump depletion and the increased loss at high pump powers. The PD current used in the calculation also includes the consideration of the back-reflected pump.

To understand the NF trend in Fig. 7, we select the SBS gain IM case as a representative example to analyze how the different noise contributions vary as a function of SBS gain, which is shown in Fig. 8, revealing that the amplified EDFA ASE noise and the SBS noise are the dominant noise contributions in the link. The back-reflected optical pump at the fiber-chip interfaces also induced additional electrical noise. The power of the shot noise is larger than that of RIN noise on probe path and thermal noise. Thermal noise depends on the

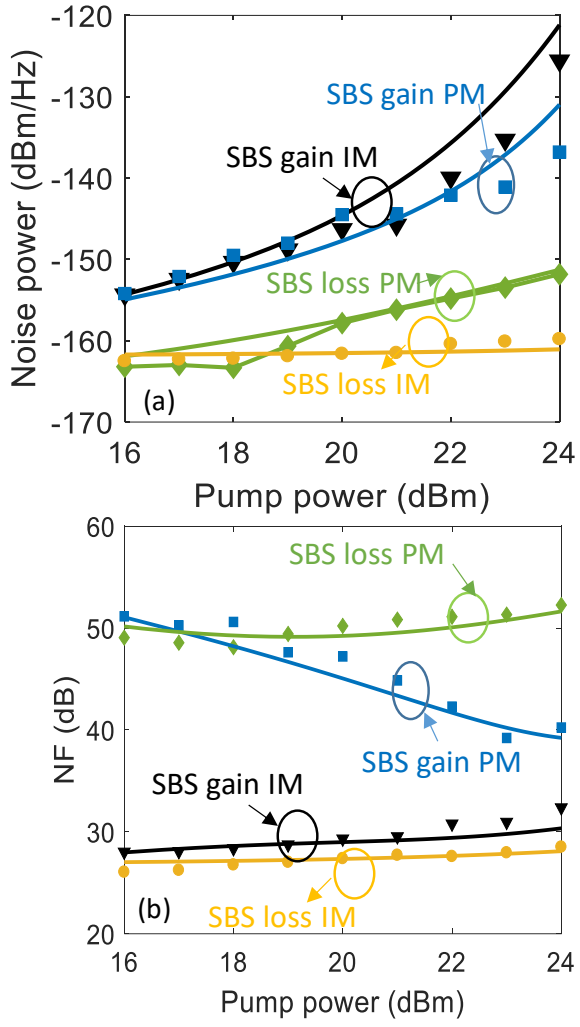


Fig. 7. Theoretical (solid line) and experimental (dots) (a) noise power and (b) noise figure vs. pump power (at the chip facet) for different schemes. On-chip pump power smaller than 24 dBm is used in the experiments (exp.), pump power larger than 24 dBm causes photo-darkening effects in the waveguide.

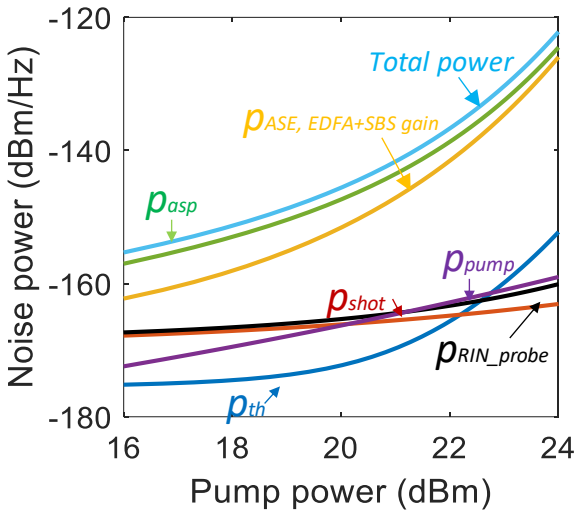


Fig. 8. The noise power contributions for SBS gain IM case (EDFA ASE noise amplified by SBS gain, SBS-induced noise, reflected noise including reflected pump noise and EDFA ASE noise on the pump path, shot noise, RIN noise of the probe, and thermal noise).

link gain, which is above 0 dB, when the pump power ranges from 20 dBm to 24 dBm. The laser RIN noise of the probe remains almost stable at around -167 dBm/Hz. In the link that employed SBS gain, the SBS noise and EDFA ASE noise are the dominant noise contributions; other noise contributions have lesser effects on the noise performance. On the contrary, in the link using SBS loss, SBS noise induced by the de-amplification process has much less impact on the total link noise.

D. Filter Resolution

Figure 9 compares the 3-dB bandwidth of the filter passband of four schemes. The expression of SBS-filter bandwidth can be found in [44]. According to Ref. [44], the bandwidth of the filter reduces on increasing the gain for low signal powers in

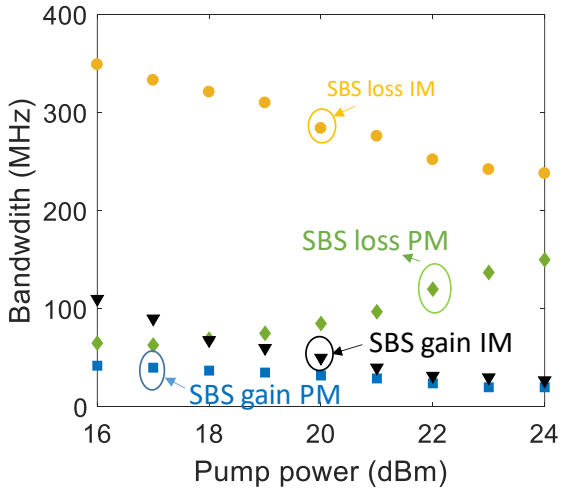


Fig. 9. Bandwidth change for different methods at a function pump power.

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

SBS gain cases, while the bandwidth of the filter increases when increasing the gain in SBS loss case. As shown in Fig. 9, the bandwidth of the filter reduces on increasing the pump power for low signal powers in SBS gain cases [46], while the bandwidth of the filter increases when increasing the gain in SBS loss case. The bandwidth for SBS loss PM case is around 150 MHz at a pump power of 24 dBm, while the bandwidth is around 25 MHz for SBS gain cases at the same pump power. For SBS loss IM case, we tuned the SBS pump power from 16 dBm to 24 dBm and obtained a decrease in filter resolution from 340 MHz down to 230 MHz, this is because the notch bandwidth increases when increasing the pump power, so that the frequency spacing between two notches decreases. In practice, the change of the bandwidth in the SBS loss IM scheme can be compensated by increasing the frequency differences between the losses.

E. Comparison of four schemes for passband filter realization

Table 3 shows a comprehensive comparison of the link performance for the four different filter synthesis techniques described in this paper. The passband using SBS gain IM has the highest link gain because the signal of interest is amplified through the SBS process. However, it suffers from high in-band amplified SBS noise associated with the SBS amplification as the signal of interest is within the SBS bandwidth. For the experimental setup, it needs an EDFA to increase the probe power before the power input into the chip for low biasing of the MZM, and an additional filter after the MZM to remove one of the sidebands. For SBS loss IM, this method suppresses the out- of-band signals to achieve a bandpass filter shape, however it has less effect due to the additional SBS noise because the passband is not processed by SBS. However, the ER is smaller than SBS gain techniques because of the low mixing products generated upon the PD. PM enables two sidebands to completely interference destructively except within the SBS bandwidth, resulting in a higher ER bandpass response than IM. However, PM has no bias control, and the link gain is lower than IM cases, it does not need a post filter.

Although the NF of PM case is much larger than that of IM case, the PM schemes have the benefit of the RF cancellation of out-of-phase signals at all frequencies, except where the signal experiences SBS gain. This effectively filters the RF signal in a very narrow bandwidth given by the SBS response with higher selectivity than achieved with single sideband (SSB) modulation (where the filter selectivity equals to the SBS gain). This enables the realization of higher selectivity filters using low on-chip power which is critical for applications that have a very limited power budget. As shown in Fig. 10, the ER using

TABLE III
COMPARISON BETWEEN DIFFERENT SCHEMES FOR BANDPASS FILTER (AT 24 DBM ON CHIP PUMP POWER)

Schemes	Link gain (dB)	Noise figure (dB)	Filter Extinction (dB)	Additional pre-filtering
SBS gain IM	17	32	30	Yes
SBS loss IM	-16	28	15	Yes
SBS gain PM	-5	40	50	None
SBS loss PM	-30	52	30	None

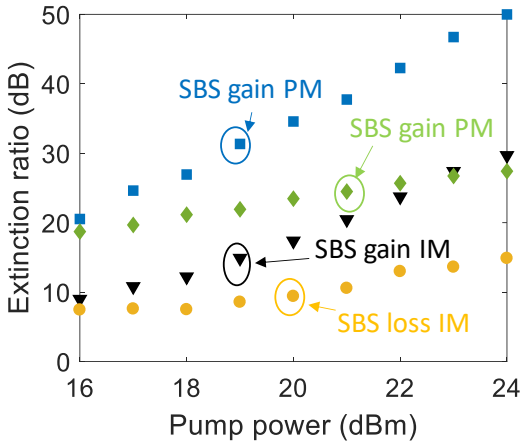


Fig. 10. Extinction ratio vs. pump power for different schemes.

SBS gain PM reaches 50 dB at pump power of 24 dB, while the ER is around 30 dB for SBS gain IM, and even lower for SBS loss IM scheme. However, PM case has lower link gain and no bias setting for noise optimization, so that the NF is larger than the IM case.

F. Bandwidth Broadening

We also investigated the broadband filter response and compared the noise performance of the filter employing the IM-based filter schemes. Here, we limit the discussion to the IM-based filter schemes, as the PM-based filter schemes are not able to produce appreciable link performance. An RF arbitrary waveform generator (AWG) is used to generate RF comb lines that modulates the pump laser to form SBS pump lines, allowing for broadening the filter shape. In practical applications, NF under 30 dB is desired [44], which is only achievable using the IM configuration according to Table III. Therefore, we limit this discussion to IM case in this section. Fig. 11 shows the bandpass filter with bandwidth of 300 MHz

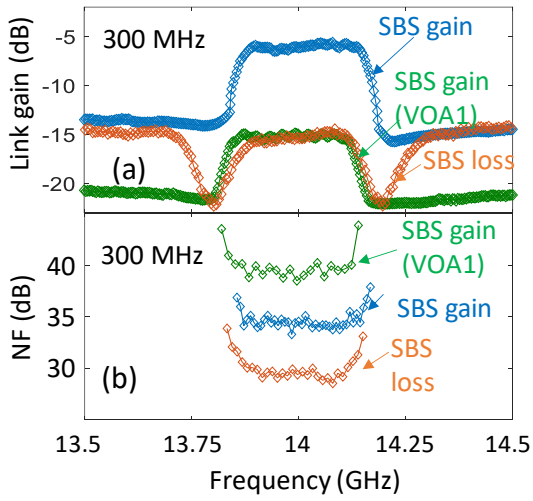


Fig. 11. (a) Link gain, (b) NF for the link using SBS losses and SBS gain (w/, w/o a VOA1) when the passband's bandwidth is 300 MHz.

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

using SBS gain and loss techniques at 14 GHz. The ripple on passband response could be affected by the noise and power reduction from the RF cable transmission, RF power reflection and the frequency dependent of components such as the LNA and the photodetector. The flatness of the passband can be optimized by tailoring the comb lines generated by the AWG [19]. Using the setup shown in Fig. 4, the maximum link gain is -5 dB at 10-dB ER.

To investigate the impact of thermal noise caused by SBS amplification, we set the passband link gain of both techniques to be the same by attenuating the optical power to the MZM through the VOA1, as shown in the green trace (SBS gain+VOA1) and orange lines (SBS loss) in Fig. 11(a). In practice, the passband generated using SBS loss IM scheme should be filtered by a passband filter with coarse roll off to allow broadband suppression. The NF of the SBS gain filter with VOA1 (green trace) is more than 10 dB larger than that using SBS losses (orange trace) as shown in Fig. 11(b). The NF of SBS gain link without using the VOA1 (blue trace) is 5 dB larger than that using SBS loss. This is mainly because the passband using SBS gain scheme suffers from in-band SBS amplification-associated noise, while the SBS noise in the SBS losses scheme locates out of passband.

VI. Discussion

NF is a key parameter to evaluate the usability of MWP links. To further investigate the improvement of the link performance, we use our numerical model to predict how the link performance can be improved by varying the chip loss and RIN of the laser.

The reduction in the propagation loss of the Brillouin waveguide efficiently increases the link gain, as the RF signal power increases quadratically with respect to the received optical power in the photodetector. In the SBS gain IM case, SBS gain increases when the chip loss reduces, resulting in a further increase of link gain. At the same time, reducing the chip loss also increases the detected electronic noise, as the photocurrent increases. Moreover, the increase in SBS gain also further increases the SBS-induced noise and amplified EDFA ASE noise which are the dominant noise sources, as shown in Fig. 8. As a result, the NF reduction is moderate,

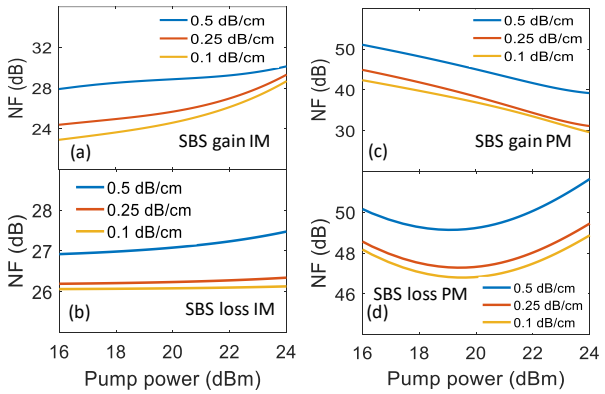


Fig. 12. NF for different waveguide loss (0.50 dB/cm, 0.25 dB/cm, 0.10 dB/cm) in (a) SBS gain IM, (b) SBS loss IM, (c) SBS gain PM, and (d) SBS loss PM, respectively.

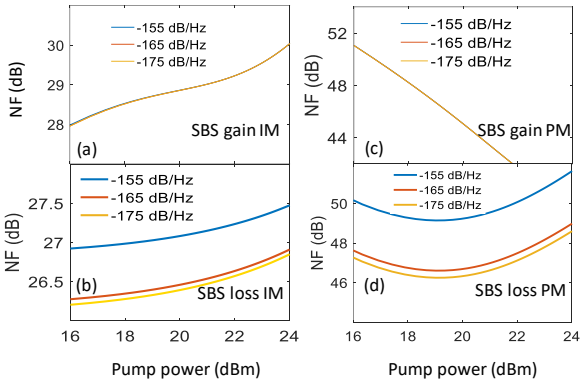


Fig. 13. NF for different laser RIN (-155 dB/Hz, -165 dB/Hz, -175 dB/Hz) in (a) SBS gain IM, (b) SBS loss IM, (c) SBS gain PM, and (d) SBS loss PM. (Three curves for SBS gain IM and SBS gain PM are overlapped as the RIN noise is not the dominate noise)

decreasing from ~28 dB to ~24 dB when the loss of the waveguide varies from 0.50 dB/cm to 0.1 dB/cm.

For the SBS loss IM case shown in Fig. 12(b), the passband noise performance avoids the effect of the SBS noise. Hence, the dominant noise is the EDFA ASE noise on the probe path. EDFA ASE noise on the probe path increases almost at the rate as the link gain. The noise figure has only around 1 dB reduction due to the effects of other noises, when the loss of the waveguide varies from 0.50 dB/cm to 0.1 dB/cm.

In Fig. 12(c), link gain of SBS gain PM case increases faster than that of SBS IM case as indicated by Eq. 1(a) and (b), but the SBS noise has the same rate as that of SBS gain IM case. As a result, the NF experiences a significant reduction of around 10 dB, when varying the chip loss from 0.50 dB/cm to 0.10 dB/cm. For the SBS loss PM case shown in Fig. 12(d), the link gain of this case increases at a slower rate than the SBS gain PM case. However, the dominant noise (SBS noise) almost at the same rate as the SBS gain PM case, resulting in a moderate NF reduction of ~3 dB when changing the chip loss from 0.50 dB/cm to 0.10 dB/cm.

Another important parameter we consider is the RIN of the light source. Fig. 13 shows the calculated NF when varying the pump power for different laser RINs. As illustrated in Fig. 8, the RIN noise on the probe path is much smaller than the dominant SBS noise and the EDFA ASE noise, so that changing the RIN has a negligible effect on the NF (Fig. 13(a) and 13 (c)). While for the SBS loss case, improving the RIN from -155 dB/Hz to -175 dB/Hz leads to a modest ~1 dB NF improvement, as shown in Fig. 13 (b) and Fig. 13 (d).

Based on the experimental and theoretical analysis, we find that the IM-based filter scheme provides an optimized noise figure, enabled by the low-biasing operation of the IM. On the other hand, PM-based RF photonic link configuration produces high filter extinction merely with moderate SBS gain, by exploiting the RF interference but lack of link optimization leads to significant NF penalty. To break this trade-off, an RF photonic filter scheme that combines the optimized link performance and RF interference technique needs to be explored in the future investigation. One possible approach is to implement an equivalent phase modulation with controllable

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

carrier-to-sideband ratio to emulate the low-biasing feature of MZM for link performance optimization, while the phase modulation enables a high filter extinction using moderate SBS gain.

VII. Conclusion

In this work, we report an experimental and theoretical investigation of the system-level performance of four architectures of chip-based Brillouin MWP bandpass filters. We evaluated and compared the overall link performance of different link architectures in terms of link gain, extinction ratio, resolution, bandwidth, and the noise figure. Through the analysis and performance comparison, the bandpass filter that pairs SBS loss with intensity modulation has the most favorable properties, as the filter passband is free from the Brillouin amplification noise. This work opens a path to achieve high-performance MWP bandpass filters for practical applications that requires versatile filters with large bandwidth and extinction tunability.

References

[1] V. J. Urick, J. D. McKinney, and K. J. Williams, *Fundamentals of microwave photonics* (Wiley, 2015).

[2] D. Marpaung, J. Yao, J. Capmany, "New opportunities for integrated microwave photonics," *J. Lightw. Technol*., vol. 30, no, 21, pp. 1813-1813, 2018.

[3] J. Capmany, P. Muñoz, "Integrated Microwave Photonics for Radio Access Networks," *J. Lightw. Technol*., vol. 32, no. 16, pp. 2849-2861, 2014.

[4] J. Yao, "Microwave photonics," *J. Lightw. Technol*., vol. 27, no. 3, pp. 314-335, 2009.

[5] J. Capmany and D. Novak, "Microwave photonics combines two worlds," *Nature Photon.*, vol. 1, no. 6, pp. 319–330, 2007.

[6] V. Vercesi, D. Onori, J. Davies, A. Seeds, and C. P. Liu, "Electronically synthesized Nyquist pulses for photonic sampling of microwave signals," *Opt. Express*, vol. 25, no. 23, pp. 29249-29259, 2017.

[7] E. A. Kittlaus, P. Kharel, N. T. Otterstrom, Zheng Wang, P. T. Rakich, "RF-photonic Filters via On-chip photonic-phononic emit-receive operations," *J. Lightw. Technol*, vol. 36, no. 13, pp. 2803-2809, 2018.

[8] Y. Xie, Z. Geng, L. Zhuang, M. Burla, C. Taddei, M. Hoekman, A. Leinse, C. G. H. Roeloffzen, K.-J. Boller, and A. J. Lowery. "Programmable optical processor chips: toward photonic RF filters with DSP-level flexibility and MHz-band selectivity." *Nanophotonics*, vol. 7, no. 2, pp. 421-454, 2017.

[9] L. Zhuang, C. G. H. Roeloffzen, M. Hoekman, K.-J. Boller, and A. J. Lowery. "Programmable photonic signal processor chip for radiofrequency applications." *Optica*, vol. 2, no. 10, pp. 854-859, 2015.

[10] T. Tanemura, Y. Takushima, K. Kikuchi, "Narrowband optical filter, with a variable transmission spectrum, using stimulated Brillouin scattering in optical fiber," *Opt. lett*., vol. 27, no. 17, pp. 1552-1554, 2002.

[11] P. Dong, N. N. Feng, D. Feng, W. Qian, H. Liang, D. C. Lee, B. J. Luff, T. Banwell, A. Agarwal, P. Toliver, R. Menendez, "GHz-bandwidth optical filters based on high-order silicon ring resonators," *Opt. Express*, vol. 18, no. 23, pp. 23784-9, 2010.

[12] E. J. Norberg, R. S. Guzzon, J. S. Parker, L. A. Johansson, L. A. Coldren, "Programmable photonic microwave filters monolithically integrated in InP–InGaAsP," *J. Lightw. Technol*., vol. 29, no. 11, pp. 1611-9, 2011.

[13] X. Xue, Y. Xuan, H. J. Kim, J. Wang, D. E. Leaird, M. Qi, A. M. Weiner, "Programmable single-bandpass photonic RF filter based on Kerr comb from a microring," *J. Lightw. Technol*., vol. 32, no. 20, pp. 3557-65, 2014.

[14] X. Xu, M. Tan, J. Wu, T. G. Nguyen, S. T. Chu, B. E. Little, R. Morandotti, A. Mitchell, and D. J. Moss, "High performance RF filters via bandwidth scaling with Kerr micro-combs," *APL Photonics*, vol. 4, no. 2, pp. 026102. 2019.

[15] X. Xu, M. Tan, J. Wu, T. G. Nguyen, S. T. Chu, B. E. Little, R. Morandotti, A. Mitchell, D. J. Moss, "Advanced Adaptive Photonic RF Filters with 80 Taps Based on an Integrated Optical Micro-Comb Source," *J. Lightw. Technol*., vol. 37, no. 4, pp. 1288-1295, 2019.

[16] B. Yu, Y. Chen, J. Pan, B. Zhang, F. Li, L. Wan, X. Guo, J. Li, Z. Li, "Silica-microsphere-cavity-based microwave photonic notch filter with ultra-narrow bandwidth and high peak rejection," *Opt. Lett.*, vol. 44, no. 6, pp.1411-1414, 2019.

[17] Y. Liu, Y. Yu, S. Yuan, X. Xu, X. Zhang, "Tunable megahertz bandwidth microwave photonic notch filter based on a silica microsphere cavity," *Opt. lett.*, vol. 41, no. 21, pp. 5078-81, 2016.

[18] Y. Stern, K. Zhong, T. Schneider, R. Zhang, Y. B. Ezra, M. Tur, and A. Zadok, "Tunable sharp and highly selective microwave-photonic band-pass filters based on stimulated Brillouin scattering," *Photon. Res.*, wol. 2, no. 4, pp. B18-B25, 2014.

[19] A. Choudhary, I. Aryanfar, S. Shahnia, B. Morrison, K. Vu, S. Madden, B. Luther-Davies, D. Marpaung, and B. J. Eggleton, "Tailoring of the Brillouin gain for on-chip widely tunable and reconfigurable broadband microwave photonic filters," *Opt. Lett*., Vol. 41, no. 3, pp. 436-439, 2016.

[20] C. Feng, S. Preussler, and T. Schneider, "Sharp tunable and additional noise-free optical filter based on Brillouin losses." *Photon. Res.*, vol. 6, no. 2, pp. 132-137, 2018.

[21] A. Mahendra, Y. Liu, E. Magi, A. Choudhary, D. Marpaung, B. J. Eggleton, "High link performance of Brillouin-loss based microwave bandpass photonic filters," *OSA Continuum*, vol. 1, no. 4, pp. 1287-1297, 2018.

[22] A. Byrnes, R. Pant, E. Li, D-Y Choi, C. G. Poulton, S. Fan, S. Madden, B. Luther-Davies, and B. J. Eggleton. "Photonic chip based tunable and reconfigurable narrowband microwave photonic filter using stimulated Brillouin scattering." *Opt. Express*, vol. 20, no. 17, pp. 18836-18845, 2012.

[23] A. Choudhary, Y. Liu, B. Morrison, K. Vu, D-Y Choi, P. Ma, S. Madden, D. Marpaung, B. J. Eggleton, "High-resolution, on-chip RF photonic signal processor using Brillouin gain shaping and RF interference" *Scientific Reports*, vol. 7, no. 1, pp. 5932. 2017

[24] S. Preußler, A.Wiatrek, K. Jamshidi, and T. Schneider, "Brillouin scattering gain bandwidth reduction down to 3.4MHz," *Opt. Express*, vol. 19, pp. 8565-8570, 2011.

[25] W. Wei, L. Yi, N. Y. Jiao, et al, "Bandwidth-tunable narrowband rectangular optical filter based on stimulated Brillouin scattering in optical fiber," *Opt. Express*, vol. 22, no. 19, pp. 23249, 2014.

[26] M. F. Ferreira, J. F. Rocha, J. L. Pinto, "Analysis of the gain and noise characteristics of fibre Brillouin amplifiers," *Optical & Quantum Electronics*, vol. 26, no. 1, pp. 35-44, 1994.

[27] D. Marpaung, C. Roeloffzen, R. Heideman, A. Leinse, S. Sales, and J. Capmany, "Integrated microwave photonics," *Laser & Photonics Reviews*, vol. 7, no. 4, pp. 506-538, 2013.

[28] Y. Liu, D. Marpaung, A. Choudhary, J. Hotten, and B. J. Eggleton, "Link performance optimization of chip-based Si_3N_4 microwave photonic filters," *J. Lightw. Technol*., vol. 36, no. 19, pp. 4361-4370, 2018.

[29] R. Brian Jenkins, Member, IEEE, Raymond M. Sova, Member, IEEE, and Richard I. Joseph, "Steady-State Noise Analysis of Spontaneous and Stimulated Brillouin Scattering in Optical Fibers," *J. Lightw. Technol*., vol. 25, no. 3, pp. 763-770, 2007.

[30] R. W. Boyd, K. Rzaewski, P. Narum, "Noise initiation of stimulated Brillouin scattering," Physical Review A, vol. 42, no. 9, pp. 5514, 1990.

[31] G. P. Agrawal, *Nonlinear fiber optics* (Elsevier academic press, 2017).

[32] A. Karim and J. Devenport, "Noise figure reduction in externally modulated analog fiber-optic links," *Photon. Technol. Lett*., vol. 19, no. 5, pp. 312–314, 2007.

[33] A. Choudhary, Y. Liu, D. Marpaung, B. J. Eggleton, "Brillouin filtering with enhanced noise performance and linearity using anti-Stokes interactions", in Conference on Lasers and Electro-Optics (CLEO), San Jose, USA, pp. STu4F. 2.

[34] E. E. Funk, M. Bashkansky, "Microwave photonic direct-sequence transmitter and heterodyne correlation receiver," *J. Lightw. Technol*., vol. 21, no. 12, pp. 2962-2967, 2003.

[35] B. J. Eggleton, C. G. Poulton, R. Pant, "Inducing and harnessing stimulated Brillouin scattering in photonic integrated circuits," *Advances in Optics and Photonics*, vol. 5, no. 4, pp. 536, 2013.

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

[36] P. C. Becker, N. A. Olsson, and J. R. Simpson, Erbium-Doped Fiber Amplifiers: Fundamentals and Technology. New York: Academic Press, 1999.
[37] X. Meng, "Designing high dynamic range microwave photonic links for radio applications," *Fiber Integrated Opt*., vol. 23, no. 1, pp. 1–56, 2004.
[38] Becker, P. C., N. A. Olsson, and J. R. Simpson. 1999.Erbium-Doped Fiber Amplifiers: Fundamentals and Technology, New York: Academic Press.
[39] M. J. Damzen, V. Vlad, A. Mocofanescu, V. Babin. *Stimulated Brillouin scattering: fundamentals and applications.* (CRC press, 2003).
[40] M. Horowitz, A. R. Chraplyvy, R. W. Tkach, J. L. Zyskind. Broad-band transmitted intensity noise induced by Stokes and anti-Stokes Brillouin scattering in single-mode fibers. *Photonics Technology Letters*, vol. 9, no. 1, pp. 124-126, 1997.
[41] F. Bucholtz, J. D. Mckinney, M. E. Godinez, et al. "Analysis of an Analog Fiber-Optic Link Employing a Low-Biased Mach–Zehnder Modulator Followed by an Erbium-Doped Fiber Amplifier," *J. Lightw. Technol*., vol. 27, no. 12, pp. 2013-2019, 2009.
[42] H. V. Roussell, M. D. Regan, J. L. Prince, C. H. Cox, J. X. Chen, W. K. Burns, G. E. Betts, and E. I. Ackerman, "Gain, noise figure and bandwidth-limited dynamic range of a low-biased external modulated link," Microwave Photonics (MWP), IEEE International topical meeting on. 1, 4224-11680-8, 2007.
[43] A. Choudhary, B. Morrison, I. Aryanfar, S. Shahnia, M. Pagani, Y. Liu, K. Vu, S. Madden, D. Marpaung, B. J. Eggleton, "Advanced integrated microwave signal processing with giant on-chip Brillouin gain," *J. Lightw. Technol*., vol. 3, no. 4, pp. 846-854, 2017.
[44] D. Marpaung, B. Morrison, M. Pagani, R. Pant, D. Y. Choi, B. Luther-Davies, S. J. Madden, and B. J. Eggleton, "Low-power, chip-based stimulated Brillouin scattering microwave photonic filter with ultrahigh selectivity", *Optica*, vol. 2, no. 2, pp.76-83, 2015.
[45] B. J. Eggleton, B. Luther-Davies, K. Richardson, "Chalcogenide Photonics," *Nature Photonics*, vol. 5, no. 3, pp. 141-148, 2011.
[46] A. Choudhary N. G. Seil, M. Pelusi, D. Marpaung, B.J. Eggleton, K.Vu, P.Ma, D.Yong-Choi, S. Madden, "Linearity and resolution of on-chip Brillouin filters for RF and optical communications," *2017 Opto-Electronics and Communications Conference (OECC) and Photonics Global Conference (PGC)*, Singapore, 2017, pp. 1-2.

Yiwei Xie received M. Sc. Degree in optical communication from Huazhong University of Science and technology, Wuhan, China, in June 2014 and the Ph.D. degree in optoelectronics from the Monash University in Melbourne, Australia in Sep. 2018. From July 2018 to Dec. 2018, she was a professional research fellow in School of Physics, University of Sydney. In 2019, she joined the Center for Optical and Electromagnetic Research, Zhejiang University as a lecturer. Her research topic is integrated photonics for high-speed, energy-efficient optical communication systems and microwave photonics.

Amol Choudhary (received a Bachelor of Engineering from the Delhi College of Engineering (now DTU), India; an Erasmus Mundus MSc in Photonics from the Gent University, Belgium, the University of St. Andrews, UK and a PhD in Optoelectronics from the University of Southampton, UK. He worked as an EPSRC Doctoral Prize Fellow at the University of Southampton and as the Australian Government's Discovery Early Career Research Award (DECRA) fellow at the University of Sydney, Australia. He is currently a faculty member at the department of Electrical Engineering at the Indian Institute of Technology (IIT) Delhi, India. His research areas of interest are photon-phonon interactions, laser physics, integrated optics, non-linear optics, ultrafast optics, optical communications and microwave photonics.

He is a Senior member of the Optical Society (OSA). He is also the 2018 recipient of the OSA's Ivan P. Kaminow Outstanding Early Career Professional Prize and is a 2019 OSA Ambassador. He has more than 100 publications including 31 Journal papers. He is a technical program committee member for CLEO USA 2017-20, SPIE Photonics Europe 2018, 2020, and SPIE Optics and Photonics 2018-19.

Yang Liu received the Bachelor's degree in optoelectronics from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2012, and the Master's degree in optical engineering from Wuhan National Laboratory for Optoelectronics (WNLO) in 2015. He received the Ph.D. degree in physics in the Nonlinear Photonics Group at the Centre for Ultrahigh Bandwidth Devices for Optical Systems (CUDOS), School of Physics, University of Sydney, Australia. He is currently a Postdoctoral Researcher leading integrated microwave photonics and hybrid integration research activities in The University of Sydney, Australia. His research interests include the fundamentals and applications of nonlinear optics and microwave photonics in integrated photonics. He was the Vice-President of OSA/SPIE student chapters in the University of Sydney from 2017 to 2018. He was a conference organization committee member for IONS KOALA 2018.

David Marpaung received the Ph.D. degree in electrical engineering from the University of Twente, The Netherlands, in 2009. From 2009 to 2012, he was a Postdoctoral Researcher in the University of Twente within the framework of the European Commission FP7 funded project SANDRA, working on microwave photonic system integration for optical beamforming. From 2015 to 2017, he was a Senior Research Fellow leading the nonlinear integrated microwave photonics research activities in CUDOS, University of Sydney, Australia. In January 2018, he joined the Laser Physics and Nonlinear Optics Group, University of Twente as an Associate Professor. His research interests include RF photonics, photonic integration, nonlinear optics, and phononics. Dr. Marpaung has authored or co-authored 50 refereed journal publications and more than 80 conference papers, and he has delivered 25 invited talks in various international conferences. He is the Technical Program Committee (TPC) Co-Chair of IEEE Topical Meeting of Microwave Photonics (MWP 2018). He received the 2015 Discovery Early Career Research Award from the Australian Research Council and the 2017 Vidi Award from the Netherlands Organisation for Scientific Research.

Khu Vu received the B.Sc. degree from Monash University, Clayton, Vic., Australia, in 2002, and the B.Sc. degree from Australian National University, Canberra, A.C.T., Australia, in 2003. In 2004, he joined the University of Southampton, Southampton, U.K., for his M.phil. degree. He received the Ph.D. degree from Australian National University in 2012. He has been a Postdoctoral Fellow with the Laser Physics Centre, Australian National University. His research interests include

0733-8724 (c) 2019 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/JLT.2019.2931077, Journal of Lightwave Technology

optics and optoelectronics from nonlinear optic, high power fiber laser amplifier, planar optical integration, and optical materials.

Pan Ma is a Postdoctoral fellow of Laser Physics Centre at the Australian National University. She received the Ph.D. degree in optical engineering at Beihang University, Beijing, China in 2016. She was a visiting student at Laser Physics Centre from 2012-2015. Her current research involves planar waveguides, mid infrared photonics, nonlinear optical material and devices.

Duk-Yong Choi biography not available at the time of publication.

Stephen Madden biography not available at the time of publication.

Benjamin J. Eggleton (F'10) received the Bachelor's (Hons.) degree in science and the Ph.D. degree in physics from the University of Sydney, Australia, in1992 and 1996, respectively.In 1996, he joined Bell Laboratories, Lucent Technologies, Murray Hill, NJ,USA, as a Postdoctoral Member of Staff, and then transferred to the Department of Optical Fiber Research. In 2000, he became the Research Director within the Specialty Fiber Business Division of Lucent Technologies. He is currently a Professor of physics at the University of Sydney, the Director of the University of Sydney Nano Institute (Sydney Nano), and Co-Director of the NSW Smart Sensing Network (NSSN). He was previously the Director of the ARC Centre of Excellence for Ultrahigh Bandwidth Devices for Optical Systems (CUDOS). He is the author or co-author of more than 500 journal publications. His research interests include nonlinear optics, all-optical signal processing, optical communications, photonic crystals, and stimulated Brillouin scattering. Dr. Eggleton is a Fellow of the Australian Academy of Science, the Australian Academy of Technology and Engineering (ATSE), and the Optical Society of America. He received the Pawsey Medal from the Australian Academy of Science, the 2004 Malcolm McIntosh Prize for Physical Scientist of the Year, the 2003 International Commission on Optics Prize, the 1998Adolph Lomb Medal from the Optical Society of America, the Distinguished Lecturer Award from the IEEE/Lasers and Electro-Optics Society, and the Re-search and Development 100 Award. He was the President of the Australian Optical Society during 2008–2010, and he is currently an Editor-in-Chief ofAPL Photonics.

0733-8724 (c) 2019 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.

AIP Conference Proceedings

Theoretical model for self-trapping of Gaussian electromagnetic beam in dusty plasma

Cite as: AIP Conference Proceedings **2136**, 060001 (2019); https://doi.org/10.1063/1.5120947
Published Online: 09 August 2019

R. Sharma, and Suresh C. Sharma


View Online

Export Citation



© 2019 Author(s).

Theoretical Model for Self-trapping of Gaussian Electromagnetic Beam in Dusty Plasma

R. Sharma[1, a)] and Suresh C. Sharma[1]
[1]*Department of Applied Physics, Delhi Technological University (DTU), Shahbad Daulatpur, Bawana Road, Delhi-110042, India.*
[a)]sharmaruchi753@gmail.com

Abstract.A theoretical model for self-trapping of Gaussian electromagnetic beam has been developed in dusty plasma. This model comprise of balancing of charge, number density, and energy exchange between plasma components having charge neutrality. Ionization of neutrals, recombination, emission and sticking of electrons to the dust particles, and the collisions between the constituents are considered herein. For numerical appreciation, curves have been drawn for variation of electron density with time and beam waist parameter with the dimensionless length of propagation for different number density and sizes of dust grains. This formulation is helpful for explaining the results of dusty plasma in space and laboratories.

1. INTRODUCTION

Several investigations [1-7] on non-linear propagation of Gaussian electromagnetic beam in dusty plasma have been published in last decades due to its numerous applications in space and industry [8-9].In this paper, non-linear heating of electrons takes place because of non-uniform irradiance of electromagnetic beam in dusty plasma. This causes more and more electrons sticking to the surface of dust grains and generates a depleted density channel which leads to self-trapping of electromagnetic beam in dusty plasma. Ionization of neutrals, recombination, stickiness and electron emission by the dust grains, and the collisions between the constituents are considered in the present model. A wave equation has been evaluated using the perturbed density of electrons and modified permittivity to study the behavior of Gaussian electromagnetic beam in dusty plasma due to non-linear ohmic heating of electrons and curves have been drawn between beam waist parameter and dimensionless length of propagation for distinct values of number density and sizes of dust grains. The remaining manuscript is framed as follows: Sec.2 is designed by the balancing equations of the charge, number density and energy exchange between dusty plasma components to give the dielectric constant of the plasma in the existence of dust grains and solving the wave equation to show the variation of beam waist parameter 'f' with the dimensionless propagation length ζ on account of propagation of electromagnetic beam in dusty plasma and Sec. 3 is structured by results of the present analysis.

2. MODEL

Assume dusty plasma, formed by electrons, ions, neutrals and dust grains having charge Ze (where Z is the charge state of dust and e is the charge on an electron); is non-linearly irradiated by Gaussian electromagnetic beam.

Balancing Equations

Neutral atoms and ions conservations

$$n_i + n_0 = n_{i0} + n_{00}\,, \tag{1}$$

where n_i, n_0 and n_{i0}, n_{00} are the ions and neutrals number densities in the presence and absence of dust grains, respectively.

Charge neutrality

$$Zn_d + n_i - n_e = 0\,, \tag{2}$$

where n_d and n_e are the dust and electrons number densities.

Charge balance of dust grains

$$\frac{dZ}{d\tau} = n_{ic} + n_{ph} - n_{ec}\,, \tag{3}$$

International Conference on Photonics, Metamaterials & Plasmonics
AIP Conf. Proc. 2136, 060001-1–060001-4; https://doi.org/10.1063/1.5120947
Published by AIP Publishing. 978-0-7354-1879-0/$30.00

where $n_{ic} = \pi r_d^2 (8k_B T_i/\pi m_i)^{1/2} n_i \left[1\text{-}Z\alpha_i\right]$ is the sticky current of ions to the dust grain surface, $n_{ph} = \pi r_d^2 n_p$ is the photocurrent emitting from surface of negatively charged dust grains, and $n_{ec} = \pi r_d^2 (8k_B T_e/\pi m_e)^{1/2} n_e \exp\left[Z\alpha_e\right]$ is thesticky current of electrons to the dust grain surface, $\alpha_i = (e^2/r_d k_B T_i)$, $\alpha_e = (e^2/r_d k_B T_e)$, r_d is radius of dust grain, k_B is Boltzmann's constant, T_i and T_e are the ions and electrons temperature, m_i and m_e are the masses of ions and electrons, respectively. n_p represent the degree of photoelectric emission per unit time per unit area from a negatively charged or neutral surface of dust grain.

Number density balance of electrons and ions

$$\frac{dn_e}{d\tau} = \beta_i n_0 - \alpha_r n_e n_i - n_d (n_{ec} - n_{ph}) \tag{4}$$

and

$$\frac{dn_i}{d\tau} = \beta_i n_0 - \alpha_r n_e n_i - n_d n_{ic}\,, \tag{5}$$

where β_i indicates the co-efficient of ionization, $\alpha_r = (\alpha_{r0}(300/T_e)^k)$ cm^3/s represent the recombination coefficient, k and α_{r0} are the constants. The initial terms on RHS of (4) and (5) correlate with the rate of increase of densities due to ionization of neutral atoms. The succeeding two terms in both the equations refer to the rate of density decay due to their recombination, and net sticking to the surface of dust grains.

Energy balance of electrons and ions

$$\begin{aligned}\frac{d}{d\tau}\left(\frac{3}{2} n_e k_B T_e\right) = {} & \beta_i n_0 \varepsilon_e - \alpha_r n_e n_i \left(\frac{3}{2} k_B T_e\right) - n_d (n_{ec}\varepsilon_{ec} - n_{ph}\varepsilon_{ph}) + \frac{n_e e^2 \nu_{eff} U_0^2}{4m_e(\omega_d^2 + \nu_{eff}^2)} \\ & - \nu_{em}\delta_{em} \frac{3}{2} n_e k_B (T_e - T_0) + \nu_{ei}\delta_{ei} \frac{3}{2} n_e k_B (T_e - T_i)\end{aligned} \tag{6}$$

and

$$\begin{aligned}\frac{d}{d\tau}\left(\frac{3}{2} n_i k_B T_i\right) = {} & \beta_i n_0 \varepsilon_i - \alpha_r n_e n_i \left(\frac{3}{2} k_B T_i\right) - n_d n_{ic} \varepsilon_{ic} \\ & - \nu_{im}\delta_{im} \frac{3}{2} n_i k_B (T_i - T_0) + \nu_{ei}\delta_{ei} \frac{3}{2} n_e k_B (T_e - T_i)\end{aligned}\,, \tag{7}$$

where ε_e and ε_i are related to electrons and ions average energy because of ionization of neutral atom, ε_{ec} and ε_{ic} represent the mean value of sticking energy of electrons and ions to the surface of dust grains. ε_{ph} is the energy corresponding to photoemission, δ is the fractional energy loss, and ν_{eff} $(=\nu_{ei}+\nu_{ed}+\nu_{em}+\nu_{sc})$ is the effective collision frequency as a result of collisions of electron with ions, dust, neutrals, and sticking collisions, and T_d is the temperature of dust assumed to be constant due to efficient thermal efficiency.

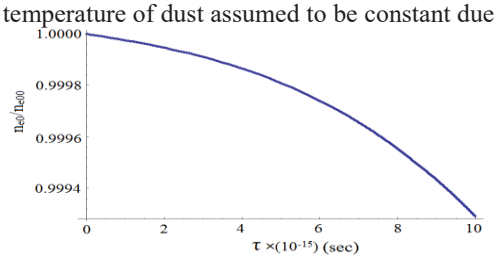


Figure 1. Dependence of dimensionless electrondensity (n_{e0}/n_{e00}) on the parameter τ (sec).

By estimating ionization coefficient and mean energy of electrons and ions through imposing the initial conditions of dust free plasma, (1)-(7) can be concurrently solved for $\tau \to \infty$ in stationary state and the variables n_e, T_e, n_i, T_i, and Z got evaluated numerically. The variation of electron density with time is shown in Fig.1.

Modified Dielectric Permittivity

Using these parameters and paraxial approximations, modified dielectric permittivity is given by

$$\varepsilon_0 = 1-\left(\frac{\omega_0^2}{\omega_d^2}\right)\left(1-\frac{\nu_{eff0}^2}{\omega_d^2}\right)\left(\frac{n_{e0}}{n_{e2}}\right) \text{ and } \varepsilon_2 = \left(\frac{\omega_0^2}{\omega_d^2}\right)\left[\left(1-\frac{\nu_{eff0}^2}{\omega_d^2}\right)\left(\frac{n_{e2}}{n_{e00}}\right)-\left(\frac{n_{e0}}{n_{e00}}\right)\left(\frac{2\nu_{eff0}\nu_{eff2}}{\omega_d^2}\right)\right], \quad (8)$$

where ε_0 and ε_2 are the modified dielectric permittivity, ν_{eff0} and ν_{eff2} are the effective collision frequencies, n_{e0} and n_{e2} are the number densities in the presence of dust grains along axis and radius , respectively., n_{e00} is the electron density in the absence of pulse and dust, and ω_0 and ω_d refer to the electron plasma frequencies with and without dust grains. (1) - (7) are accompanied by unison solution to use $n_{e0}, n_{e2}\ \nu_{eff0}, \nu_{eff2}$ in (8), and to get ε_0 and ε_2, in term $\beta_0 U_{00}^2$.

Propagation of Gaussian Electromagnetic Beam in Dusty Plasma

Assume a y-axial uniformly polarized Gaussian electromagnetic beam propagating in dusty plasma along the z-axis; the cylindrical coordinate of its electric field vector U is defined as

$$U(z)=\hat{j}\, U_0(\rho,z)e^{i\omega_d t}, \quad (9)$$

where $(U_0)_{z=0} = U_{00}e^{\left(-\rho^2/\rho_0^2\right)}$,and $\hat{j}$ is the y-axial unit vector.
Concerned wave equations for the transverse and cylindrically symmetrical beam is given by

$$\frac{\partial^2 \vec{U_0}}{\partial z^2}+\frac{1}{\rho}\frac{\partial \vec{U_0}}{\partial \rho}+\frac{\partial^2 \vec{U_0}}{\partial \rho^2}=\frac{\omega_d^2 \varepsilon_{eff}(\rho,z)}{c^2}\vec{U}_0. \quad (10)$$

Following Akhmanov*et al.*[10] and Sodha *et al*. [11] solution of (10)can be represented as

$$U_0(\rho,z)=A_0(\rho,z)e^{-ikz}, \quad (11)$$

$$A_0(\rho,z)=A_{00}(\rho,z)e^{-iks(\rho,z)}, \quad (12)$$

where $k=(\omega_d/c)\varepsilon_0^{1/2}$. A_{00} and s are real. By employing Ansatz for Gaussian beam $A_{00}^2=\frac{U_{00}^2}{f^2}e^{-\frac{\rho^2}{\rho_0^2 f^2}}$ and solving in paraxial ray approximation using eikonal function $s(\rho,z)$ as $s=(\rho^2/2)\,b(z)+\Psi(z)$, where $\Psi(z)$ is an arbitrary function of z and $b(z)=\frac{1}{f}\frac{df}{dz}$, we get

$$\frac{d^2 f}{d\zeta^2}=-\frac{1}{R_s^2}\frac{1}{f^3}+\frac{1}{f^3}, \quad (13)$$

where $R_s\left(=\frac{2\varepsilon_0\rho_0^2}{\varepsilon_2 U_{00}^2}\right)^{1/2}$ is the typical length of self-trapping, $\zeta=\frac{z}{R_d}$ and $R_d=k\rho_0^2$.The initial term on the right –hand side in (13) represent the nonlinear self-trapping term although later is the diffraction term.

3. RESULTS AND DISCUSSIONS

To study the characteristics of propagation of Gaussian electromagnetic beam in dusty plasma, curves have been drawn for variation in electron density with time and beam waist parameter with the dimensionless length of propagation for distinct number density and sizes of dust grains in the following domain : $\alpha_{r0}=5.5\times10^{-7}cm^3/s$, $n_{e0}=n_{i0}=2.45\times10^6cm^{-3}$, $n_0=n_{00}=0.9\times10^{12}cm^{-3}$, $k=1.1$, $n_d=10^2cm^{-3}$, $\rho_0=5.5\times10^{-2}cm$, $k_B=1.38\times10^{-16}ergsk^{-1}$, $T_{e0}=1001K$, $m_e=9.1\times10^{-28}g$, $m_i=29\times1.6\times10^{-24}g$, $m_d=39\times1.6\times10^{-24}g$, $\omega_d=1.1\times10^8s^{-1}$, $e=4.8\times10^{-10}$statcoloumb , $T_d=300K$, $r_d=10^{-5}cm$, $U_{00}=10^5$ statvolt cm^{-1} and 6×10^5 statvolt cm^{-1}.

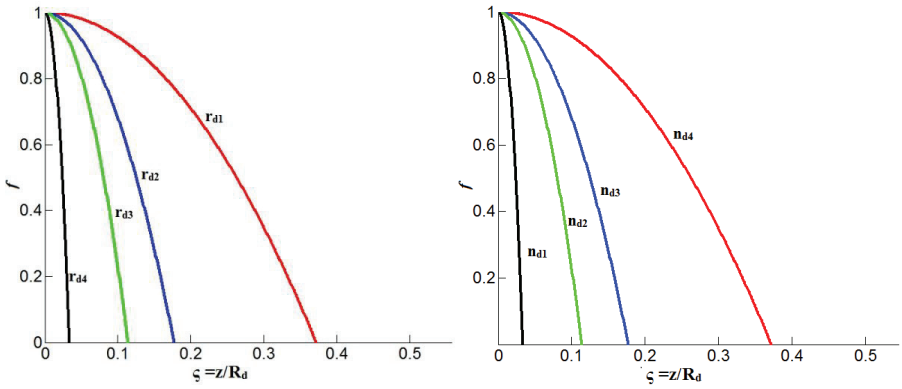


Figure 2. Deviation in beam waist parameter (*f*) with the dimensionless length of propagation $\zeta = z/R_d$ for distinct values of dust grain sizes corresponds to $r_{d1}=4\times10^{-6}cm$, $r_{d2}=6\times10^{-6}cm$, $r_{d3}=8\times10^{-6}cm$, $r_{d4}=10\times10^{-6}cm$ in (a) and number density of dust grains corresponds to $n_{d1}=10^3cm^{-3}$, $n_{d2}=10^4cm^{-3}$, $n_{d3}=10^5cm^{-3}$, $n_{d4}=10^6cm^{-3}$in(b).

Figure1indicates the electron density decay with time due to their larger stickiness to the surface of dust grains with its embryonic growth. Figure 2 show the self-trapping of the beam increases with increasing dust grain size but reverse the case of number density. This is because of increase in non-linearity on account of predominant self-trapping term with the increasing dust grain size and decreasing number density.

ACKNOWLEDGEMENTS

The authors are thankful to Prof. M.S. Sodha for remarkable ideas. The author (Ruchi Sharma) is thankful to her colleagues for their supportive approach during the composition.

REFERENCES

[1] M.S. Sodha, S. K. Mishra, and Shikha Misra, Phys.Plasmas **16**, 123701(2009).
[2] Ruchi Sharma and Suresh C. Sharma, Contrib. Plasma Phys. **59**, 211–225(2019).
[3] Y. Hong , C.Yuan , J. Jia , R. Gao , Y. Wang, Z. Zhou , X. Wang , H. Li , and J. Wu, Phys.Sci.Technol.**19**, 055301 (2017).
[4] Jieshu Jia, Chengxun Yuan, Ruilin Gao, Ying Wang, Yaoze Liu, Junying Gao, Zhongxiang Zhou, Xiudong Sun, Jian Wu and Hui Li, and Shaozhi Pu,J.Phys.D: Appl.Phys.**48**, 465201 (2015).
[5] M. S. Sodha, A. Dixit, S. Srivastava, S. K Mishra, M. P. Verma, and L. Bhasin, Plasma Sources Sci. Technol. **19,**015006 (2010).
[6] Suresh C. Sharma, A.Gahlot, and R.P.Sharma,Phys. Plasmas **15**, 043701 (2008).
[7] S. K. Mishra, Shikha Misra, and M. S. Sodha, Phys. Plasmas**18,** 043702(2011).
[8] R.L. Merlino and J.A. Goree, Physics Today **57**,32-38 (2004).
[9] A. Barkan, N. D'Angelo, and R.L. Merlino, Planet. Space Sci. **43**, 905-908(1995).
[10] S. A. Akhamanov, A. P. Sukhorukov, and R. V. Khokhlov, Sov. Phys. Usp. **10,** 619(1968).
[11] M.S.Sodha, A.K.Ghatak, and V.K.Tripathi, Prog.Optics (North Holand and Co.)**13**, 169(1976).