

Devendra Kumar Sharma ·  
Valentina Emilia Balas · Le Hoang Son ·  
Rohit Sharma · Korhan Cengiz *Editors*

# Micro-Electronics and Telecommunication Engineering

Proceedings of 3rd ICMETE 2019

# Lecture Notes in Networks and Systems

Volume 106

## Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences,  
Warsaw, Poland

## Advisory Editors

Fernando Gomide, Department of Computer Engineering and Automation—DCA,  
School of Electrical and Computer Engineering—FEEC, University of Campinas—  
UNICAMP, São Paulo, Brazil

Okyay Kaynak, Department of Electrical and Electronic Engineering,  
Bogazici University, Istanbul, Turkey

Derong Liu, Department of Electrical and Computer Engineering, University  
of Illinois at Chicago, Chicago, USA; Institute of Automation, Chinese Academy  
of Sciences, Beijing, China

Witold Pedrycz, Department of Electrical and Computer Engineering,  
University of Alberta, Alberta, Canada; Systems Research Institute,  
Polish Academy of Sciences, Warsaw, Poland

Marios M. Polycarpou, Department of Electrical and Computer Engineering,  
KIOS Research Center for Intelligent Systems and Networks, University of Cyprus,  
Nicosia, Cyprus

Imre J. Rudas, Óbuda University, Budapest, Hungary

Jun Wang, Department of Computer Science, City University of Hong Kong,  
Kowloon, Hong Kong

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

**\*\* Indexing: The books of this series are submitted to ISI Proceedings, SCOPUS, Google Scholar and Springerlink \*\***

More information about this series at <http://www.springer.com/series/15179>

Devendra Kumar Sharma · Valentina Emilia Balas ·  
Le Hoang Son · Rohit Sharma · Korhan Cengiz  
Editors

# Micro-Electronics and Telecommunication Engineering

Proceedings of 3rd ICMETE 2019

 Springer



*Editors*

Devendra Kumar Sharma  
Department of Electronics  
and Communication Engineering  
SRM Institute of Science and Technology,  
Delhi NCR Campus  
Ghaziabad, India

Valentina Emilia Balas  
Department of Automatics  
and Applied Software  
“Aurel Vlaicu” University of Arad  
Arad, Romania

Le Hoang Son  
Department of Multimedia  
and Virtual Reality, Institute  
of Information Technology  
Vietnam National University  
Hanoi, Vietnam

Rohit Sharma  
Department of Electronics  
and Communication Engineering  
SRM Institute of Science and Technology,  
Delhi NCR Campus  
Ghaziabad, India

Korhan Cengiz  
Department of Electrical and Electronics  
Engineering  
Trakya University  
Edirne, Turkey

ISSN 2367-3370

ISSN 2367-3389 (electronic)

Lecture Notes in Networks and Systems

ISBN 978-981-15-2328-1

ISBN 978-981-15-2329-8 (eBook)

<https://doi.org/10.1007/978-981-15-2329-8>

© Springer Nature Singapore Pte Ltd. 2020, corrected publication 2021

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

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

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

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

# Preface

The book presents high-quality papers from the Third International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE 2019). It discusses the latest technological trends and advances in major research areas such as micro-electronics, wireless communications, optical communications, signal processing, image processing, Big Data, cloud computing, artificial intelligence, and sensor network applications. This book includes the contributions of international scientists, researchers, and engineers from both academia and the industry. The contents of this volume will be useful to researchers, professionals, and students alike.

Ghaziabad, India  
Arad, Romania  
Hanoi, Vietnam  
Ghaziabad, India  
Edirne, Turkey

Devendra Kumar Sharma  
Valentina Emilia Balas  
Le Hoang Son  
Rohit Sharma  
Korhan Cengiz

# Contents

<b>Radio Direction-Finding Techniques for an Unmanned Aerial Vehicle</b> . . . . .	1
Henricus Augustus Cook, Mohamed Tariq Ekeramodien Kahn and Vipin Balyan	
<b>Wine Quality Analysis Using Machine Learning Algorithms</b> . . . . .	11
Mahima, Ujjawal Gupta, Yatindra Patidar, Abhishek Agarwal and Kushall Pal Singh	
<b>Internet Traffic Detection and Classification Using Machine Learning</b> . . . . .	19
Mrudul Dixit, Ritu Sharma, Saniya Shaikh and Krutika Muley	
<b>Secure Intelligent Optimized Link Heuristic in Cross-Network Handover for IoT</b> . . . . .	31
Anita Sethi and Sandip Vijay	
<b>Remote Monitoring of Vital Parameters with IoT-Based Sensing System</b> . . . . .	41
Rohini, M. A. Ansari and Nidhi Singh Pal	
<b>Economic Load Dispatch Using PSO</b> . . . . .	51
Satyam Tiwari, Nidhi Singh Pal, M. A. Ansari, Dilip Yadav and Nivedita Singh	
<b>Effective Vibration Damping Using Self-tuning Smart Material</b> . . . . .	65
Gayatri R. More and Sharada N. Ohatkar	
<b>Performance Analysis of BIPV Solar Panel Under the Effect of External Conditions</b> . . . . .	75
Ravi Sagar, Nidhi Singh Pal, M. A. Ansari, Nivedita Singh and Dilip Yadav	

<b>Classification of Prediabetes and Healthy Subjects in Plantar Infrared Thermal Imaging Using Various Machine Learning Algorithms . . . . .</b>	<b>85</b>
Usharani Thirunavukkarasu and Snehalatha Umapathy	
<b>Fog Computing using Interoperability and IoT Security Issues in Health Care . . . . .</b>	<b>97</b>
P. Karthika, R. Ganesh Babu and P. A. Karthik	
<b>Comparison of Manual and Semi-automated Method in Measurement of Joint Space Width Measurement in Feet Region of RA Patients . . . .</b>	<b>107</b>
Meghna Sampath, Snehalatha Umapathy, Sakshi Srivastava and Nelufer Shamsudim	
<b>Translation into Pali Language from Brahmi Script . . . . .</b>	<b>117</b>
Neha Gautam, Soo See Chai and Megha Gautam	
<b>The Dataset for Printed Brahmi Word Recognition . . . . .</b>	<b>125</b>
Neha Gautam, Soo See Chai and Megha Gautam	
<b>A Fractal Boundary Wideband Antenna with DGS for X-Band Application . . . . .</b>	<b>135</b>
Shashi Bhushan Kumar and P. K. Singhal	
<b>An Approach to Automated Spam Detection Using Deep Neural Network and Machine Learning Classifiers . . . . .</b>	<b>143</b>
Shubham Vashisth, Ishika Dhall and Garima Aggarwal	
<b>Analysis and Implementation of IWT-SVD Scheme for Video Steganography . . . . .</b>	<b>153</b>
Urmila Pilonia and Prinima Gupta	
<b>Handling Sparsity in Cross-Domain Recommendation Systems: Review . . . . .</b>	<b>163</b>
Nikita Taneja and Hardeo Kumar Thakur	
<b>Enabling Edge Computing in an IoT-Based Weather Monitoring Application . . . . .</b>	<b>175</b>
Kavita Srivastava and Sudhir Kumar Sharma	
<b>Green Cloud Job Scheduling and Load Balancing Using Hybrid Biogeography Based Optimization and Genetic Algorithm: A Proposed Approach . . . . .</b>	<b>185</b>
Yashika Sharma and Sachin Lakra	
<b>Detection of Hazardous Analyte Using Transparent Gate Thin-Film Transistor . . . . .</b>	<b>197</b>
Ajay Kumar, Amit Kumar Goyal, Manan Roy, Neha Gupta, MM Tripathi and Rishu Chaujar	

**A Robustness Analysis of Different Nonlinear Autoregressive Networks Using Monte Carlo Simulations for Predicting High Fluctuation Rainfall** ..... 205  
 Tien-Thinh Le, Binh Thai Pham, Vuong Minh Le, Hai-Bang Ly and Lu Minh Le

**Daily Rainfall Prediction Using Nonlinear Autoregressive Neural Network** ..... 213  
 Vuong Minh Le, Binh Thai Pham, Tien-Thinh Le, Hai-Bang Ly and Lu Minh Le

**Demographic Analysis of World Population Using Various Predictive Algorithms** ..... 223  
 Ishan Bhatt and Kartik Ramaswamy

**Attribute-Based Access Control Schemes in Cloud: Performance and Research Directions** ..... 235  
 S. Sabitha and M. S. Rajasree

**Junctionless Gaussian Doped Negative Capacitance SOI Transistor: Investigation of Device Performance for Analog and Digital Applications** ..... 245  
 Hema Mehta and Harsupreet Kaur

**Development of a Down-Converter for a Software-Defined Radio Environment** ..... 255  
 Bruce Alistair van Niekerk, Tariq Ekeramodien Kahn and Vipin Balyan

**Analysis for Time-Synchronized Channel Swapping in Wireless Sensor Network** ..... 265  
 M. Prathap, Gashaw Bekele, Melkamu Tsegaye and P. Karthika

**Adiabatic Design Implementation of Digital Circuits for Low Power Applications** ..... 275  
 Bhavika Mani, Shaloo Gupta and Hemant Kumar

**A Review: Emerging Trends of Big Data in Higher Educational Institutions** ..... 289  
 Raza Hasan, Sellappan Palaniappan, Salman Mahmood, Vikas Rao Naidu, Aparna Agarwal, Baldev Singh, Kamal Uddin Sarker, Ali Abbas and Mian Usman Sattar

**Realization of a Dual Stop Band for ‘S’ and ‘X’ Bands From the Complimentary Geometry of the Dual Pass Band (S and X Band Application) FSSs** ..... 299  
 Amanpreet Kaur and Komalpreet Kaur

<b>Defect Prediction of Cross Projects Using PCA and Ensemble Learning Approach</b> .....	307
Lipika Goel, Mayank Sharma, Sunil Kumar Khatri and D. Damodaran	
<b>Performance Analysis of Square and Triangular CNT Bundle Interconnects Driven by CNTFET-Based Inverters</b> .....	317
P. Uma Sathyakam, Ananyo Banerjee and P. S. Mallick	
<b>A Current Tunable Third-Order Oscillator Using CCDDCC</b> .....	325
Sunil Kumar Yadav, Manoj Joshi and Ashish Ranjan	
<b>Parametric Classification of Dynamic Community Detection Techniques</b> .....	333
Neelu Chaudhary and Hardeo Kumar Thakur	
<b>A Low-Profile Compact Ultra-Wideband Antenna for Wireless Applications</b> .....	341
Preeti Pannu and Devendra Kumar Sharma	
<b>Weather Monitoring System Using Smart Sensors Based on IoT</b> .....	351
Suresh Kumar, M. A. Ansari, Shivam Pandey, Pragati Tripathi and Mukul Singh	
<b>Brain Tumor Detection Using Image Processing Based on Anisotropic Filtration Techniques</b> .....	365
Aditya Garg, Aditya Bajaj and Roshan Lal	
<b>Sales Terminal Interactive Device for Disabled People</b> .....	375
Priyam Shah, Harsh Patel, Roshni Rao and Manan Shah	
<b>Ranking of E-Commerce Sites in India Using Decision-Making Approach</b> .....	387
Kajal Sharma and Sanjay Kumar Dubey	
<b>Computing Diet Composition for Patients with High Cholesterol Using Decision-Making Approach</b> .....	395
Garima Rai and Sanjay Kumar Dubey	
<b>Automatic Vehicular Number Plate Recognition (VNPR) for Identification of Vehicle Using OCR and Tesseract</b> .....	403
J. S. Nirmala, Rahul Banerjee and Rajath S. Bharadwaj	
<b>High-Frequency CNTFET-Based Voltage-Controlled Oscillator for PLL Application</b> .....	413
Yogesh Kumar, Ashish Raman, Ravi Ranjan and R. K. Sarin	
<b>Trajectory Tracking Control of Unmanned Aerial Vehicle for Autonomous Applications</b> .....	421
Aditi Zear and Virender Ranga	

**Evaluation and Study of IoT Entrances** . . . . . 431  
 E. Sai Sravani, A. V. Sreehitha, A. Konda Babu and Durgesh Nandan

**Survey on the Impact of FSM Design for High-Performance Architecture Evaluation** . . . . . 439  
 K. Sowmya, P. Bujji Babu and Durgesh Nandan

**A Taxonomy and Survey of Data Partitioning Algorithms for Big Data Distributed Systems** . . . . . 447  
 Quadri Waseem, Mohd Aizaini Maarof, Mohd Yazid Idris and Amril Nazir

**Voice-Based Automation Control Platform for Home Electrical Devices** . . . . . 459  
 Constance Izuchukwu Amannah and Promise Nlerum

**Modeling and Simulation of Inertial Navigation System** . . . . . 471  
 Madhavi Vedpathak, Prachi Mukherji and Balkrishna Prasad

**Dual Watermarking for Colour Images’ Copyright and Authentication Using DWT Technique** . . . . . 483  
 Rajesh Thakare, Sandeep Kakde and Prashant Mani

**A Transition Toward Green IT: An Initiative** . . . . . 493  
 Yashika Sharma and Sachin Lakra

**Vision-Based Real-Time Human–Computer Interaction on Hand Gesture Recognition** . . . . . 499  
 Poorvika Singh Negi, Riya Pawar and Roshan Lal

**Investigations of Rectangular Dielectric Resonating Antenna Excited by CPW Feed** . . . . . 509  
 Alina Khan and Sovan Mohanty

**Design and Performance Analysis of Circular Microstrip Patch Array (2 × 2) for S-Band Wireless Applications** . . . . . 521  
 Saptarshi Gupta, Neeta Awasthy and R. L. Sharma

**A Secure Key Agreement Protocol for Data Communication in Public Network Based on the Diffie-Hellman Key Agreement Protocol** . . . . . 531  
 Chukhu Chunka, Subhasish Banerjee, Soumyajit Nag and Rajat Subhra Goswami

**A Secure Steganographic Scheme Based on Chaotic Map and DNA Computing** . . . . . 545  
 Bhaskar Mondal

**Rice Disease Detection and Classification Using Deep Neural Network Algorithm** . . . . . 555  
 S. Ramesh and D. Vydeki

<b>Relative Investigation of Methods to Generate Millimeter Wave in Radio-Over-Fiber Communication</b> .....	567
M. Vinoth Kumar and Vinod Kumar	
<b>A Compact Microstrip Patch Antenna for Mobile Communication Applications</b> .....	575
A. Sanega and P. Kumar	
<b>Microstrip Patch Antenna with Enhanced Gain for 2.4 GHz Wireless Local Area Network Applications</b> .....	583
N. L. Nhlengethwa and P. Kumar	
<b>An Investigation on Drain Current of Junction and Junctionless Surrounding Gate MOSFET</b> .....	593
Aditya Agarwal, R. L. Sharma and Prashant Mani	
<b>Improving the Performance of Video Content Genuineness Using Convolution Neural Network</b> .....	601
Bharat Gupta, Vasvi Bajaj, Rajat Bhusan Panda and Lalit Garg	
<b>A Novel IN-Gram Technique for Improving the Hate Speech Detection for Larger Datasets</b> .....	611
Bharat Gupta, Nikita Goel, Dhruv Jain and Namita Gupta	
<b>Design of Configurable Analog Block-Based Oscillator and Possible Applications</b> .....	621
Kushaagra Maheshwari, Sudhanshu Maheshwari and Piyush Yadav	
<b>Enhanced Reliability of Polarity Controllable–Ferroelectric–FETs under the Impact of Fixed Trap Charges</b> .....	631
Priyanka Pandey and Harsupreet Kaur	
<b>Particle Swarm Optimization for Training Artificial Neural Network-Based Rainfall–Runoff Model, Case Study: Jardine River Basin</b> .....	641
Vikas Kumar Vidyarthi and Shikha Chourasiya	
<b>Text Generation Using Long Short-Term Memory Networks</b> .....	649
Ishika Dhall, Shubham Vashisth and Shipra Saraswat	
<b>Socio-medic Drone with Integrated Defibrillator</b> .....	659
Shivam Pandey, Rahul Kumar Barik, Aritra Karan, P. Phani Kumar, D. Haripriya and N. Kapileswar	
<b>Big Data Processing Based on Machine Learning for Multi-user Environments</b> .....	669
Kamel H. Rahouma and Farag M. Afify	
<b>Integrating Chatbot Application with Qlik Sense Business Intelligence (BI) Tool Using Natural Language Processing (NLP)</b> .....	683
Vipul Vashisht and Pankaj Dharia	



**Area Efficient Multilayer Designs of XOR Gate Using Quantum Dot Cellular Automata** ..... 693  
 Rupali Singh and Devendra Kumar Sharma

**Improved Design of Digital IIR Second-Order Differentiator Using Genetic Algorithm** ..... 707  
 Amit Bohra, Rohit Sharma and Vibhav Kumar Sachan

**Easy Synthesis of Nanostructures of ZnO and ZnS for Efficient UV Photodetectors** ..... 713  
 Vipin Kumar, Ishpal Rawal and Vinod Kumar

**Smart Government E-Services for Indian Railways Using Twitter** ..... 721  
 Mukta Goyal, Namita Gupta, Ajay Jain and Deepa Kumari

**Correction to: Smart Government E-Services for Indian Railways Using Twitter** ..... C1  
 Mukta Goyal, Namita Gupta, Ajay Jain and Deepa Kumari

**Author Index** ..... 733

## About the Editors

**Dr. Devendra Kumar Sharma** specializes in VLSI design & Signal Processing. He is currently a Professor & Dean of SRM Institute of Science and Technology, Delhi NCR Campus, Ghaziabad, India. He holds a Ph.D. degree in Electronics & Communication Engineering (VLSI design) from National Institute of Technology, Kurukshetra, India. Dr. Sharma has authored many papers in SCOPUS/SCI index journals and international conferences. His research interests are in VLSI interconnects, electronic circuits, digital design, testing, and signal processing. He has been the Editor to the 2nd International Conference on Microelectronics and Telecommunication. He is a member of professional societies like LISTE, LIETE, and VSI and is reviewer of many international journals/conferences. He has served as General Chair and Convener of the 2nd International Conference on Microelectronics and Telecommunication (ICMETE-2018) held in India. Dr. Sharma has participated in many international and national conferences as Session Chair and a member in Steering, Advisory or International Program Committees.

**Dr. Valentina Emilia Balas** is currently a Full Professor in the Department of Automatics and Applied Software at the Faculty of Engineering, “Aurel Vlaicu” University of Arad, Romania. She holds a Ph.D. in Applied Electronics and Telecommunications from Polytechnic University of Timisoara. Dr. Balas is author of more than 250 research papers in refereed journals and international conferences. Her research interests are in intelligent systems, fuzzy control, soft computing, smart sensors, information fusion, modeling, and simulation. She is the Editor-in-Chief to International Journal of Advanced Intelligence Paradigms (IJAIP) and to International Journal of Computational Systems Engineering (IJCSysE), a member in editorial board member of several national and international journals, and the director of Intelligent Systems Research Centre in Aurel Vlaicu University of Arad. She is a member of EUSFLAT and SIAM and a senior member IEEE, a member in TC –Fuzzy Systems (IEEE CIS), a member in TC – Emergent Technologies (IEEE CIS) and a member in TC – Soft Computing (IEEE SMCS). E-mail: [balas@drbalas.ro](mailto:balas@drbalas.ro).

**Dr. Le Hoang Son** obtained the Ph.D. degree in Mathematics and Informatics at VNU University of Science, Vietnam National University (VNU), in 2013. He has been promoted to an Associate Professor in Information Technology since 2017. Dr. Son worked as senior researcher and vice director at the Center for High Performance Computing, VNU University of Science, Vietnam National University, during 2007–2018. From August 2018, he is the Head of the Department of Multimedia and Virtual Reality, VNU Information Technology Institute, VNU. His major fields include artificial intelligence, data mining, soft computing, fuzzy computing, fuzzy recommender systems, and geographic information system. He is a member of many professional and academic bodies. Dr. Son serves as Editorial Board of Applied Soft Computing (ASOC, in SCIE), International Journal of Ambient Computing and Intelligence (IJACI, in SCOPUS), and Vietnam Journal of Computer Science and Cybernetics (JCC). He is also associated with many other journals in different roles. Dr. Son served as a reviewer for various international journals and conferences. Up to now, he has 131 publications in prestigious journals and conferences including 74 ISI papers (SCI: 18, SCIE: 50, ESCI: 6) and 01 SCOPUS paper and undertaken more than 20 major joint international and national research projects. He has published 6 books and book chapters. So far, he has awarded “2014 VNU Research Award for Young Scientists,” “2015 VNU Annual Research Award,” 2015.

**Dr. Rohit Sharma** is working as an Assistant Professor of Electronics and Communication Engineering, SRM University, Delhi NCR Campus, Ghaziabad, India. He has completed his Ph.D. in Electronics and Communication Engineering from Teerthanker Mahaveer University, Moradabad, India, M.Tech. (Communication Engineering) from Shobhit University, India, and B.Tech. (Electronics and Communication Engineering) from Uttar Pradesh Technical University, Lucknow, India. He has a teaching experience of over 7 years at SRM University and Dewan V.S. group of Institution, India. Dr. Sharma is an active member of ISTE, IEEE, ICS, IAENG, and IACSIT. He is an editorial board member and reviewer of more than 8 international journals and conferences. He has published about 38 research papers in international/national journals and about 09 research papers in international/national conferences. Dr. Sharma has been the Editor to 2nd International Conferences. He is also a reviewer of many international journals/conferences.

**Dr. Korhan Cengiz** received his B.S. degree at Electronics and Communication Engineering in 2008 from Kocaeli University, Turkey. He received his Ph.D. degree in Electronics Engineering in 2016 from Kadir Has University, Turkey. Dr. Cengiz has more than 30 articles related to wireless sensor networks and wireless communications. Dr. Cengiz serves as a TPC member for more than 20 conferences. He is the Editor-in-Chief of 2 journals and the Editor of several journals. His honors include Tubitak Priority Areas Ph.D. scholarship and two best paper awards in conferences ICAT 2016 and ICAT 2018.

# Radio Direction-Finding Techniques for an Unmanned Aerial Vehicle



Henricus Augustus Cook, Mohamed Tariq Ekeramodien Kahn  
and Vipin Balyan

**Abstract** This paper's aim is to improve the operation of tracking and locating transmitters through leveraging of the emergence of cost-effective software-defined radios (SDRs) and the additional degrees of freedom of an unmanned aerial vehicle (UAV) platform. Different radio direction-finding (RDF) systems and techniques are investigated to find a suitable solution that fits within the constraints of a suitable UAV platform.

**Keywords** Software-defined radio · Unmanned aerial vehicle · Radio direction-finding

## 1 Introduction

Radio frequency (RF) monitoring has been attracting more research interest as accurate and legal usage of spectrum becomes more important. It is also of interest to be able to provide protection of legal spectrum which is allocated to licensees at a fee. Thus, interference resulting from illegal use of spectrum can result in degraded quality of service (QoS) if not total service disruption or jamming of operations. Illegal use of spectrum occurs as a result of illegal installation of repeaters without the consent of the regulatory authority or ignorance on the part of consumers where by-products are imported into a country or region which is not licensed for use in that region. This results in the devices interfering with licensed spectrum users and requires intervention from the regulatory authority to mitigate against interference.

The scope of this paper will be detecting, tracking and managing interference in a timely manner with minimal intrusion. Thus, this paper proposes tracking RF signals by using an unmanned aerial vehicle (UAV) platform fitted with a software-defined radio (SDR) and a computing unit configured as a spectrum analyzer. This will assist the operator to track RF signals by getting multiple bearings quickly and

---

H. A. Cook (✉) · M. T. E. Kahn · V. Balyan  
Department of Engineering—Electrical,  
Cape Peninsular University of Technology, Cape Town 7535, South Africa  
e-mail: [cook.henricus@gmail.com](mailto:cook.henricus@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_1](https://doi.org/10.1007/978-981-15-2329-8_1)

safely to triangulate the source and with the additional degrees of freedom be able to transverse multiple floors to isolate the source of the RF signal in minimal time and without intruding on residence.

## ***1.1 Background***

UAVs are flying robotic platforms that have begun to garner a lot of attention lately, and they can be used in a variety of applications depending on how they are equipped [1]. This was again demonstrated in a paper where a UAV was used to test network spectrum [2]. This paper evaluates the use of a low-cost SDR with the use of a computing unit and the RDF techniques [3], to construct a cost-effective RDF system, coupled with a UAV platform and information streamed to a ground station with a suitable telemetry solution for the operator to use [2].

Currently, locating the source of an illegal transmitter in a built-up area is time consuming due to the difficulties caused by reflections from walls and floors resulting in false positives. This in turn leads to intrusion into the wrong properties and invasion privacy. These problems lead to delayed response to the interference resulting in continued disruption of service or service quality resulting in loss of revenue for clients. This study seeks to leverage UAV technology to enhance the capability of the operator to identify and isolate a RF transmitter in the 25 MHz–1.55 GHz frequency range.

## **2 RDF Literature Review**

### ***2.1 Tracking, Radio Direction-Finding Systems and Methods***

Radio direction-finding (RDF) involves the act of determining which direction of an illegal or unknown radio transmitter. RDF systems attempt to determine the incident angle of arrival (AoA) of a radio signal by calculating the time of arrival or the phase of an RF signal [4].

The classical method of using a directional antenna is explored [3] and evaluated with other AoA-type RDF systems which have three main measurement approaches, scanning beam or Doppler RDF, RF amplitude evaluation or simultaneous multiple beam, and phase interferometer techniques (Table 1).

**Table 1** RDF technique evaluation

RDF technique evaluation			
Technique	Complexity	Sensitivity/directivity	Cost
Directional antenna	Low	High/high	Low
Mechanically spinning antenna (doppler DF)	Medium	Medium/low	Medium
Interferometer (phase and time delay) multiple antennae	High	High/high	High

## 2.2 Classical Directional Antennas

Calculating the maximum received voltage of a rotating directional antenna in relation to the direction is the simplest implementation of RDF. Implementing this technique allows the bearing to be determined from the characteristics of the measured signal in relation to the antenna rotation angle. Upon measurement of an incoming signal, the measured voltage results in a directional pattern dependent on the antenna direction (rotation angle). Thus, the signal strength relative to the antenna position or rotation angle is how the bearing is measured [5].

Since the directivity of the receiving antenna is as a result of superimposing partial waves where the phase differences rely on the incoming wave angle, this technique for RDF can be considered a form of phase direction-finding. During manual operation, the antenna is rotated by the operator and the bearing determined based on signal strength and signal pattern by the operator. As a method to automate this process, the directional antenna, while detecting minimum or maximum patterns, is continuously rotated through attaching a motor to the antenna, this received voltage is illustrated in relation to the angle of rotation, and in this way a rotating direction finder is constructed (see Fig. 1) [5].

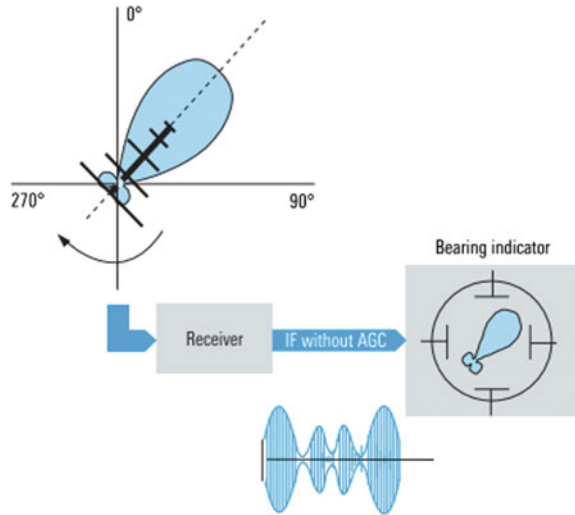
**Advantages and Disadvantages.** The subsequent benefits are typical to variations of this RDF technique:

- Due to the directivity of the antenna, this technique has high sensitivity.
- Since only one receiver is required on a single-channel implementation, the solution is simple and cost-effective implementation.
- The same antenna can be used to perform monitoring as well as direction-finding.

Alternatively, as a result of the directivity of the antenna, as well as the antenna's rotation speed, which is primarily determined by the use of a mechanical rotator, this technique has restricted angular detection range.

- As a result, directivity is linked to probability of intercept.
- This technique fails in the case of intermittent signals, since the rotation of the antenna may not sync up with the signal.

**Fig. 1** RDF using a directional antenna [5]



### 2.3 Doppler Direction Finder

While the antenna turning in a circular fashion which has radius of  $R$ , the measured signal has a frequency  $\omega_0$  which is frequency-modulated with the rotational frequency  $\omega_r$  of the rotating receive antenna as a result of the Doppler effect. The received frequency will increase when the antenna element moves in the direction of the transmitting source, and decrease when the antenna element travels from the transmitting antenna.

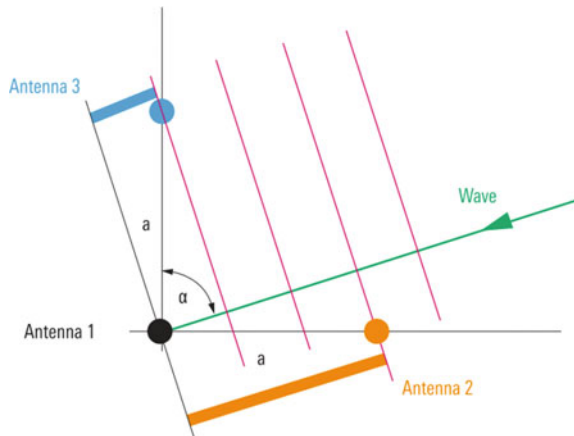
As having antenna element mechanically rotating is not always practical or suitable, numerous elements (monopoles, dipoles, crossed loops) are organized in a circle and then electrically sampled through RF switches also called cyclic scanning [5].

### 2.4 Interferometer (Phase Delay and Time Delay)

The interferometer radio direction finder was initially implemented in the radio astronomy field. With the use of this technique, the resolution power is increased as well as the sensitivity of the RDF system by taking the signals of a few antenna elements that were spaced many wavelengths apart and superimposing those signals [5].

With this phase interferometer technique, at least two receiving antennas are required with a known fixed spacing between them. As the incoming signal arrives at the two different antennae at slightly different times, making the two received signals slightly out of phase. Since this time difference is depending on the spacing

**Fig. 2** Three-element interferometer [5]



and wavelength, the AoA can easily be calculated from the data available. This technique is highly accurate, but at substantial cost, and can be quite cumbersome [4].

With at least three omnidirectional antenna receivers, a calculation of the azimuth and elevation can be made as long as the spacing between antennas is equal to or less than half the wavelength (Fig. 2).

In practice, the three-antenna configuration is improved by adding more antenna elements in order for the distance between the antennae to be adapted to the antenna spacing of  $a > \lambda/2$  for the intended operating frequency range and through this small aperture RDF system accuracy can be increased.

### 3 UAV Literature Review

UAVs are garnering a lot of attention from various industries, hobbyists and researches exploring applications with which to utilize a UAV. Small UAVs are designed to fly at altitudes below 1000 m, and this provides close observation and interaction with ground objects. Low-altitude flight in densely populated areas makes UAVs difficult to pilot and easy to crash. For this reason, a reliable and accurate autopilot system is crucial for small UAVs to effectively perform operations like low-altitude surveillance and free the operator up to perform other tasks.

UAVs typically have two control methods, an autopilot control and a remote control (RC) mode. When using remote control operation, a human operator is required to control the UAV with a handheld radio transceiver, while the autopilot control mode can autonomously maintain a UAV at the desired position or on a preprogrammed flight route. Another control mode which is semiautonomous flight is also available, and this is where the onboard autopilot takes control of the altitude and other flight stabilizing functions and the human pilot navigates the UAV flight path [6].



### 3.1 *Fight Controller*

A flight controller (FC) is the control center for all UAVs. It is essentially the processing board that has all the necessary built-in sensors that detects changes in orientation for controlled flight. In addition, it receives all operator inputs and communicates to the motors via the ESCs in order to relay the operators' intention to the UAV and thereby steering it in the right direction [1]. Typically, most flight controllers have common sensors such as gyroscopes (Gyro) and accelerometer. Other FCs could include more sophisticated sensors such as magnetometer (compass) and barometers (barometric pressure sensors). This allows for more precise control by allowing more accurate reading of the surroundings and environmental effects [7].

The FCs are also a central hub for many other peripherals, such as Global Positioning System (GPS) inputs, LED and sonar sensor, allowing for effective automation and GPS-stabilized flights. This allows for specialized controlled tasks to be performed on UAV platforms since the operator does not have to concentrate continuously on correcting for changes in environmental conditions like wind.

GPS-assisted autopilot systems consist of features such as uncontrolled waypoint flights, loiter in position flights and return home flights. The recommended FCs to use in these applications are Pixhawk, ArduPilot, Naza or iNav flight controllers. These types of FCs are mainly used by aerial photographers and videographers and for autonomous missions. These are the ideal types of FCs for the purpose that is proposed in this paper [7].

### 3.2 *Autopilot Control*

Autopilot system is a tightly integrated mesh involving both hardware and its supportive software. The Sperry Corporation in 1912 developed the first aircraft autopilot [8]. They later demonstrated it with hands-off flight after two years. UAV autopilot systems' main goal is to reliably navigate the UAV to track a reference flight path or route it through various waypoints [6].

UAV autopilot systems are typically closed-loop control systems, and this is made out of two main software modules, the state observer and the controller. In general, the micro-inertial guidance systems act as the state observer, and this includes the gyro, accelerometer and magnetic sensors also known as the microelectromechanical systems (MEMS). The MEMS sensor readings in relation to the GPS location data are fed to a filter to calculate an approximation indicating the current states to be used in the control stage [6].

The state estimator and flight controller (see Fig. 3) are typically comprised of a GPS receiver, a MEMS inertial guidance system and an onboard processor which makes up the UAV autopilot. The state estimation and control input producer are the two essential functions for autopilot systems, as it correlates with the current state.

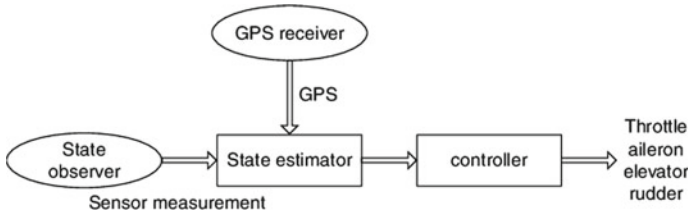


Fig. 3 Functional structure of the UAV autopilot [6]

### 3.3 Software Platforms

The inertial sensor measurements are sent to the autopilot computer for additional filtering and control processing. Depending on the control objectives, the autopilot may subscribe to additional services from the relevant sensors available [6].

**State Observation.** To obtain an effective state observation, the autopilot process will collect all the sensor outputs in real time. This needs to be processed in relation to the waypoint target by passing the state observations for processing [6].

**Open-Source Autopilot.** Due to the popularity in motion control in the gaming industry, MEMS inertial sensors experienced massive growth and subsequent price drop, and this in conjunction with low-cost RC airframes made it possible for several open-source UAV autopilot projects to bloom. Due to the flexibility in the software and hardware, open-source autopilots have a greater advantage over its competition. Researchers can without difficulty change the autopilot based on their own specific project requirements [6].

The Pixhawk platform running the PX4 flight stack software has a fully featured autopilot system. The key benefit of using the Pixhawk is that it operates on a 32-bit architecture compared to the 8-bit architecture implemented by the older APM. In addition, the Pixhawk is using a more powerful CPU with increased memory giving it more processing capacity and speed, and this makes the Pixhawk the superior choice [9, 10].

**Ground Station.** The ground control station (GCS) software application monitors and controls flight paths loaded to the UAV autopilot, which is in charge of controlling the UAV. Although GCS software is the de facto tool for UAV control, not all GCS packages operate with all autopilots. For the Pixhawk autopilot in conjunction with the PX4 FMU, the QGroundControl (QGC) [11] station and MAVproxy [12] are well-integrated control stations. Mission Planner, DroidPlanner and APM Planner are other well-known GCSs that can also be used for similar applications. The aforementioned GCS packages all use the MAVLink protocol but are incompatible with the PX4 flight stack. In addition to being used for creating flight paths, the GCS is also used for controller tuning and sensor calibration [10].

**MAVLink Protocol.** For a successful autonomous mission to be performed, the UAV should be able to transmit and receive information to and from the GCS. For this to work the autopilot's communication protocol, MAVLink is used. Micro Air

Vehicle Link (MAVLink) is a lightweight messaging protocol for interfacing with UAVs to the GCS as well as interfacing onboard on add-on UAV components. To this end, MAVLink employs a design pattern which is a hybrid publish–subscribe model and the point-to-point model [12]. These message structures are defined in XML files. Every XML file sets out a message set used by the specific MAVLink system, which is called the “dialect” [12].

## 4 Results

The main aim is to investigate a cost-effective and compact RDF unit that can be integrated to and leverage of UAV technology. To that end, the focus has to be on two major aspects: one being the RDF techniques and the other being flight automation and hardware and software. Tracking of RF signals with the use of SDRs has been addressed in some form [4]; in addition to the advent of cost-effective UAV technology, it opens the window to combine the two disciplines.

### 4.1 *RDF Technique Evaluation*

In this paper, a comparison is drawn between the classical methods of using a directional antenna explored above, evaluated against other AoA-type RDF systems which have three main measurement approaches, scanning beam or Doppler RDF, RF amplitude evaluation or simultaneous multiple beam, and phase interferometer techniques.

Irrespective of disadvantage of the directional antenna DF technique, this technique of a manually rotated directional antenna is still prevalent nowadays since other techniques are considerably more expensive and a lot more difficult to implement. When operating in the microwave frequency range, the manual RDF technique is most likely the only suitable choice when comparing the noise, gain and expenditure [5]. With the advent of low-cost SDRs and using the various direction-finding techniques as discussed in this paper, one can drastically reduce the size and complexity required by RDF systems [3, 4].

### 4.2 *UAV Evaluation*

For a UAV project of this nature, the key requirements are that the autopilot implemented in the aircraft must fulfill the following criteria:

- Small dimensions and weight;
- Cost-effective;

**Table 2** UAV flight controller evaluation

UAV evaluation			
Properties	Pixhawk	DJI	iNav
Cost	Low	High	Low
Software interface	Open source	API interface	Open source—limited
Advance autopilot features	PX4 autopilot	Yes	Limited
Availability	Local	International	Local

- Waypoint following capabilities;
- Auto-takeoff and landing;
- Configurable;
- Software interfaceable.

An important part of this project is the requirement for interfacing with the flight management unit (FMU). This is only capable with the open-source autopilot systems. This capability is enabled by the use of the PX4 flight stack software and the MAVLink protocol (Table 2).

### 4.3 Proposed Hardware and Software Configuration

For the hardware configuration, it is proposed that a SDR is used to connect to a Raspberry Pi computing unit, and this is to read surrounding RF spectrum data. To obtain location and orientation data, it would be best obtained through a connection with the FC, since it already houses the GPS interface as well as the other sensor peripherals to aid in obtaining orientation, bearing and altitude. Once the platform is airborne, an interface is required between the computing unit and the ground station or operator, and this is to be implemented with the telemetry interface to the ground station.

On the software development, various communication protocols are required to effectively communicate with the Pixhawk and the SDR. This communication is critical since all positioning and orientation information needs to be obtained from the FC, and secondly through the interface with the SDR all related RF spectrum information can be obtained.

## 5 Conclusion

Different RDF systems and techniques were evaluated to find suitable solution that fits within the constraints of a suitable UAV platform. From the results, it is evident that the use of the classical method with directional antenna would be an ideal fit for

this type of application, and this in conjunction with the open-source FMU like the Pixhawk with access to the MAVLink communication protocol opens up the ability to integrate RF data with flight orientation data to build a RDF system for live RF signal tracking.

## References

1. Rahman MF, Ani AI, Yahaya SZ, Hussain Z, Ahmad A (Universiti T. M. P.) (2017). Implementation of quadcopter as a teaching tool to enhance engineering courses. In: 2016 IEEE 8th international conference on engineering education: enhancing engineering education through academia-industry collaboration, ICEED 2016. Institute of Electrical and Electronics Engineers Inc., pp 32–37. <https://doi.org/10.1109/ICEED.2016.7856089>
2. Kuester DG, Jacobs RT, Ma Y, Coder JB (2016) Testing spectrum sensing networks by UAV. In: 2016 United States National Committee of URSI National Radio Science Meeting, USNC-URSI NRSM 2016. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/USNC-URSI-NRSM.2016.7436234>
3. Chen J-C (2004) A virtual RSNS direction finding antenna system. National Defense University Chung Cheng Institute of Technology
4. Schwarz R (2012) Introduction into theory of direction finding
5. Chao H, Cao Y, Chen Y (2010) Autopilots for small unmanned aerial vehicles: a survey. *Int J Control Autom Syst* 8(1):36–44
6. Liang O (2014) Choose flight controller for quadcopter, oscarliang.com. [Online]. Available <https://oscarliang.com/best-flight-controller-quad-hex-copter/>. Accessed 07 May 2017
7. Stevens BL, Lewis FL (2003) Aircraft control and simulation
8. Pixhawk (2018) “Pixhawk” [Online]. Available <https://pixhawk.org/>. Accessed 03 Mar 2018
9. Lidbom A, Kiniklis E (2015) Providence—UAV system to support search and rescue. Chalmers University of Technology
10. QGroundControl (2018) “QGroundControl” [Online]. Available <http://qgroundcontrol.com/>. Accessed: 03 Mar 2018
11. Tridgell A, Barker P, Dade S (2015) “MAVProxy” [Online]. Available <http://ardupilot.github.io/MAVProxy/html/index.html>. Accessed 03 Mar 2018
12. Meier L (2015) MAVLink micro air vehicle communication protocol [Online]. Available <https://mavlink.io/en/>. Accessed 03 Mar 2018
13. Abdessamad W et al (2016) An SDR platform using direction finding and statistical analysis for the detection of interferers. In: 8th international congress on ultra modern telecommunications and control systems and workshops (ICUMT), pp 43–48

# Wine Quality Analysis Using Machine Learning Algorithms



**Mahima, Ujjawal Gupta, Yatindra Patidar, Abhishek Agarwal and Kushall Pal Singh**

**Abstract** Wines are being produced since thousands of years. But, it is a complex process to determine the relation between the subjective quality of a wine and its chemical composition. Industries use Product Quality Certification to promote their products and become concern for every individual who consumes any product. It is not possible to ensure quality with experts with such a huge demand of product as it will increase the cost. Wine-makers need a permanent solution to optimize the quality of their wine. This paper explores the space to easy out and make the whole process cost-effective and more trustworthy using machine learning. It allows to build a model with user interface which predicts the wine quality by selecting the important parameters of wine which play a significant role in determining the wines quality. Random forest algorithm is used in determining wines' quality whose correctness would further be escalated using KNN which makes our model dynamic. Output of this proposed model is used to determine the wines' quality on a scale of Good, Average or Bad. This proposed model can further be applied to several other products which need quality certification. Our prediction model provides ideal solution for the analysis of wine, which makes this whole process more efficient and cheaper with less human interaction.

**Keywords** KNN · CART algorithm · Random forest · Wine quality · Machine learning

---

Mahima (✉) · U. Gupta · Y. Patidar · A. Agarwal · K. P. Singh  
SRM Institute of Science and Technology, NCR Delhi, Sonipat, India  
e-mail: [mahimasaini1997@gmail.com](mailto:mahimasaini1997@gmail.com)

U. Gupta  
e-mail: [audaciousujjawal@gmail.com](mailto:audaciousujjawal@gmail.com)

Y. Patidar  
e-mail: [yatindrapati@gmail.com](mailto:yatindrapati@gmail.com)

A. Agarwal  
e-mail: [a97agarwal@gmail.com](mailto:a97agarwal@gmail.com)

K. P. Singh  
e-mail: [kpall090@gmail.com](mailto:kpall090@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_2](https://doi.org/10.1007/978-981-15-2329-8_2)

# 1 Introduction

Nowadays, demand of wine in market is growing day by day, in order to back up with the rise in demand by accepting new inventions. Along with producing wines, quality assurance certification is also a crucial issue for wine-makers. Currently, within wine industries, quality is estimated through physio-chemical data (e.g. PH levels) and sensory data (e.g. expert critics' involvement) [1]. Analytical data is provided through sensitivity analysis, i.e. the response measured when input variable is compared with its domain value. To produce distinct type and kind of wine producers, use diverse range of grapes and varieties of yeasts. To produce other than flavour, several other factors are also considered to improve quality of wine.

To analyse the quality of wine, a large dataset is taken which consists of huge variety of chemical and acidic aggregation of both red wine and white wine. By occupying smart business science techniques, we can discover the essential and exotic vision which could be productive for better-quality wine; this could be beneficial in economical, in financial and in business sectors in wine production companies [2]. Evolution of business strategies and values added were result shown by grade in excellence in refined model. Resulted produced product can be admiring if and only if it is produced with minimum cost along with maximum quality. The discovered model could be used independently either as wine quality prediction or as a replacement for human wine tasting appraisal by wine critics and could help in development of better wines by industries.

Section 2 is literature survey which describes our research. Section 3 is methodology which is going to tell about the whole process flow and data description which provides insights about the data. Section 4 is result analysis section which is going to tell about the advantages of using this method over the existing methods. Section 5 describes the conclusion and future scope of this model.

## 2 Literature Survey

### 2.1 Documentary Research

Linear regression is easy and simple to implement practically for making predictions in many fields. Using linear regression, the correlation between the attributes was determined. This helped in determining the important parameters with respect to quality [3]. After data analysis, it was found that alcohol shows maximum variation than other parameters. Higher the concentration of alcohol leads to better quality of wine and lowest density [4]. Two different machine learning techniques can be used to develop the prediction model, i.e. neural network and support vector machine. The two used is divided into two parts: red wine and white wine datasets. Both of them consist of 12 different physio-chemical characteristics [5].

While using KNN algorithm, it evaluates Strassen's matrix to calculate the maximum and minimum values of attributes that consist in dataset. K-nearest neighbour, random forests and support vector machines are evaluated on datasets. It shows precision to predict that wine quality can be improved to 90–92% from 75% [6]. How decision tree is formed from the dataset used and mean values are evaluated from 12 different attributes [7].

There are several machine learning algorithms which are analysed to distinguish the quality for both red wine and white wine such as k-nearest neighbour and random forests. The best fortunate to classify data should done using random forest algorithm, where the precision for prediction of good-quality wine is 96% and bad-quality wine is almost 100%, which give overall precisions around 96%. It also helps us to classify different parameters of wine with rating from 1 to 10 or good–bad. From the existing rating, 1–4 predicts bad quality, 5–6 gives average and 7–10 predicts good quality of wine [2].

## 2.2 *Algorithm Analysis*

It gives insights of the dependency of target variables on independent variables using machine learning techniques to determine the wine quality because it gives the best outcome for the assurance of quality of wine. The dependent variable is “quality rating”, whereas other variables, i.e. alcohol, sulphur, etc., are assumed to be predictors or independent variables [6]. While hindering the effectiveness of the data model, various types of errors have occurred like over fitting, introduced from having too large of a training set and bias occur due to too small of a test set.

### 2.2.1 **Random Forest**

CART is a decision tree used for analysing both datasets (red and white wines). CART always generates binary decision trees, which consist of two branches, each for decision node. The tree grows by organizing data in each decision node, by splitting in all possible directions and selecting an optimal split. The decision tree supports a tool based on outcome that accesses a tree-like structure for making decisions and their desirable consequence, along with all the outcome chances, overall resource cost and efficiency. It is one of the ways for demonstration of algorithms which consist of provisional curb statement [4].

It is generally used for exploration operation, especially while analysing decision, helps to identify a strategy that reaches the desired goal and has proved to be an important tool in machine learning [7]. It is like a flowchart structure that consists of internal node which represents “Test” attribute (e.g. if we flip a coin, it comes out to be Heads or Tails), and every Branch represents conclusion for every tested data, as well as leaf node represents label of class (after computing, all parameters' decision is taken). The classification rule is represented from root to leaf.



Random forest is a method of classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees [8].

Following are some of the features of random forest algorithm:

1. It runs efficiently on large databases.
2. It gives estimates of what variables are important in the classification.
3. It generates an internal unbiased estimate of generalization error as the forest building progresses occur. Random forest is similar to the decision tree method in that it builds trees—that is why known as “random forest” [7]. This is a learning method which creates a multitude of decision trees, and outputting the class that occurs most frequently among them and classify the output.

### 2.2.2 K-Nearest Neighbour

This classifier technique is depended on learning by analogy; this means a comparison between a test tuple with similar training tuples. The training tuples are described by  $n$  attributes. Each tuple corresponds a point in an  $n$ -dimensional space. All the training tuples are stocked in an  $n$ -dimensional pattern space. For an unknown tuple, a  $k$ -nearest neighbour classifier searches the pattern space for the  $k$  training tuples that are closest to the unknown tuple.  $K$  training tuples are called as the  $k$ -nearest neighbours of the unknown tuple [2].

“Closeness” is a metric distance, likewise Euclidean distance between two points or tuples, say,  $X_1 = x_{11}, x_{12}, \dots, x_{1n}$  and  $X_2 = x_{21}, x_{22}, \dots, x_{2n}$  is:

$$\text{dist.}(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}.$$

Standalone random forest algorithm is modelling the data with a RMSE of 0.6430 for white wine and 0.6322 for red wine. The proposed system is working in combination with KNN which reduces the RMSE of the above system. Quality formula is developed that connects random forests with KNN algorithm.

## 3 Methodology

This section gives insights the dependency of target variables on independent variables using machine learning techniques to determine the quality of wine because it gives the best outcome for the assurance of quality of wine [4]. The dependent variable is “quality rating,” whereas other variables, i.e. alcohol, sulphur, etc., are

assumed to be predictors or independent variables [4]. The analysis on these variables is done in two different ways:

1. Firstly, regression algorithms are used to depict the importance of each and every independent attributes or predictors.
2. Secondly, random forest and k-nearest neighbour techniques are used to evaluate the value of target variable, i.e. wine quality.

### ***3.1 Model Overview***

This architectural diagram shows the overall process flow along with the components of the system.

#### **3.1.1 Dataset**

Dataset is divided into test set and training set according to the splitting ratio. Training set then is taken to build a machine learning model which establishes the relation from the data of the dataset that issued to predict the quality of wine, i.e. output (Fig. 1).

#### **3.1.2 Model Engineering**

Machine learning model is using the random forest and k-nearest neighbour to build the prediction model. KNN is used dynamically with random forest. Output of random forest is further processed by KNN to predict output.

#### **3.1.3 Training**

Build model will undergo in training phase, which will train the model corresponding to the dataset provided with the help of algorithms used.

#### **3.1.4 Testing**

The final model output undergoes testing for its predicted output with the help of test set that was splatted prior building the model. If the tested output have the desired accuracy and shows it as the output, otherwise it undergoes in training.

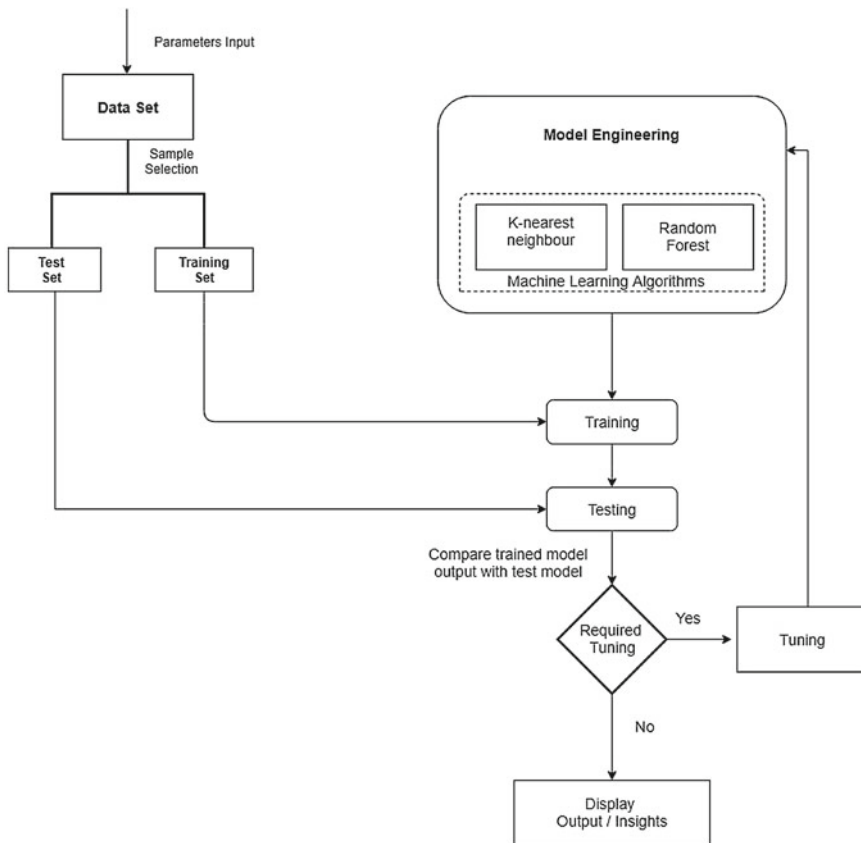


Fig. 1 Architectural diagram

### 3.2 Data Description

To analyse the quality of wine, a large dataset is taken from the research done by Paulo Cortez, in UCI Machine Learning Repository, contributed by University of Minho, Portugal [9]. This dataset consists of chemical information of 6499 types of Portugal wines, in which 4989 varieties are of white wines and 1650 varieties are of red wines. These datasets contain 1599 observations with 12 different feature variables/attributes such as alcohols, residual sugar, chloride, density, free sulphur dioxide, total sulphur dioxide and pH present in both red and white wines [9]. The quality of wine is analysed as follows:

$$\text{Quality} = \alpha_0 + \alpha_1 \text{alcohol} + \alpha_2 \text{volatile acidity} + \alpha_3 \text{density} + \alpha_4 \text{chlorides} + \alpha_5 \text{pH} + \epsilon$$

Conclusion based on the analysis of datasets is as follows:

1. The two most important features among all 12 attributes are sulphur dioxide (both free and total) and alcohol.
2. The most important factor to decide the quality of wine is alcohol; higher concentration of alcohol leads to better quality of wine and lower density of wine.
3. Sulphates are added by wine-makers to prevent spoilage and has positive correlation to wine quality.
4. Volatile acidity contributes to acidic tastes and has negative correlation to wine quality. Citric acid is added to give a freshness test and hence has a positive impact on wine quality.

## 4 Result and Analysis

After literature survey, we found that previously built models were using only single algorithm which gives root mean square error (RMSE) minimum of 0.6430 for white wine and 0.6322 for that of red wine by using random forest, and they were using all the attributes of the wine to build the model (Table 1).

Also, we found the research gap in which we come to know the fact that accuracy of the model can be further raised if we select only important features which play important role in determining wine quality; along this, KNN is used dynamically with random forest which minimizes the RMSE up to 0.541 for white wine and 0.584 for red wine of this model and is capable of predicting bad wine with 100% accuracy and good wine with 92% accuracy. KNN RMSE calculated the same to that of random forest and used along with it (Table 2).

**Table 1** RMSE values for random forest

Algorithm	White wine RMSE	Red wine RMSE
Random forest	0.6430	0.6322

**Table 2** RMSE values for random forest and KNN

Algorithm	White wine RMSE	Red wine RMSE
Random forest	0.541	0.584
K-nearest neighbour	0.541	0.584

## 5 Conclusion and Future Work

The classification tree provides the information that only 43% of the red wine tuples were classified in tree created using red wine dataset. Further, only 58% of white wine tuples were classified in tree created using white wine dataset. The quality value ranges from 4 to 7 which is classified in the decision tree. The limitation of this is that it does not classify extreme quality values, i.e. 0–3 and 7–10. Regression tree among another machine learning algorithms provides the best result with more accuracy. To make wine analyser model more dynamic, KNN algorithm is used through which we can predict quality of any produced wine. The quality value ranges from 4 to 7 which is classified in the decision tree.

Quality of wine is closer to the original value when we use only selected parameters to determine the quality which mainly influences the result. Wine manufacturers can use results to enhance the quality of wine by analysing the ranges in which different constituents should be for best-quality wine.

The parameters used in our dataset form a complex dimensional representation of each type of wine. But it can be possible that there are co-relations that cannot be visible immediately or need some calculation to be more specific and classified, e.g. PH and fixed acidity show a similar relation on quality, and hence, the datasets can be merged, to simplify the problem. As we know random forest tree is the best algorithm to analyse the datasets of wine and shows more accuracy then other algorithm, it can be used to improve the dataset and to reduce the number of dimensions formed while analysing wine quality.

## References

1. Er Y (2016) The classification of white wine and red wine according to their physicochemical qualities. *Int J Intell Syst Appl Eng* 4(1):23–26
2. Executive Summary, Wine Process Monitoring, Wine Quality, Wine Safety, and Wine Complexity (2016) Wine analysis :from ‘Grape to Glass’ an analytical testing digest of the wine manufacturing process
3. Palmer J, Chen B (2018) Wine informatics : regression on the grade and price of wines through their sensory attributes
4. Tajini B, Paris OC (2017) BadrTajini—On campus Paris—DSTI 2017 47(4):547–553
5. Ghosh A (2018) Project report : red wine quality analysis final 3. An empirical red wine quality analysis of the Portuguese ‘Vinho Verde’ wine (2017, 2018)
6. A. Co “Final report,” Apr (2001)
7. Gupta Y (2018) Selection of important features and predicting wine quality using machinelearning techniques. *Procedia Comput Sci* 125:305–312
8. Predictive Model, Linear Regression, and RapidMiner Studio (2018) Building and evaluating a predictive model w/ linear regression in RapidMiner studio
9. Cortez P (2010) Wine quality dataset. <https://archive.ics.uci.edu/ml/datasets/Wine+Quality>

# Internet Traffic Detection and Classification Using Machine Learning



Mrudul Dixit, Ritu Sharma, Saniya Shaikh and Krutika Muley

**Abstract** Growth of Internet resulted in increased number of Internet users along with wide use of Internet. Besides its advantages, the disadvantage of this exponential rise is excess data flooding on the network. To ensure good quality of service, to maintain the speed of Internet, to secure the data flowing on the network, it has become essential to monitor and control the data traffic. Analysis of dataflow involves categorizing it into different types and further filtering it. On the basis of port numbers, payload information, source and destination IP address or statistical information, the data packets are categorized. This paper discusses classification of Internet traffic into different transaction protocols categories, on the basis of statistical parameters such as inter-packet arrival time, time to live, duration of packets and number of packets on the network. Categorizing using statistical parameters prevents invasion of packet data and preserves data privacy. Use of machine learning reduces human intervention in monitoring the Internet traffic. Classification of Internet traffic in the UNSW NB 15 data set is done using five machine learning algorithms, which are K-nearest neighbours, Naïve Bayes, artificial neural network, decision tree and random forest. The aim is to achieve maximum accuracy in minimum execution time. Amongst all algorithms, random forest algorithm gives best result with classification accuracy of 85%. Decision tree requires least execution time and gives accuracy almost equal to random forest algorithm.

**Keywords** Artificial neural network · Decision tree · Internet traffic detection · K-nearest neighbours · Machine learning · Naïve Bayes · Random forest

---

M. Dixit (✉) · R. Sharma · S. Shaikh · K. Muley  
Department of Electronics and Telecommunication, MKSSS's Cummins College of Engineering for Women, Pune, India  
e-mail: [mrudul.dixit@cumminscollge.in](mailto:mrudul.dixit@cumminscollge.in)

R. Sharma  
e-mail: [ritu.sharma@cumminscollge.in](mailto:ritu.sharma@cumminscollge.in)

S. Shaikh  
e-mail: [saniya.shaikh@cumminscollge.in](mailto:saniya.shaikh@cumminscollge.in)

K. Muley  
e-mail: [krutika.muley@cumminscollge.in](mailto:krutika.muley@cumminscollge.in)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_3](https://doi.org/10.1007/978-981-15-2329-8_3)

## 1 Introduction

Internet is network of networks, and Internet traffic is the data flowing through these networks in the form of packets. Internet traffic detection means to identify this data flow and to categorize it into different types. These different types of traffic can be classified using machine learning, to reduce the need of human operator in network management operations. During the training process of model, it establishes relationships between different features of data set. To check the accuracy of trained model, it makes output predictions on the testing data set. Supervised machine learning algorithms train the model using labelled data. In this paper, the approach of identifying Internet traffic is based on statistical parameters, and the detected traffic is classified according to its transaction protocols. The considered statistical parameters are as follows:

- Inter-packet arrival time: as the packets do not arrive continuously at the host, there is delay in between their arrival. Inter-packet arrival time is the time span between arrivals of two packets on the host.
- Time to live (TTL) of the packets: it is the number of hops that a packet is permitted to travel towards the destination before being discarded by a router.
- Packet count: it is the total number of packets travelling from source to destination or vice versa.
- Packet duration: this is the time for which the packet is present on the network.

Internet traffic detection and classification finds its application in static packet filtering firewalls. In these firewalls, packets can be classified on the basis of:

- IP addresses: from the packet header, the source and destination IP addresses are checked. If the transmitted and received packets verify the IP addresses, then the packet is said to be from the authentic source.
- Port numbers: this approach focuses on the port number of the application and then categorizes into that type commonly associated with the port. Example, port number 80 belongs to Hypertext Transfer Protocol (HTTP). This approach is not reliable because it cannot detect dynamic port numbers.
- Protocols: they define the rules and regulations of sharing of information between two or more resources. The incoming and outgoing packets are filtered on the basis of their transaction protocols.

## 2 Methodology

The methodology for Internet traffic detection is as follows:

- Input original data set: UNSW NB-15 data set is generated by Cyber Range Lab in Australian Centre for Cyber Security (ACCS) [1]. The data set comprises of

175,335 records having 49 statistical as well as non-statistical features with 133 different types of transaction protocols.

- Extraction of statistical parameters: out of all available 49 features, this paper uses the following statistical features to train and test the machine learning model:
  1. Source/destination inter-packet arrival time ( $s_{inpkt}/d_{inpkt}$ )
  2. Source to destination/destination to source packet count ( $s_{pkts}/d_{pkts}$ )
  3. Source to destination/destination to source time to live value ( $s_{ttl}, d_{ttl}$ )
  4. Duration of packet (dur).

To prevent dominance of any one feature in resulting output, the inputs are scaled.

- Segregation of data set into input and output: the data set is segregated into input and output. The statistical parameters ( $s_{inpkt}, d_{inpkt}, s_{ttl}, d_{ttl}, s_{pkts}, d_{pkts}, dur$ ) are the input features used to train the machine learning classification model. The results of this model will categorize the testing data according to the transaction protocol used.
- Splitting into training and testing data set: the available data set is divided into training data set and testing data set in the ratio of 7:3, 8:2, 9:1.
- Applying classification algorithm: the paper focuses on five algorithms, namely K-nearest neighbours, Gaussian Naive Bayes, artificial neural network, decision tree and random forest. These algorithms establish relationship between features and target protocols. Thus, the output is predicted by the trained model.
- Comparing results: the results of both the algorithms are compared on the basis of execution time, accuracy, precision, recall and F1 score values [2]. Figure 1 shows the steps for Internet traffic detection.

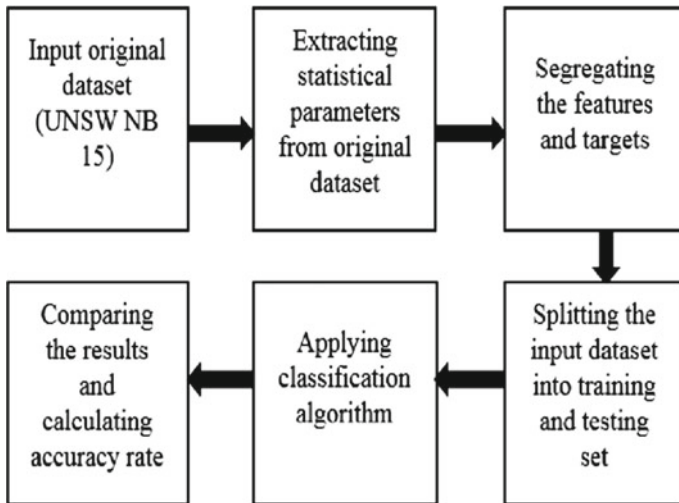


Fig. 1 Block diagram of methodology used to detect Internet traffic



### 3 Machine Learning Algorithms

#### 3.1 Gaussian Naïve Bayes

Naïve Bayes algorithm is based on applying Bayes theorem to detect Internet traffic with assumption that the features are independent of each other [3]. Gaussian classifier assumes that the attributes follow a normal distribution. So, the probability of the new inputs is calculated using the Gaussian probability density function.

The formula of PDF of Gaussian Naïve Bayes algorithm is given as:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)} \quad (1)$$

- $\mu$  is the mean calculated for features of target protocol.
- $\sigma$  is the standard deviation for features of target protocol.
- $x_i$  is target protocol.
- $y$  is sample which contains features of target protocol.
- $P$  is the probability of  $x_i$  such that probability of  $y$  is given.

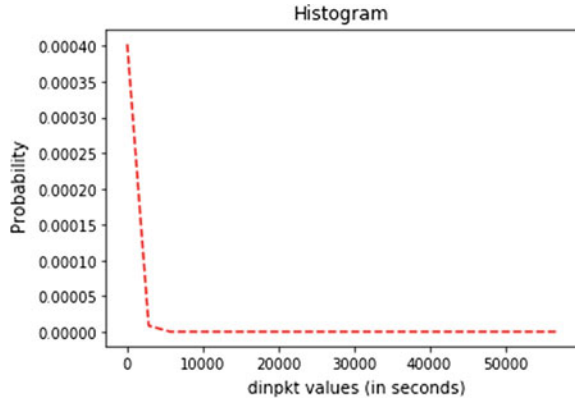
The steps for Naïve Bayes algorithm are as follows:

- In training stage, training data consisting of both input and target output is given to the Naive Bayes classifier.
- The prior probabilities of the target output classes are calculated from the training data.
- The likelihood probabilities of input features, i.e. probability of input features given that probability of target output is known, are calculated.
- In testing stage, testing data consisting of new inputs is been given to the classifier.
- The posterior probability, i.e. probability of output classes given that probability of input features is known, is calculated. Its formula is given as:

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)} \quad (2)$$

- $P(A|B)$  is the posterior probability
- $P(B|A)$  is the likelihood
- $P(A)$  is the class prior probability
- $P(B)$  is the prior probability.

Figure 2 shows the histogram of  $d_{\text{inpkt}}$  feature of the UNSW NB 15 data set.

**Fig. 2** Histogram of  $d_{\text{inpkt}}$ 

### 3.2 *K-Nearest Neighbours*

KNN is a classification algorithm which groups together all the members of data set showing similar properties. KNN algorithm stores the training data set in memory and each time an input from testing data set is to be categorized and its comparison with all other members of training data set is done [4].

To train the model:

- The similarity amongst the members is calculated by finding Euclidian distance between their corresponding parameters.

$$d_{\text{pkts}} = \sqrt{(s_{\text{pkt1}} - s_{\text{pkt2}})^2 + (d_{\text{pkt1}} - d_{\text{pkt2}})^2} \quad (3)$$

$$d_{\text{inpkt}} = \sqrt{(s_{\text{inpkt1}} - s_{\text{inpkt2}})^2 + (d_{\text{inpkt1}} - d_{\text{inpkt2}})^2} \quad (4)$$

$$d_{\text{ttl}} = \sqrt{(s_{\text{ttl1}} - s_{\text{ttl2}})^2 + (d_{\text{ttl1}} - d_{\text{ttl2}})^2} \quad (5)$$

- $d_{\text{inpkt}}$  is the Euclidian distance between inter-packet arrival time of two members.
- $d_{\text{ttl}}$  is the Euclidian distance between time to live of two members of data set.
- $d_{\text{pkts}}$  is the Euclidian distance between number of packets of two members.

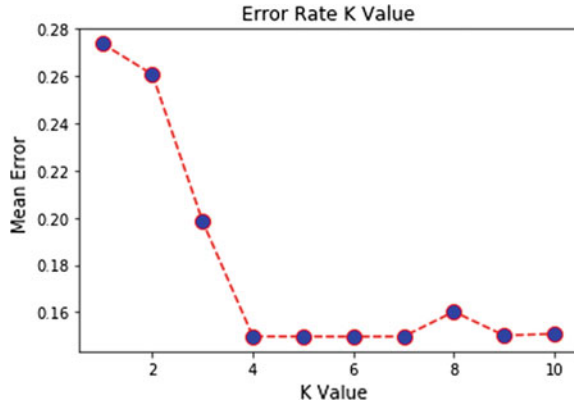
To test the model:

- All above distances are calculated with respect to each member of the data set. Minimum distance indicates maximum similarity.
- Thus, all the members of the training data set showing similar properties are categorized into one group.

Steps for determining the ideal value of number of nearest neighbours:

For values of  $k$  ranging from 1 to 10, the model was trained with KNN algorithm, and its accuracy and error rate were calculated. Error rate is when the predicted value

**Fig. 3** Graph of error versus  $K$ -value



of output is not equivalent to the actual value from the data set. It was found that minimum error rate was obtained for  $k = 6$  value of nearest neighbours. Initially for first three values of  $k$ , error rate is found to be considerably high due to under-fitting. For  $k$  values greater than 6, again error rate increases. This indicates the model is over-fitting due to generalization. Figure 3 shows the graph of error rate versus  $K$  values.

### 3.3 Artificial Neural Network

Artificial neural networks (ANN) are an abstract simulation of a real nervous system that consists of a set of neural units connected to each other. A multilayer perceptron (MLP) is a class of feedforward artificial neural network. It consists of three layers of nodes: an input layer to receive the signal, an arbitrary number of hidden layer and an output layer that makes decision or prediction about the input.

Steps of implementing ANN are as follows:

- Input consists of features of Internet traffic, namely number of packets, arrival time, time to live, etc.
- These input features are then multiplied by weights which are real valued numbers. A bias is added to the summation to set threshold weight that increases the steepness of activation function.
- Activation function decides, whether a neuron should be activated or not. The purpose of the activation function is to introduce nonlinearity into the output of a neuron.
- ReLU activation function equation is:

$$f(x) = \max(0, x)$$

$$\text{ReLU}(x) = \{0 \text{ if } x < 0; x \text{ if } x \geq 0\} \tag{6}$$

- In a neural network, update the weights and biases of the neurons on the basis of the error at the output. This process is known as back-propagation.

### 3.4 Decision Tree

Decision tree has a flow chart-like tree structure in which the nodes are the test on attributes, and the branches are the outcome of the test, while the leaf node represents the target label. The root node is the attribute which has the least uncertainty. The root node splits into nodes until they cannot be split any further. For splitting, the difference between entropy, i.e. purity of a branch before the split and after the split is calculated. This difference is called the information gain. The attribute which has the maximum information gain has been selected to split. The goal is to reduce uncertainty.

Formula of entropy is:

$$\text{Entropy}(S) = \sum_{i=1}^c -p_i \log_2(p_i) \quad (7)$$

- $P_i$  is the proportion of values falling into a class level  $i$
- $c$  is a number of class levels.

The below formula represents information gain.

$$\text{Gain}(S, A) \equiv \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (8)$$

- $A$  is an attribute we want to split
- $S$  is the number of examples
- $S_v$  is the number of examples for current value of attribute  $A$ .

### 3.5 Random Forest Algorithm

This algorithm comes under ensemble supervised learning classification category. It comprises of multiple decision trees, formed on the basis of each feature under consideration [5]. This avoids dominance of any particular feature as the result depends on all of them equally. The advantage of random forest algorithm over decision tree is that instead of increasing depth of a single tree, random forest has multiple trees which avoid over-fitting. Over-fitting means the algorithm fits very closely to a smaller data set, resulting in decreased accuracy and increased complexity. As random forest does not closely fit with data set, a high variance in the data set is

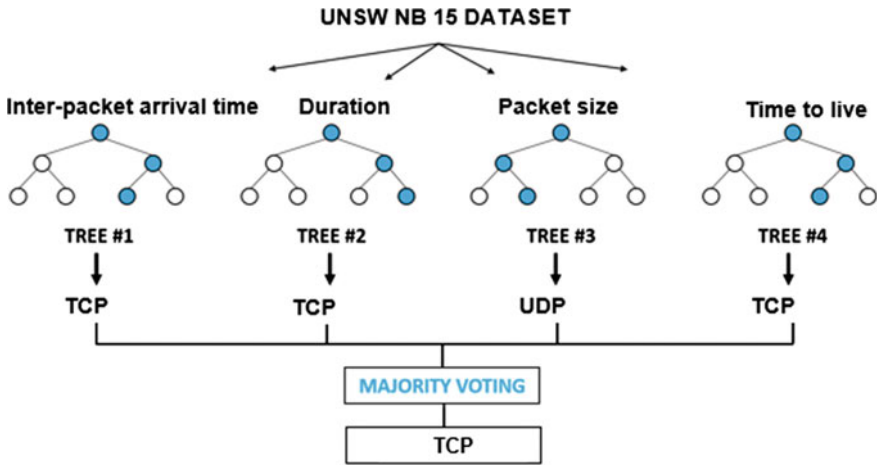


Fig. 4 Random forest algorithm

not generalized, thus making the model more flexible. This algorithm provides good results with nonlinear data.

- Training in Random Forest algorithm:  
the model is trained using concept of bagging that is bootstrap aggregation. In bagging, at the time of making a decision of splitting at a node, any member of data set is chosen randomly. Also, during training, this same member can be chosen again with its replacement again in data set.
- Testing in Random Forest algorithm:  
implementing this algorithm, from the sci-kit learn library provides additional feature of OOB, out of bag testing. Due to this, the members of data set which are used for training according to bagging algorithm will not be used for testing (Fig. 4).

## 4 Results

Table 1 shows the results of all five algorithms.

Minimum accuracy is obtained from Gaussian Naïve Bayes algorithm. Maximum accuracy is obtained for random forest algorithm when 90% of data set is used for training the model. With the increase in percentage of training data set, the accuracy increases. But for KNN maximum accuracy is obtained for 80% training data set and it decreases at 90% training data set due to over-fitting

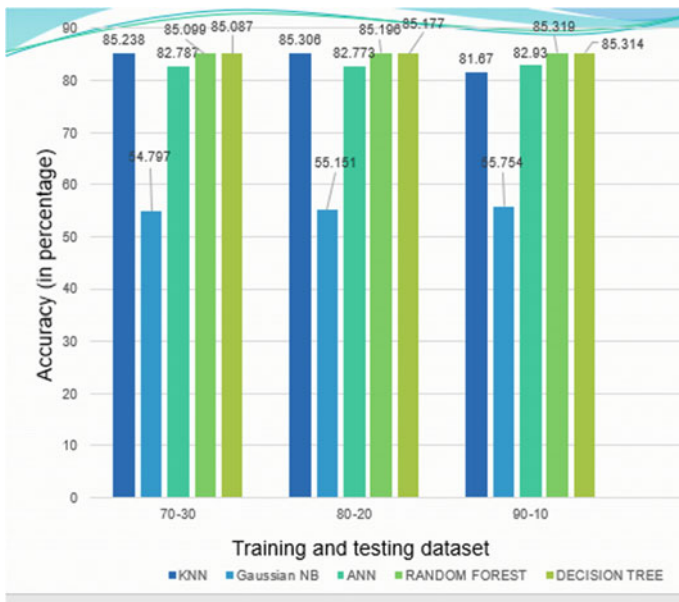
All the above results are computed on Quad-Core 2 GHz speed processor.

Figure 5 shows the comparative analysis of accuracy of all five algorithms.

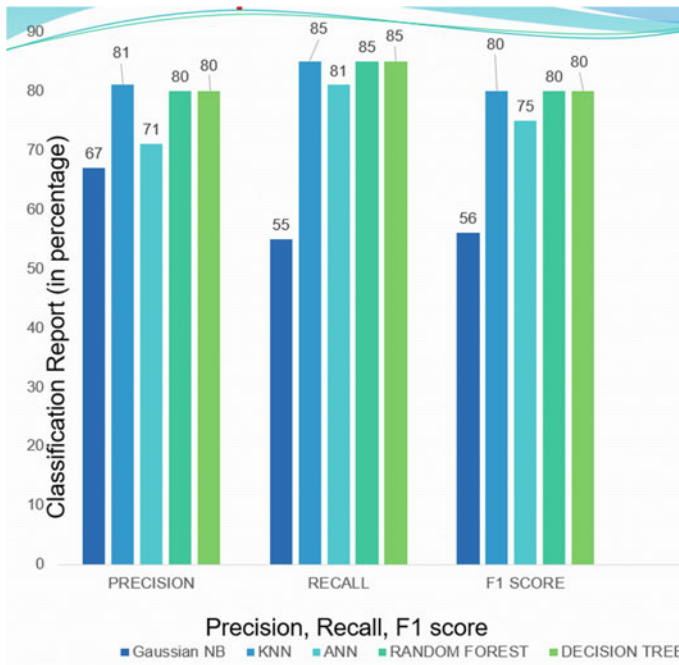
Figure 6 shows comparative study of all five algorithms on the basis of precision, recall and F1 score.

**Table 1** Results

Train-test split data (in %)	Algorithm	Accuracy (in %)	Training time (in seconds)	Testing time (in seconds)
70-30	KNN	85.238	31.50	14.0
	Gaussian NB	54.797	0.380	3.70
	ANN	82.787	51.482	0.272
	Decision tree	85.087	3.336	0.384
	Random forest	85.099	5.132	2.203
80-20	KNN	85.306	45.42	11.69
	Gaussian NB	55.151	0.410	2.660
	ANN	82.773	39.297	0.156
	Decision tree	85.177	2.884	0.0937
	Random forest	85.196	5.813	1.77
90-10	KNN	81.647	61.30	7.190
	Gaussian NB	55.754	0.570	0.750
	ANN	82.930	113.40	0.0937
	Decision tree	85.314	4.583	0.0468
	Random forest	85.319	6.589	0.663



**Fig. 5** Comparative analysis of accuracy of all algorithms



**Fig. 6** Comparative study of all five algorithms on the basis of precision, recall and F1 score

## 5 Conclusion

The project focuses on four statistical parameters out of the 50 parameters available in original data set UNSW NB 15 and successfully achieves accuracy of around 85% with random forest algorithm. Minimum accuracy is obtained from Gaussian Naïve Bayes algorithm. This algorithm assumes the features to be mutually independent. So this assumption proves to be false for UNSW NB 15 data set. Also, the histogram of the features does not follow Gaussian probability distribution curve, which justifies the reason for obtaining low accuracy. Accuracy increases with increase in proportion of training data set, and hence, maximum accuracy is obtained when 90% of the data set is used to train the model. But, for KNN algorithm, maximum accuracy is obtained when 80% of data set is used for training and at 90% training data set the accuracy is found to decrease due to over-fitting. KNN requires maximum training time as compared to other four algorithms. To reduce the execution time, four comparisons are done simultaneously on the Quad-Core machine under consideration. Also, to avoid capturing noise of the data set, the optimum number of neighbours to achieve maximum accuracy was found out to be 6. Gaussian Naïve Bayes requires least training time amongst all five algorithms. Decision tree classifier requires least testing time. Random forest algorithm shows maximum accuracy. Random forest and decision tree classifiers have almost equal values of precision, recall and F1 score.

Artificial neural network gave optimum results for three hidden layers, and rectified linear unit (ReLU) activation function for a batch size of 3000 records iterated 700 times. Since the above results are with respect to UNSW NB 15 static data set, the scope of the project can be further extended by implementing it on other data sets. Also, accuracy can be further increased by ensembling different algorithms.

Thus, Internet traffic classification on the basis of transaction protocols finds its applications in packet filtering firewalls. Deep packet inspection fails when the contents of packet are encrypted. Hence, Internet traffic detection and classification on the basis of statistical parameters overcomes this disadvantage of deep packet inspection and thereby preserves privacy of user data flowing on the network. This makes network management less complex and reduces human intervention in it.

## References

1. Smit D, Millar K (2017) Clinton page, using deep learning to identify internet communications traffic, macquarie matrix: special edition, ACUR 2017. <https://students.mq.edu.au/study/my-study-program/undergraduate-research-journal/acur2017/Smit.pdf>
2. Adekitan AI, Abolade J, Shobayo O (2019) J Big Data 6: 11. <https://doi.org/10.1186/s40537-019-0176-5>
3. Casas P, Fiadino P, D'Alconzo A (2016) Machine-learning based approaches for anomaly detection and classification in cellular network, TMA
4. Lashkari AH, Draper Gil G, Mamun MSI, Ghorbani AAA (2017) Characterization of tor traffic using time based features, ICISSP. [https://www.researchgate.net/publication/314521450\\_Characterization\\_of\\_Tor\\_Traffic\\_using\\_Time\\_based\\_Features](https://www.researchgate.net/publication/314521450_Characterization_of_Tor_Traffic_using_Time_based_Features)
5. Lokhande S, Jaiswal RC Machine learning based internet traffic recognition with statistical approach. <https://ieeexplore.ieee.org/document/6726074>



# Secure Intelligent Optimized Link Heuristic in Cross-Network Handover for IoT



Anita Sethi and Sandip Vijay

**Abstract** Complexity management and system performance optimization are achieved using machine learning in high level of digitized automation systems. In this paper, we worked on enhancement of the link state heuristic performance in terms of reliability, scalability, power consumption and capture effect with outcomes that have demonstrated the usefulness, flexibility and configurability with security. Handover in heterogeneous ad hoc network plays an important role in the performance of the network. As technology advances, massive IoT devices communicate with each other with different technologies and transmit the data to the desired device through cloud or fog by using different technologies. Cross-domain, cross-platform, cross-network with cross-layering can enhance the operation of smart projects. QoS is desired for video and heavy data traffic. In this paper, we presented the handover within HetNet with its performance. Observation shows that link state heuristic for medium and dense area performs better. The impact of queue size, distance and packet size is represented in the graph.

**Keywords** HetNet · Handover · Jitter · Throughput

## 1 Introduction

Real-time interaction arises the creation of on a large scale and complexity level using advanced communication capabilities by connecting billions of devices with petaflops of computing resources. Without human interaction, high degree of intelligent automation is required for operation and management of digitized industries. IoT services, licensed and unlicensed are converged on a single next-generation packet core platform which is automated to simplify service creation. Network resources allotment based on per customer, per service level rather than implementing a network

---

A. Sethi (✉)  
ICFAI University, Dehradun, Uttarakhand, India  
e-mail: [seep4g@gmail.com](mailto:seep4g@gmail.com)

S. Vijay  
Shivalik Engineering College Dehradun, Dehradun, Uttarakhand, India

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_4](https://doi.org/10.1007/978-981-15-2329-8_4)

slicing for all IoT services and another for business services. Cross-domain automation and cross-work network and mass-scale network automation provide service automation for multi-vendor networks. Distributed architecture enables multi-access edge to compute by delivering customization optimization across entire radio access network. Self-organizing network SON supports fully automated machine learning system to deliver improved user satisfaction and expanding the value of radio technology [1].

Numerous smart systems including network connectivity which perform specific tasks or enable use models like smart city, smart grid/utility, industrial automation, building automation, transportation, agriculture, connected supply chain, critical infrastructure protection and control and wide area gaming for industry. Vertical applications required highly engineered network and operate on different lifecycles. Vertical network is required to remain in service for longer time than a service provider network. Over the top, services and advanced mobile applications using Internet connectivity become ubiquitous with high-quality services and are more challenging tasks for mobile networks. Hard real-time applications are limited due to the present networking architecture, for example autonomous driving. The present architecture of VANET in 5G can resolve the challenges of air interface, security, congestion, backhaul networking and mobility management. Intelligently utilization of pre-fetching and caching techniques in fog computing, transforming the simple connectivity provision to content and service provision. Network function virtualization (NFV)/software-defined network (SDN) supports cloudification of network resources in 5G enable the vehicles to bring in cloud [2] (Fig. 1).

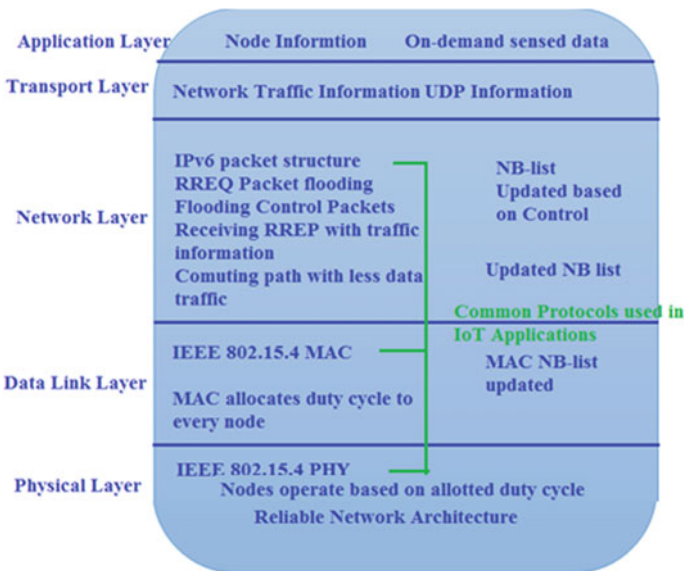


Fig. 1 Application based IoT network layering

At the application layer, CoAP, MQTT, XMPP, RESTFUL Services, AMQP and WebSockets are some of protocols used in communication in IoT architecture. Constrained Application Protocol (CoAP) supports constrained resource devices with the synchronous request/response services keeping the overall implementation lightweight using UDP. Both synchronous and asynchronous responses are used in CoAP protocol in client/server architecture with unicasting and multicasting techniques, as opposed to TCO to reduce the bandwidth requirements [3]. It uses its own mechanism for reliability over UDP connection, and for quality of service, two bits are used in the header and comfortable, non-comfortable, acknowledgement and reset messages are communicated. CoAP was developed for M2M and IoT communications which is lightweight protocol, so no built-in security feature is included. For security purpose, Datagram Transport layer Security (DTLS) runs at the top of UDP and like TLS in TCP. Automatic key management, authentication, cryptography algorithm, confidentiality and data integrity are supported by DTLS [4]. Additional packets increase the network traffic required for handshake purpose and shorten the lifespan of mobile device.

With the objective of lightweight M2M communication, IBM released Message Queue Telemetry Transport (MQTT) protocol runs at the top of TCP stack. Instead of using request/response primitives used in client/server architecture, asynchronous publish/subscribe protocol meets the requirement of IoT environment which uses minimum bandwidth and computational resources' utilization and no requirement of updating client software. Quality of service is maintained by using fire-and-forget, delivered at least once and delivered exactly once signals which support the reliability to network [5] in Extensible Messaging and Presence protocol (XMPP). In network disruption, Advanced Message Queuing protocol (AMQP) provides reliable communication using asynchronous publish/subscribe communication with messaging using store and forward feature. At most once, at least once and exactly once messages ensure reliability.

## 2 State of the Art

Maximization of user data throughput, Cell-EDGE bit rate and minimization of cost per bit and delay for both transmission latency and connection establishment with improved spectral efficiency, reasonable power consumption and smooth mobility are essential features of radio technology. Strengths of both optical and wireless technology provide the seamless operation of network function virtualization, IoE, IoT and mobile content delivery network (CDN) and software-defined network in 5G [6]. Prime objective of this paper is scalability which is a key important factor for due to the transmission range and number of devices may reach a given gateway simultaneously. Due to the random access, MAC protocol combined with large time-bandwidth product reduces the scalability and capturing the performance of node far away from the gateway. Another important factor is the collision of the packets and further degrades the performance of the network. Acknowledgement of

the transmitted frame plays a vital role in performance of the network. Due to duty cycle regulations and lost acknowledgement, reason of external interference, nodes required to retransmit the frame which degrades the network reliability. Resources of downlink channel are more valuable, discouraging any fine-grained scheduling of nodes performed by gateways demands more downlink channel resources [7]. All these factors arise the requirement of a novel heuristic at MAC and network layer to improve the scalability and reliability of heterogeneous network.

Traffic flow plays an important role in 5G network, so modelled it carefully. Distributed data processing, network virtualization, transparency of data and time combined with mobility of node origin unpredictable and dynamic traffic pattern change behaviour. Numerous traffic patterns are generated by traffic source which is executed by traffic shaping applied by network services, SDN functionalities and overlay network, observed under virtual platform, application, IP and physical m/c depending of type of communication (delay-tolerant applications, personal area networks, M2M, client/server, P2P). Modern traffic pattern requires the deployment of intelligent network management heuristics with customized and suitable modelling to handle variety in flow behaviour [8]. Traffic engineering and regulating, failover algorithms, reliability, scheduling, topology revisions are required for management of next-generation networks. Challenge for traffic modelling is dynamic nature and diversity of present services and applications. Some scenarios generate static and probable patterns, e.g. software updates or replicated storage backups, and other generate difficult to predict and high variable patterns, M2M data flows, sensor data and distributed media services. For example, dense Zigbee sensor network, although generating payload of few bytes in simulation, can generate high intense traffic in real-time applications. In this scenario, traffic pattern generated in and across data centre is very complex and represents big data applications which require high security and investigation.

Numerous performance metrics have been defined for routing protocols that can be partitioned into node and link metrics like throughput, link quality, latency, hop count, delay, Jitter, node energy. In this paper, we are addressing the congestion problem in MANETs and enhancing the network performance when congestion occurs. In the previous chapter, we analytical model for optimization and find out network conditions with congestion [9]. Most of the packets are lost in the network due to poor scheduling, and high network traffic which is not properly handled by queue management heuristic and overflow occurs. Packet lost occurring in the network can be handled by using best queue management heuristic. Objective function based on congestion with link quality can enhance the performance of the network which change the traffic based on congestion.

Between different technologies used at link layer and upper layers, media independent entity is defined for service access points (SAP) by IEEE 802.21 standard to use the media dependent services. In IEEE 802.21 standard, handover technique with the help of remote interfaces and architecture with network-network and terminal-network interfaces are used to transmit the information. In IEEE 802.21 framework by using protocol stack supports seamless handover between heterogeneous technologies with objective of optimization in handover decision using minimum necessary

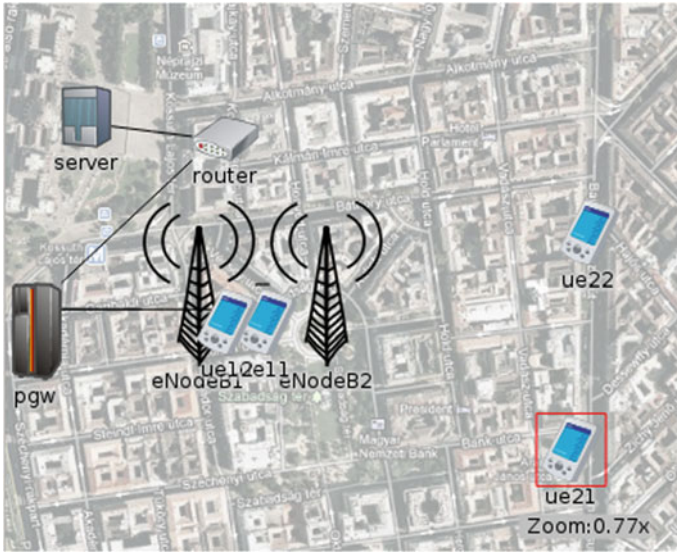


Fig. 2 Handover within HetNet at different locations

interactions among devices [10]. Link layer functions, independent of technology specifications, are provided in new link layer SAP with the help of mapping to the corresponding technology-specific primitives. Figure 2 represents the handover based on mobility at different locations.

Essential functionalities are added for enhanced handovers with a set of primitives for triggering between local and remote link layer with mobility management protocols Mobile IP in 802.21 framework. These are primary objectives of 802.21 framework while service continuity during the handover procedure without restarting is also an important objective of the framework. During handover procedure, 802.21 framework supports handover aware application which helps in decisions; for example, after a silence period, handover is executed in voice. Based on QoS criteria, framework supports the necessary functions in order to take handover decisions, e.g. handover to a new network which provides the desired QoS [11]. Information on the candidate neighbours for handover is provided to user using 802.21 feature for network discovery. Depending on throughput, QoS, billing or policies, handover decisions are made for network selection.

At node level and network level, protocol stack uses the different entities for interaction in IEEE 802.21, for example it uses media independent handover function (MIHF). Between upper and lower layers, MIHF works as intermediate layer, whose main objective is to coordinate between the different devices using different commands and functions involved in handover decisions. Mobile node query is transferred by the MIF to the information server by using four messages which provide the complete information required for handover while in the same network or in roaming through this specific geographical area. After getting the reply, it contains

information regarding a possible WLAN network interface and starts listening for beacons [12]. On receiving of beacon, IEEE 802.11 link layer defines the primitives indicating the detection of a new link which is mapped with a service access point (SoP). After receiving indication of link, handover is started with point of service (PoS) primitive. Depending on the list of resources availability and QoS requirements of user, serving PoS and other candidate PoS's result of queries are sent to MN through messages. Media independent handover (MIH) user selects the target network according to requirement and transmits a switch command which will trigger a WLAN L2 connection.

In terms of processing and communication, latency is significantly increased while maintaining the security by exchanging keys and encryption techniques. In dense network, processing speed is increased, and computational overhead is increased due to high beacon rate due to authentication, apart from congestion. 43.2% of computation time is spent on message check, which can be reduced by redesigning of checking/authentication method. Overall time can be reduced by lightweight filter security mechanism, which reduces the amount of information for checking. Group signature and hybrid pseudonym heuristics are privacy-enhancing heuristics, used to protect genuine users in unsecured systems [13]. Due to time delay of 1 ms limitation, faster computation techniques for security are used at the physical layer by system on chip (SoC) or propagation randomness is exploited in secret keys. In 5G architecture, cellular and DSRC should be tightly coupled; thus, the security protocols used can easily share and compatible with the vehicles to communicate seamlessly.

### 3 Network Simulation

Optimized link heuristic with simulation attributes is given in Table 1 using NS3 simulator. In this simulation work, we used grid size of 5 \* 5 with number of nodes equal to 25 and protocol used for traffic model is UDP. Constant position mobility model is used with RTS limit of 1500.

During the transmission of data from source to destination using  $OL_{OF}$  where IP address assigned to source is 10.1.1.1 and destination address is 10.1.1.25. During this flow information, average delay, packet loss ratio, average delay and jitter and other statistics are given in Table 2.

From Table 2, we can observe that transmitted bit rate is more than as compared to received bit rate due to packet loss. First packet is transmitted at 3.1 ns and received at 3.1127 ns, which represents that packet is transmitted in 127 ps. Simulation results are obtained using NS3 simulator to validate simulation result for different node densities, queue size with different offered load (Figs. 3, 4 and 5).

Simulation attributes and parameter values are represented in Table 1. In the simulation,  $MRH_{OF}$  is the main function which is observed with  $E_{TX}$  metrics. Simulation time is set to 25 s, and in this period, every node periodically transmits the data packet to sink node with packet length of 109 (Fig. 6).

**Table 1** Simulation attributes and their values

Simulation attributes	Values
Grid size of wireless ad hoc network	5 X 5 and 2D grid
Routing protocol	OLoF
Traffic flows	UDP
Traffic scheduling time	1, 1.5 and 2 s
Distance	500 m
Packet size	600
RTS limit	1500
IEEE 802.11	802.11b
Propagation model	Friis propagation loss model
Speed	Constant speed propagation delay model
Mobility model	Constant position mobility model
Transport layer	UDP

**Table 2** Statistics received for link heuristic using grid topology

Flow ID	10.1.1.1 → 10.1.1.25
Tx BitRate	948.226 kbps
Rx BitRate	64.9956 kbps
Mean delay	2058.04 ms
Packet loss ratio	91.9704%
delaySum	1.83166e + 12 ns
jitterSum	6.64465e + 10 ns
lastDelay	1.83166e + 12 ns
txPackets	13,023
rxPackets	890
lostPackets	10,194
timesForwarded	6230
Packet dropped no route	381
Bytes dropped no route	239,268
delayHistogramnBins	3173
jitterHistogramnBins	2603
packetSizeHistogramnBins	32

## 4 Result Analysis

In routing protocols when a control message is queued for a long duration of time, then piggybacks all the control messages in a single transmission. Multiple transmissions of TCs in short interval of time cause the value of jitter to zero and lead to collision

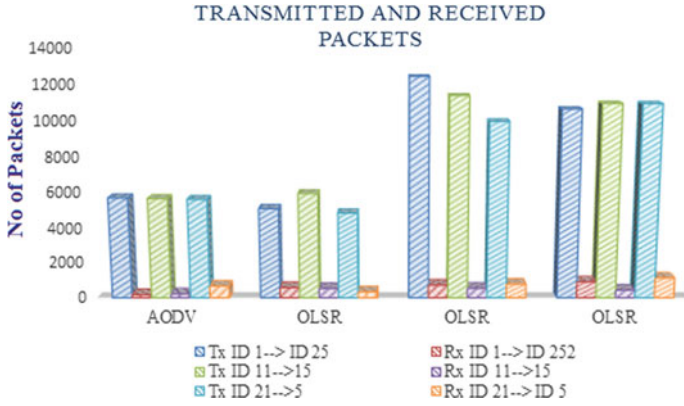


Fig. 3 Transmitted and received packets

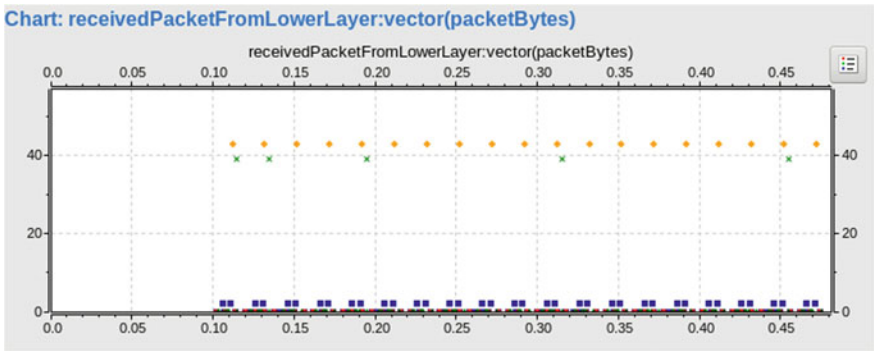
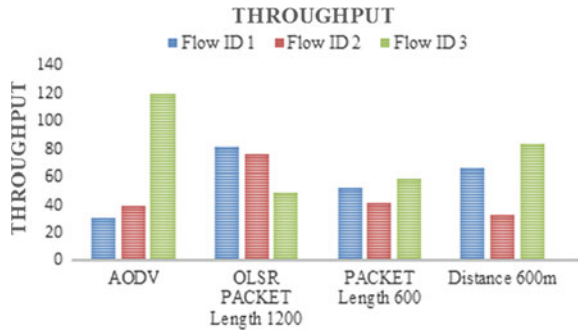
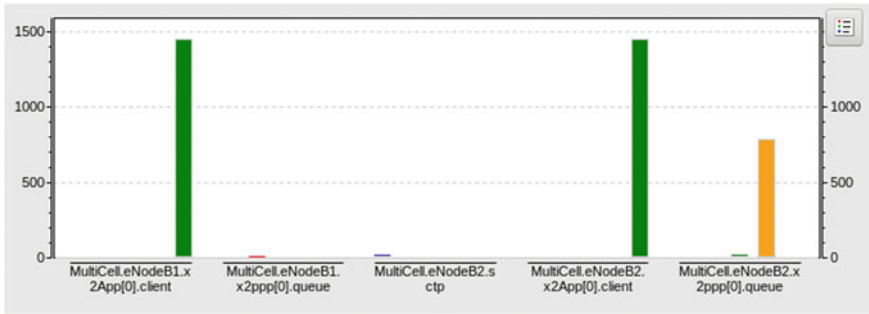


Fig. 4 Received packets from lower layer to upper layer at different time

Fig. 5 Throughput with varying packet size and distance







**Fig. 6** Packet transmission at different locations with queue performance

at the link layer. Transmission of topology control with zero or small value of jitter represents the message that is forwarded by an attached node before the message if forwarded by legitimate node. The affected node can inject the malicious content in TC causes identification spoofing results in discarding the original message as duplicate message.

## 5 Conclusion

Scalability, throughput, goodput, delay and packet delivery ratio are important parameters for performance measurement of the routing protocols for heterogenous ad hoc network. Proactive routing protocols have lower delay as compared to reactive routing protocols because path information is easily maintained at each router. Depending on scalability and transmission time, performance of OLSR, AODV, DSDV and DSR is represented in different diagrams. Variation of pause time from 0 to 100 s within a permanent topography of  $500 \times 500$  metres and node density is changed from 20 to 100. Throughput, goodput, packet delivery ratio, average E2E delay and Jitter are critical parameters in evaluating any routing protocol performance. For high mobility scenarios, OLSR is the best routing protocol in terms of all parameters. DSDV is very effective and simple for small-sized network and for moderate traffic scenario AODV is the best choice.

## References

1. Chatzikokolakis K, Kaloxylos A, Spapis P et al (2015) On the way to massive access in 5G: challenges and solutions for massive machine communications. CrownCom
2. Da Silva I, Mildh G et al (2015) Tight integration of new 5G air interface and LTE to fulfill 5G requirements. In: IEEE 81st vehicular technology conference. VTC Spring, Glasgow
3. Kovatsch M (2013) Coap for the web of things: from tiny resource-constrained devices to the web browser. In: Proceedings of the 2013 ACM conference on pervasive and ubiquitous

- computing adjunct publication, UbiComp '13 Adjunct, ACM, New York, pp 1495–1504
4. Kothmayr T, Schmitt C, Hu W, Brünig M, Carle G (2013) Dtls based security and two-way authentication for the internet of things. *Ad Hoc Netw* 11(8):2710–2723
  5. Chu H-T, Chen W-S, Huang Y-H, Chen J-Y (2004) A novel design of instant messaging service extended from short message service with XMPP. In: Fifth IEE international conference on 3G mobile communication technologies, IEEE Computer Society Press, pp 504–508
  6. Salahuddin M, Al-Fuqaha A, Guizani M (2015) Software-defined networking for rsu clouds in support of the internet of vehicles. *IEEE Internet Things J* 2(2):133–144. <https://doi.org/10.1109/JIOT.2014.2368356>
  7. Reynders B, Wang Q, Tuset-Peiro P, Vilajosana X, Pollin S (2018) Improving reliability and scalability of lorawans through lightweight scheduling. *IEEE Internet Things J* 5(3):1830–1842. <https://doi.org/10.1109/jiot.2018.2815150>
  8. Meneguetto RI, Filho GPR, Guidoni DL, Pessin G, Villas LA, Ueyama J (2016) Increasing Intelligence in inter-vehicle communications to reduce traffic congestions: experiments in urban and highway environments. *PLoS ONE* 11(8):e0159110. <https://doi.org/10.1371/journal.pone.0159110>
  9. Sethi A, Vijay S, Aeron A (2019) Secure cross layer architecture for IOT devices in NGN. *IJRTE* 8(1):2533–2537
  10. Di Caro G, Giordano S, Kulig M, Lenzarini D, Puiatti A, Schwitter F A cross-layering and autonomic approach to optimized seamless handover
  11. Melia T, Corujo D, de la Oliva A, Vidal A, Aguiar R, Soto I (2007) Impact of heterogeneous network controlled handovers on multi-node mobile device design
  12. Choi S, Hwang G, Kwon T, Lim A, Cho D (2005) Fast handover scheme for real time downlink services in IEEE 802.16e BWA system
  13. Lopez Y, Robert E (2010) OpenMIH, an open-source media-independent handover implementation and its application to proactive pre-authentication. In: Pentikousis K, Blume O, Agüero Calvo R, Papavassiliou S (eds) *Mobile networks and management. MONAMI 2009. Lecture notes of the institute for computer sciences, social informatics and telecommunications engineering*, vol 32. Springer, Berlin, Heidelberg

# Remote Monitoring of Vital Parameters with IoT-Based Sensing System



Rohini, M. A. Ansari and Nidhi Singh Pal

**Abstract** Being the major influential aspect within the lives of the people, health has become an increasing cause of consciousness and concern today. In this paper, a healthcare system has been set up and designed which enables patients to collect daily, the vital parameters at home and sending them over the cloud with the use of IoT i.e. Internet of things. Identification of a set of five parameters—weight, temperature, heart rate, pulse oximetry, and attentiveness of alcohol in breathe, is done by using biomedical sensors which are interfaced with Arduino Uno microcontroller that transmits the recorded data through ESP8266 Wi-Fi module on IoT platform i.e. ThingSpeak. The vital parameters can be visualized and monitored on devices that include desktops, laptops, or smart phones which are connected under similar networks.

**Keywords** Remote monitoring · Arduino · ESP 8266 · IoT · ThingSpeak

## 1 Introduction

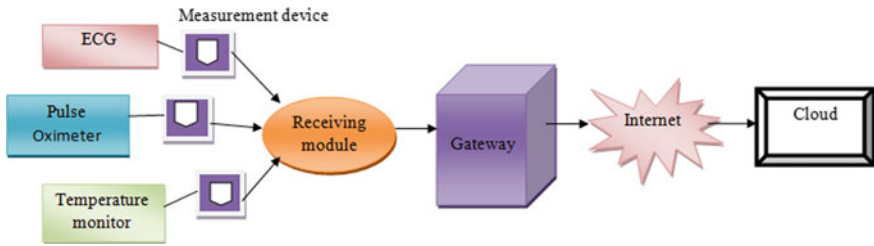
With growing time, there is a gradual decline in the immunity of the body that leads to many diseases and health issues in people. Continuous monitoring of the patient is required at both hospital and home [1]. In the past number of years, a massive improvement in the methods of diagnosis, health monitoring and alerting the patients as well as medical professionals with improved accuracy has taken place [2]. At present, most of the healthcare systems are based on hospitals and require the patient to do periodic visits in order to get the vital parameters checked that affect their health. This in turn has become a difficult and exhaustive job for the patients

---

Rohini (✉) · M. A. Ansari · N. S. Pal  
Gautam Buddha University, Greater Noida, India  
e-mail: [rohinishingh0710@gmail.com](mailto:rohinishingh0710@gmail.com)

M. A. Ansari  
e-mail: [ma.ansari@ieee.org](mailto:ma.ansari@ieee.org)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_5](https://doi.org/10.1007/978-981-15-2329-8_5)



**Fig. 1** Simple telehomecare system

living far away from the place where the hospitals or clinics with the required equipments are located. Telemedicine has emerged as the newest technology that combines information technology and telecommunication for medical purposes [3, 4].

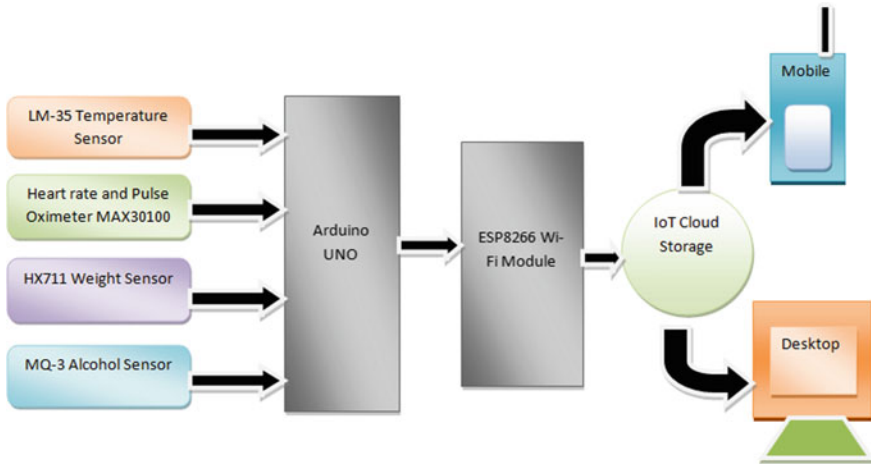
## 2 Telemedicine

Telemedicine includes the exchange of medical information between two sites via electronic communications in order to improve patient's health status [4]. It provides a better and new improved way to deliver the much-needed healthcare service efficiently to the patients when the distance between the doctor and patient is significantly large. A simple telehomecare system that includes integrated measurement devices and receiving module is shown in Fig. 1 [5]. The server of the telehomecare system is responsible for receiving data uploaded by measurement devices, reprocessing the data, and saving the data to the corresponding database location [6].

In this paper, a remote healthcare system has been designed for the patient to use at home that helps to monitor and measure all the vital parameters that need to be checked at regular intervals for maintaining the health status and lifestyle quality [7].

## 3 Internet of Things

IoT can be referred to the network consisting of appliances and devices inclusive of everything from electronics to actuators allowing connectivity, interaction, and exchange of data between them. The extension of Internet connectivity between standard devices to even non-Internet enabled devices is facilitated by IoT [8, 9]. These technology-embedded devices can be independently controlled and monitored remotely. IoT has evolved as an ideal technology influencing the Internet as well as communication technologies. With the efficient use of IoT, people can connect with anything and at anytime by using any service, path, or network. Smart health care plays a very eminent role in medical applications through embedding the sensors and actuators for the purpose of monitoring and keeping track of patient's health status



**Fig. 2** Block diagram of the IoT-based healthcare system

[10]. For monitoring the physiological status of the patient under consideration, the clinical care units use IoT for the collection and analysis of concerned information through various sensors.

## 4 Methodology and Discussion

This work is an integrated healthcare system that allows the patients to daily monitor and collect vital signs sitting at home and sending them using IoT. Heart rate, weight, temperature, saturation of oxygen in blood, and attentiveness of alcohol in breathe are the five parameters that are measured by using biomedical sensors. These sensors are connected to Arduino Uno board that consists of AVR microcontroller—Atmel (8-bit). Figure 2 shows the block diagram of IoT-based healthcare system. All the sensors are programmed using the Arduino that stores the potential data measured by the sensors and transmits it for further processing or visualization. ESP 8266 is the Wi-Fi module that acts as a server and sends data wirelessly on the IOT platform—ThingSpeak [11].

## 5 Software and Hardware Requirements

The software and hardware requirements for the healthcare system include the biomedical sensors and other equipments. Table 1 shows the list of components and their respective ratings [12].

**Table 1** Ratings of circuit components of the IoT-based smart healthcare system

S. No.	Component	Ratings
1	Arduino Uno	14 digital input/output, 6 analog inputs, a 16 MHz quartz crystal, on board. 3.3–A 3.3 V
2	ESP 8266	32-bit microcontroller, 80 MHz, 16 GPIO pins, input voltage—3.3–3.6 V
3	MQ-3 alcohol gas sensor	Operating voltage: 5 V $\pm$ 0.1, current: 150 mA, operating temperature: 10–70 °C
4	Weight sensor (HX711)	24-bit (ADC), normal operation <1.5 mA, power down <1 $\mu$ A, supply voltage range: 2.6–5.5 V
5	Temperature sensor (LM35)	Input voltage—2.7–5.5 V, linear at 10 mV/°C
6	Pulse oximeter and heart rate sensor IC MAX30100	VDD = 1.8 V, VIR_LED+ = VR_LED+ = 3.3 V, $T_A$ = 25 °C
7	LCD display JHD 162A	8-bit data pins, supply voltage—5 V (4.7–5.3 V)

## 6 Description of System Components

### 6.1 IoT Webpage

‘ThingSpeak’ is one of the many IoT platforms that offer a wide variety of counteraction capabilities for analyzing and monitoring the data. It can be referred to a channel that stores and processes the data collected from concerned devices. It also provides a number of useful applications to facilitate integration with different Web services and social network. Figure 3 shows the ThingSpeak channel that has been created for the storage of recorded sensor data in different fields that defined each sensor.

### 6.2 ESP8266 Wi-Fi Module

The Wi-Fi Module ESP8266 gives access to any microcontroller desired Wi-Fi network. Figure 4 shows the ESP8266 Wi-Fi Module. This Wi-Fi module is integrated with TCP/IP protocol stack and is a SOC (self-contained). It is a very cost-effective board and is pre-programmed with AT command set. This makes it very easy to connect with the Arduino directly [13].

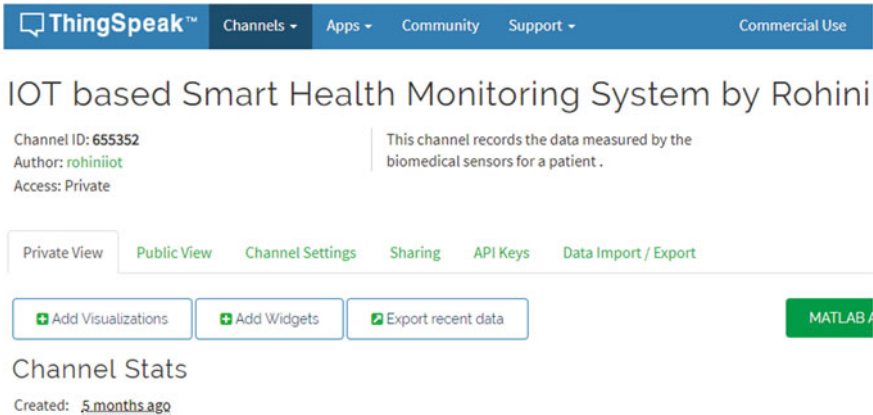
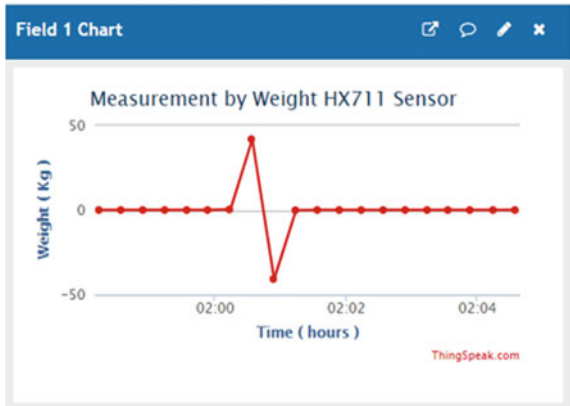


Fig. 3 ThingSpeak channel for storage of sensor data

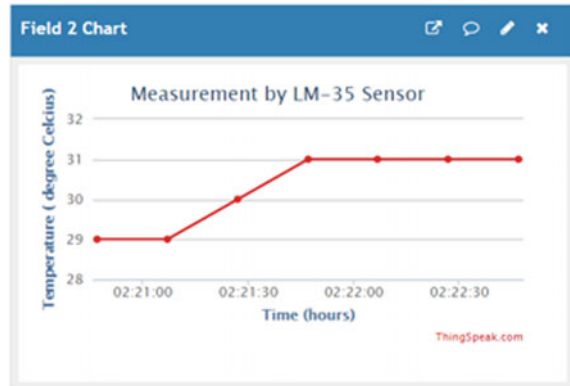
Fig. 4 Measurement by weight HX711 sensor for 41.93 kg



### 6.3 Arduino Uno Board

The system works with Arduino Uno microcontroller board. Arduino has total of 14 pins that are used as input/output pins. Six pins are used for PWM outputs and six as analog inputs. It also has a power jack, a reset button, an ICSP header, a USB connection, 16 MHz quartz crystal. USB or external power supply can be used to power the Arduino. Figure 5 shows the Arduino Uno board used in the system.

**Fig. 5** Measurement by LM-35 temperature sensor for 31 °C



## 6.4 HX711 Weight Sensor

HX711 is an ADC of 24-bit based on patented technology by Avia Semiconductor. It has been designed to work as weigh scales and to be interfaced with a bridge sensor directly in industrial control applications. Load cell—a type of transducer used for creating an electrical signal having its magnitude in direct proportion to the measured force. There are four strain gauges in a load cell in Wheatstone bridge configuration. HX711 acts as an instrumentation amplifier that amplifies the electrical signal output which is in order of millivolts (mV).

## 6.5 LM 35 Temperature Sensor

The temperature sensor LM35 is high precision-integrated circuit sensor. The voltage output of the LM35 temperature sensor has linear proportionality to the degree Celsius temperature. Because of this characteristic, LM35 sensor is advantageous in comparison with linear temperature sensors which are calibrated in ° Kelvin. Thus, the need for subtracting a large value of output voltage of sensor to obtain temperature on Centigrade scale is avoided. It possesses low self-heating capability less than 0.1 °C in air, and only 60  $\mu$ A current is drawn by it from the supply. The operation range for this sensor is  $-55$  to  $+150$  °C.

## 6.6 Alcohol Sensor (MQ-3)

MQ-3 alcohol sensor has a perspective layer of tin oxide inside micro-tubes of aluminum oxide and a heating element that is inside a casing which is tubular in shape. The ethyl alcohol present in the breadth is oxidized into acetic acid and that passes to



the heat element. As the sensing tin oxide layer is cascaded with ethyl alcohol, there is a decrement in the resistance. The resistance variation conversion into suitable voltage variation can be achieved with the use of external load resistance [14].

## 6.7 Heart Rate and Pulse Oximetry Sensor IC

The MAX30100 is an ultra-low power operating-integrated heart rate and pulse oximetry sensor, and it comes in a tiny 5.6 mm × 2.8 mm × 1.2 mm 14-pin optically enhanced system in package. It is a combination of two LEDs, a photodetector, analog low-noise signal processing and optimized optics. The operating range of power supply for MAX30100 is 1.8–3.3 V. It can be powered using a software.

## 7 Results and Discussion

### (a) Readings of HX711 Weight Sensor (Weight measured = 41.93 kg)

The weight HX711 sensor measures up to 100 kg. The readings are shown on ThingSpeak.com. For patient 1 (age = 15 years), Fig. 4 shows the reading on Field 1 which is equal to 41.93 kg. The *x*-axis shows the time of the day at which the reading is taken, and *y*-axis shows the value of weight measured in Kg.

### (b) Readings of LM-35 Sensor (Measured Temperature = 31 °C)

The LM-35 temperature sensor measures the temperature in degree Celsius. For patient 1 (age = 13 years), Fig. 5 shows the reading on Field 2 which is equal to 31 °C. The *x*-axis shows the time of the day at which the reading is taken, and *y*-axis shows the value of temperature measured in degree Celsius.

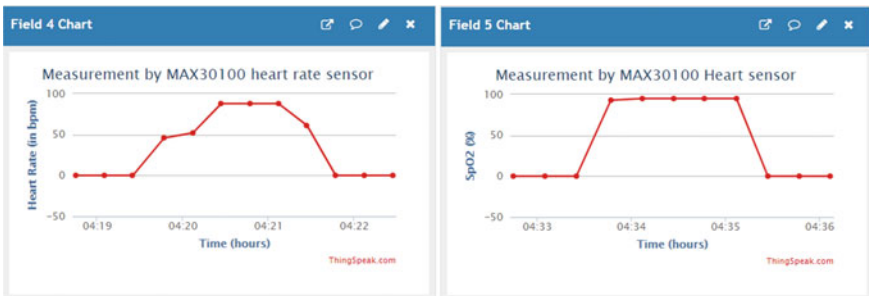
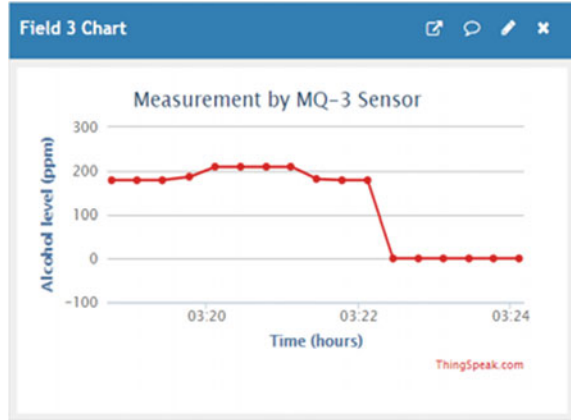
### (c) Readings of MQ-3 Alcohol Sensor (Alcohol Level = 209 ppm)

The MQ-3 alcohol sensor measures the level of alcohol in ppm. The readings are shown on ThingSpeak.com. For patient 1 (age = 18 years), Fig. 6 shows the reading on Field 3 which is equal to 209 ppm. The *x*-axis shows the time of the day at which the reading is taken, and *y*-axis shows the value of alcohol attentiveness in breadth measured in ppm.

### (d) Readings of Heart Rate Sensor (a. Heart rate = 88 bpm b. SpO2 = 95%)

The heart rate and pulse oximeter IC MAX30100 sensor measure the heart rate in bpm and saturation of oxygen in blood in percentage. The readings are shown on ThingSpeak.com. For patient 1 (age = 13 years), Fig. 7 shows the reading on Field 4 for heart rate which is equal to 88 bpm and measurement of saturation of oxygen in blood for patient 1 on Field 5 which comes out to be 95%. The *x*-axis shows the

**Fig. 6** Measurement by MQ-3 alcohol sensor for 209 ppm



**Fig. 7** Measurement by heart rate sensor, i.e., 88 bpm; SpO2 = 95%

time of the day at which the reading is taken, and y-axis shows the heart rate in bpm and % SpO2.

In the reflective method there is some fixed light reflected back to the sensor due to finger. With each heartbeat, there is sudden increase in blood volume in the finger and this results in more light reflection back to the sensor.

## 8 Conclusion

A wireless system for health monitoring with the use of sensors and mobile devices can be implemented in a global network with the efficient use of Arduino and ESP8266. The devices and IoT gather and exchange information facilitating the collection, analysis, and monitoring of data with more accuracy. Thus, IoT can be used for monitoring the patient and providing services in time. The proposed healthcare system can be modified, enhanced, and extended by using other invasive as well as noninvasive sensors for picking up essential medical potentials of a patient. This can

be analyzed further, stored, and transferred on a global platform. More biomedical sensors can also be added to enhance the functionality of the model.

**Acknowledgements** In regard to the clarification sought on the captioned matter, it is to convey that no patient was directly enrolled for this work. However, the results of various parameters as depicted in the paper pertain to the close group members of author only and with their permission. The objective of this work is not to diagnose any illness or health issues of the person but to demonstrate how the vital parameters can be measured, recorded, and transmitted with the help of designed Wi-Fi-embedded IoT-based system on the cloud.

## References

1. Uplenchwar K, Vedalankar A (2017) IoT based health monitoring system using raspberry pi and Arduino. *Int J Innovative Res Comput Commun Eng* 5(12)
2. Kajornkasirat S, Chanapai N, Hnusuwan B (2018) Smart health monitoring system with IoT. In: 2018 IEEE symposium on computer applications & industrial electronics (ISCAIE). Penang, pp 206–211
3. Matlani P, Londhe ND (2013) A cloud computing based telemedicine service. In: 2013 IEEE point-of-care healthcare technologies (PHT). Bangalore, pp 326–330
4. Le Glédic S, Fournier M, Malek S, Dupraz-Poiseau A (2011) Telemedicine and e-health systems. In: The 14th international symposium on wireless personal multimedia communications (WPMC). Brest, pp 1–5
5. Ji H, Wang J, Gao J, Liu X (2016) Research on telemedicine technology and implement based on virtual reality. In: 2016 IEEE advanced information management, communicates, electronic and automation control conference (IMCEC). Xi'an, pp 1581–1586
6. Wang CS (2013) The implementation of a tele-homecare system with service oriented architecture and HL7 message transmission standard. *Am J Public Health Res* 1(1):18–26
7. D'Souza M, Ros M, Postula A (2006) Wireless medical information system network for patient ECG monitoring. In: Digital system design: architectures, methods and tools, DSD 2006, 9th EUROMICRO conference, pp 617–624
8. Kumar R, Rajasekaran MP (2016) An IoT based patient monitoring system using raspberry pi. In: 2016 International conference on computing technologies and intelligent data engineering (ICCTIDE'16). Kovilpatti, pp 1–4
9. Biswas AR, Giaffreda R (2014) IoT and cloud convergence: opportunities and challenges. In: 2014 IEEE world forum on internet of things (WF-IoT). Seoul, pp 375–376
10. Benazzouz Y, Munilla C, Günalp O, Gallissot M, Gürgen L (2014) Sharing user IoT devices in the cloud. In: 2014 IEEE world forum on internet of things (WF-IoT). Seoul, pp 373–374
11. Kale S, Mane S, Patil P (2017) IOT based wearable biomedical monitoring system. In: 2017 International conference on trends in electronics and informatics (ICEI). Tirunelveli, pp 971–979
12. Talukdar G, Ansari MA, Sarkar A, Samvat K, Kumar K (2010) GSM based remote patient monitoring system. In: Proceedings of the international conference on biomedical engineering and assistive technologies (BEATS'10)
13. Banik A, Anand RS, Ansari MA (2008) Remote monitoring and analysis of human lung sound. In: Proceedings of the 3rd IEEE international conference on industrial and information systems (ICIIS'08), pp 1–6, 8–10 Dec 2008
14. Sarathi MP, Ansari MA (2015) A scheme for patient study retrieval from 3D brain MR volumes. In: Proceedings of the IEEE international conference on green computing and internet of things (ICGCIoT'2015), pp 214–218, 8–10 2015

# Economic Load Dispatch Using PSO



Satyam Tiwari, Nidhi Singh Pal, M. A. Ansari, Dilip Yadav  
and Nivedita Singh

**Abstract** Nowadays, electrical energy is playing vital role in human life. The economic operation of the power system is always desirable and this can be achieved by economic load dispatch. The economic load dispatch means allocating the power to different generating units to minimize or reduce the total fuel cost satisfying the different power system constraints. The main purpose of economic load dispatch (ELD) is to allocate the total power at generating units to meet total load demand with minimum operational cost fulfilling all operational obstacles (constraints). For optimal dispatch of power, different evolutionary optimization techniques are being used like CSO, PSO, etc. In this paper, PSO has been implemented on MATLAB to minimize the total cost for 3-unit and 6-unit systems and the results are compared with CSO and other evolutionary optimization techniques for different power demands.

**Keywords** Particle swarm optimization (PSO) · Operational obstacles (constraints) · Transmission losses

## 1 Introduction

To provide electricity at lowest price in power system, economic scheduling of generating units is very significant. With increase in cost of fuel, generation of power is becoming expensive. So, this is a major concern of modern system to reduce the cost of generation as well as transmission and distribution while satisfying the electrical constraints. Economic load dispatch gives allocation of the total demand of the system among the various units in service (on state) so that total cost of generation is minimum [1]. Different optimization techniques are there to optimize the given problem and to find the least cost of generation. Nowadays, Evolutionary techniques are being used for better result and less time consumption [2]. Practically, ELD problem is non-linear-type problem due to inclusion of equality and inequality

---

S. Tiwari (✉) · N. S. Pal · M. A. Ansari · D. Yadav · N. Singh  
Department of Electrical Engineering, Gautam Buddha University, Greater Noida, India  
e-mail: [tsatyam740@gmail.com](mailto:tsatyam740@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_6](https://doi.org/10.1007/978-981-15-2329-8_6)

constrains. Many evolutionary search techniques like modified Hopfield network, simulated annealing (SA) [3], genetic algorithm (GA) [4], and tabu search algorithm (TSA) [5] are being used to solve non-linear complex ELD problem efficiently. Particle swarm optimization (PSO) is an evolutionary computational technique which is based in swarm intelligence and inspired by social behavior of creatures (Bird flock or fish) in nature to optimize a wide range of continuous linear and non-linear problems [6].

## 2 Mathematical Modeling

The main goal of economic dispatch problem is that to maintain power at generating unit in such a way that the total cost of generation at generating units is minimum [7]. For  $n$  number of generating unit, cost function can be given as

$$F_i(P_{gi}) = a_i P_{gi}^2 + b_i P_{gi} + c_i \quad (1)$$

where  $P_{gi}$  is termed as real power generation for  $i$ th unit, (a) (\$/MW<sup>2</sup>), (b) (\$/MW), and (c) (\$) are fuel cost coefficients. For calculation of total fuel cost of the generation, Eq. (2) can be used as:

$$F_T(P_{gi}) = \sum_{i=1}^N F_i P_{gi} \quad (2)$$

where;  $F_T(P_{gi})$  is the total fuel cost (\$/h),  $F_i(P_{gi})$  is cost of fuel of  $i$ th unit, and  $N$  is the total number of generating unit. The ELD problem has objective function as:

$$\text{Minimize } F_T(P_{gi}) \quad (3)$$

where,  $F_T(P_{gi})$  is the total fuel cost subject to different constraints.

1. **Equality constraints**—The power balance equation is given by [8]

$$\sum_{i=1}^N P_{gi} P_{gi} = P_D + P_L \quad (4)$$

$P_L$  is stated as total transmission loss and can be given by

$$P_L = \sum_{k=1}^N \sum_{i=1}^N P_{gi} B_{ik} P_{gk} \quad (5)$$

where'  $B_{ik}$  is called transmission loss coefficient,  $P_{gk}$  is the power generation at  $i$ th unit, and  $P_D$  is total active power demanded.

2. **Inequality constraints**—Inequality constraints may be classified as

- (a) **Voltage constraints**—Phase angle and voltages values at different nodes should be in their limits. If limit violates, the performance of generating unit will be affected [9].

$$V_{i\min} < V < V_{i\max} \text{ and } \delta_{\min} \leq \delta \leq \delta_{\max}$$

- (b) **Transformer tap setting**—The transformer tap setting lies in between 0 and 1.

$$0 \leq t \leq 1$$

For secondary side, the relation of tap setting can be given as

$0 \leq t \leq n$  where  $n$  is turn ratio.

- (c) **Generator constraints**—In this, the parameter of active power and reactive power must be in their stipulated limit. If thermal factor is considered, then generation of active power is limited to its maximum and minimum limits [10].

$$P_{i\min} \leq P_i \leq P_{i\max}$$

In the same way, heating of winding inhibits the maximum reactive power and stability inhibits the minimum reactive power.

$$Q_{i\min} \leq Q_i \leq Q_{i\max}$$

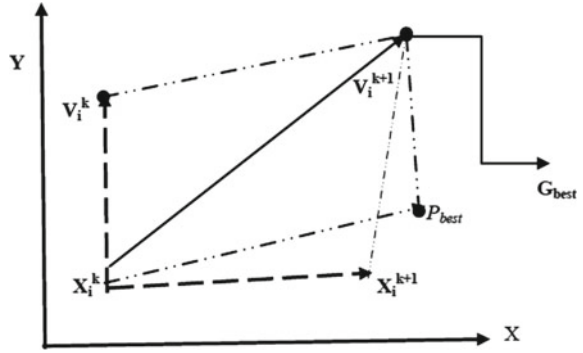
- (d) **Ramp rate limit**—Output power of generator cannot change instantly. This limit tells about how power varies with time.

### 3 Evolutionary Optimization Techniques

Evolutionary techniques are better than a conventional technique because it requires less computational time and gives more accurate results. Some of evolutionary techniques are discussed as below [11]:

- a. **Cuckoo search optimization**—CSO is inspired by the hatch parasitic tendencies of new types of cuckoo, which gives its eggs in the host bird's nests.
- b. **Genetic algorithm**—GA is a search optimization technique that is used to solve sophisticated optimization problem with a large number of parameters whose analytical solution is difficult to obtain.

**Fig. 1** Modification of velocity and position in PSO [8]



- c. **Particle swarm optimization**—It is one of the best evolutionary optimization techniques. PSO was actuated by social conduct (behavior) of biological organism. [12] (Fig. 1).

## 4 PSO Methodology

In the simulation of PSO, the population number is denoted by  $N$  in the ELD problem. All particles move with specified velocity  $V_i$  and their position can be represented by  $X_i$ . The updating position and velocity of the particles can be shown in the following equation [13]

$$V_i^{k+1} = wV_{ik} + C1 \times \text{rand}()_1 \times (P_{best_i} - X_{ik}) + C2 \times \text{rand}()_2 \times (G_{best_i} - X_{ik}) \quad (6)$$

$$X_i^{k+1} = X_i^k + V_i^k \quad (7)$$

where  $V_i^{k+1}$  = velocity of particle  $i$  after modification,  $V_i^k$  = current velocity of particle,  $X_i^{k+1}$  = position of particle after modification,  $X_i^k$  = current position of particle, and  $\text{rand}()$  = random numbers which lies in between 0 and 1. The term  $\text{rand}()_1 \times (P_{best_i} - X_{ik})$  is called memory influence of particle and the term  $\text{rand}()_2 \times (G_{best_i} - X_{ik})$  is swarm influence.  $V_i^k$  should be in between  $V_{\min}$  and  $V_{\max}$ . By  $V_{\max}$ , the fitness value can be calculated. If value of  $V_{\max}$  will be excessive high, then particle may fly previous good solution. If  $V_{\min}$  will be too low, then particle may not investigate sufficiently beyond local solution.  $C_1$ ,  $C_2$  is the acceleration constant.  $w$  is termed as the inertia weight constant.

## 5 PSO Algorithm

Step by step algorithm of PSO [2] for the minimum cost operation can be given as

- Step 1 Specify min. and max. power limit at each generating unit. Initialize the cost function which is total cost function of different generating units.
- Step 2 Initialize all the parameters of PSO, i.e.  $C_1$ ,  $C_2$ ,  $W_{\max}$ , and  $W_{\min}$  including position and velocity and  $\text{iter} = 1$ .
- Step 3 Input the cost function, total power demand, and loss coefficient matrix.
- Step 4 Evaluate cost function while substituting the initial position of each particle and in the population, any particle which possesses best value will be called as  $P_{\text{best}}$ .
- Step 5 Compare the each individual's value of each evaluation with its  $P_{\text{best}}$ . The best value among the  $P_{\text{best}}$  is denoted by  $G_{\text{best}}$ .
- Step 6 Update the velocity of each individual particles according to Eq. (6).
- Step 7 Update each particle positions according to Eq. (7).
- Step 8 At each stage, the error gradient is checked and the value of  $G_{\text{best}}$  is plotted, i.e. convergence curve  $3Z\ 0 + 2581$  till it comes with in the minimum-specified range.
- Step 9 This value of the  $G_{\text{best}}$  will be the best optimal solution. The value of  $G_{\text{best}}$  will be the best cost.

The objective function for the total production cost can be written in the equation given below [4]

$$F_i(P_{gi})_{\min} = a_i P_{gi}^2 + b_i P_{gi} + c_i + |d_i * \sin(e_i (P_{ig}^{\min} - P_g))| \quad (8)$$

where  $e_i$  and  $d_i$  are cost -coefficient of valve point loading effect. For accurate-modeling of fuel cost function, the effect of valve point has been considered. The fix value of  $e_i$  and  $d_i$  is taken as in standard format. The main aim of this paper is to get the economically less fuel consumption  $F_T(P_{gi})$  considering the parameters like fuel cost coefficient, loss coefficient, and coefficient of valve point loading effect.

### Case-1: 3-Unit System

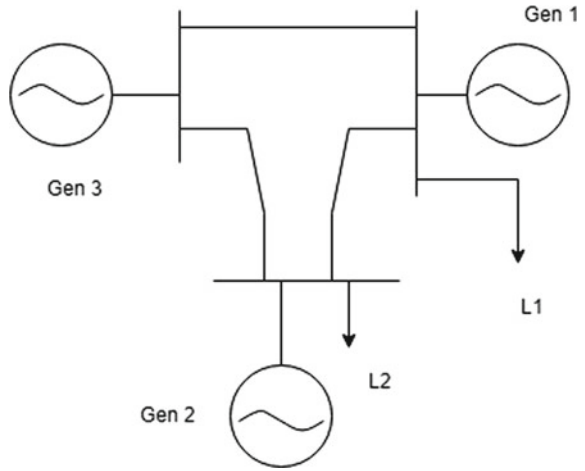
In Fig. 2, the 3-generators are connected at 3-buses and load L1 and L2 are having the power demand of  $P_D$ . In ELD problem, the total power demanded must be distributed to the generators so that objective of minimum cost achieved. The minimum and maximum power limits and cost coefficient associated with above 3-unit system can be tabulated as in Table 1.

PSO parameters are tabulated in Table 2 and the loss coefficient associated with 3-unit test system is given as follows:

$$B_{ik} = \begin{bmatrix} 0.000218 & 0.000093 & 0.00002 & 0.000093 & 0.000218 \\ 0.000017 & 0.000028 & 0.00001 & 0.000179 & \end{bmatrix}$$



**Fig. 2** Three-unit 3-bus system

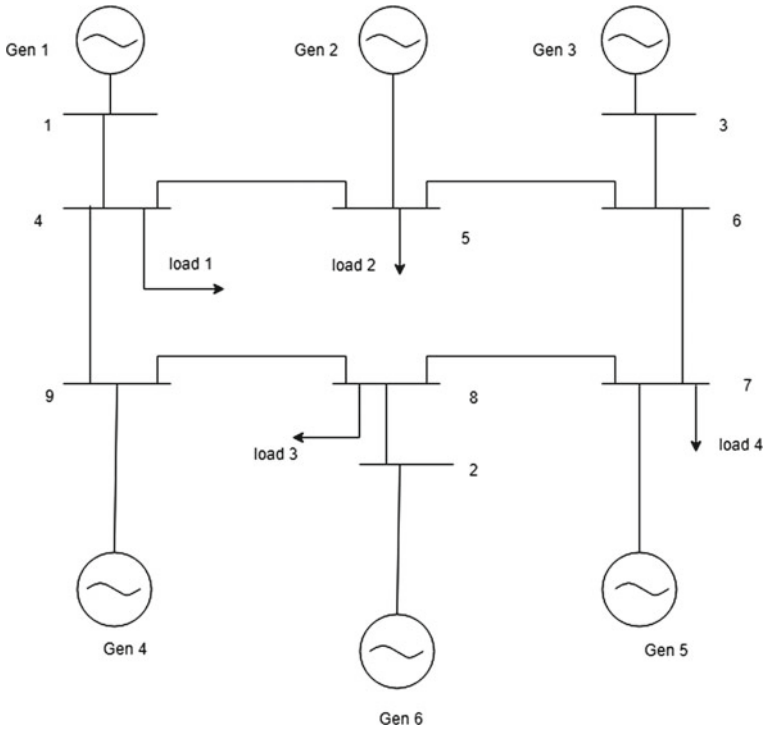


**Table 1** Fuel cost coefficient of 3-unit system

Unit no.	$P_i$ (MW)	$P_j$ (MW)	$a_i$	$b_i$	$c_i$
<i>3-bus system</i>					
1	40	80	200	7	0.008
2	60	120	180	6.3	0.009
3	80	190	140	6.8	0.007
<i>6-bus system</i>					
125	756.79886	38.53	0.15240		
150	451.32513	46.15916	0.10587		
215	1049.9977	40.39655	0.02803		
225	1242.5311	38.30443	0.3546		
325	1658.5696	36.32782	0.2111		
315	1356.6592	28.27041	0.1799		

**Table 2** Parameters of PSO

Parameter	Values
$C_1$	2
$C_2$	2
$W_{min}$	0.9
$W_{max}$	0.4
Iter	80



**Fig. 3** Six-unit 9-bus system

PSO parameters are tabulated in Table 2. The value of  $C_1$  is 8765 and  $C_2$  is 36780.  $C_1$ ,  $C_2$  and weight parameters are taken from reference paper [4].

### **Case-2: 6-Unit System**

In Fig. 3, a typical 9-bus six generator system is shown. For the load demand of  $P_D$  generator has to supply the power according to the power balance Eq. (4). The minimum and maximum power limits and cost coefficient associated with above 6-unit system can be tabulated as in Table 1. PSO parameters are same as case (1) and tabulated in Table 2. The loss coefficient associated with 6-unit test system is as following

$$B_{ik} = \begin{bmatrix} 0.00014 & 0.00017 & 0.000150 & 0.00019 & 0.00026 & 0.00022 \\ 0.00017 & 0.00060 & 0.000130 & 0.00016 & 0.00015 & 0.00020 \\ 0.00015 & 0.00013 & 0.000650 & 0.00017 & 0.00024 & 0.00019 \\ 0.00019 & 0.00016 & 0.000170 & 0.00072 & 0.00030 & 0.00025 \\ 0.00026 & 0.00015 & 0.000240 & 0.00030 & 0.00069 & 0.00032 \\ 0.00022 & 0.00020 & 0.000190 & 0.00025 & 0.00032 & 0.00085 \end{bmatrix}.$$

## 6 Results and Discussion

The PSO algorithm has been implemented on MATLAB while taking the PSO parameters as shown in Table 2. The result of  $G_{best}$  means best optimized cost has been calculated for both 3-unit and 6-unit system for different load demands. The result of PSO has been then compared with other evolutionary methods.

### Case 1: 3-Unit System

Taking the  $B_{ik}$  and data of fuel cost coefficient from Table 1 for 3-unit into consideration, the result of 3-unit system for the load demand of 150, 180, and 200 MW the minimum cost of generation is calculated and compared with other methods for the detailed analysis of PSO. For the different demand, minimum cost of operation can be obtained as in Table 3. Comparison of results for 3-unit cost for the load demand of 150 MW is tabulated in Table 3.

It can be observed that from Table 3 that for 150 MW power, the value of cost obtained from PSO gives the minimum fuel cost in comparison to other methods. Similarly, for 180 and 200 MW fuel cost obtained from PSO method is minimum as compared to CSO and GA methods as shown in Tables 4 and 5, respectively. Comparison results for 3-unit cost for the load demand of 180 and 200 MW are tabulated in Tables 4 and 5, respectively (Table 6).

The convergence of cost curve of PSO for case-1 is shown in Figs. 4, 5, and 6. It can be observed from the above figures that PSO has fast convergence characteristics.

### Case 2: 6-Unit System

Taking the  $B_{ik}$  for 6-unit into consideration, the result of 6-unit system for the load demand of 700, 800, and 900 MW, the minimum cost of generation is calculated and

**Table 3** PSO results for different load demands

N	Load (MW)	$P_1$ (MW)	$P_2$ (MW)	$P_3$ (MW)	Cost (\$/h)
1	150	62.410	50.323	36	1584.6
2	180	28.910	80.000	70	1799.10
3	200	80.513	44.884	71	1851.2

**Table 4** Comparison of PSO with other method

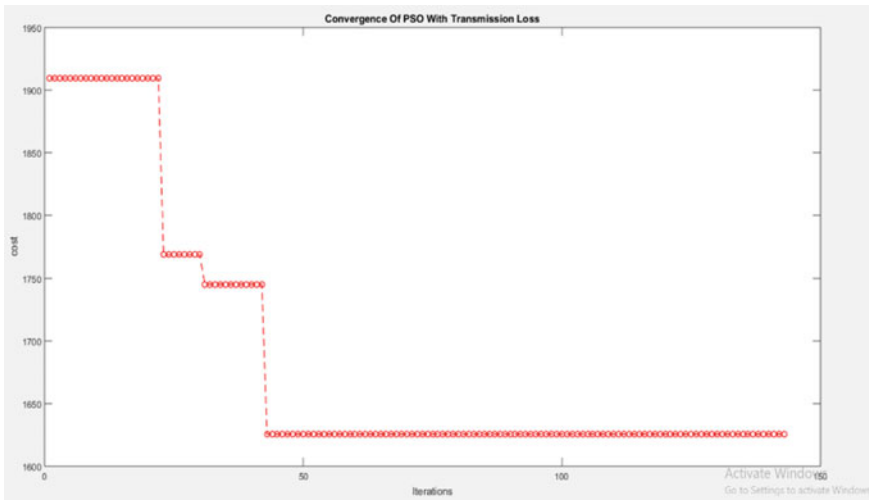
Load (MW) (150 MW)			
Load (MW)	PSO	CSO	GA
$P_1$	62.41	32.70	30.01
$P_2$	50.32	63.71	65.95
$P_3$	36.00	55.90	56.53
$P_{loss}$	1.217	2.391	2.506
Total fuel cost (\$/h)	1584.6	1597.77	1628.75

**Table 5** Comparison of PSO with other method (180 MW)

Load (MW)	PSO	CSO	GA
$P_1$	28.91	46.33	39.04
$P_2$	80	71.37	75.76
$P_3$	70	65.64	70.06
$P_{loss}$	1.10	3.582	3.65
Total fuel cost (\$/h)	1799.10	1833.54	1917.84

**Table 6** Comparison of PSO with other method (for 200 MW)

Load (MW)	PSO	CSO	GA
$P_1$	80.51	54.20	54.76
$P_2$	44.88	79.98	80
$P_3$	71.43	69.99	70
$P_{loss}$	3.17	4.604	4.64
Total fuel cost (\$/h)	1851.2	1946.33	2024.6



**Fig. 4** Convergence of 3-unit system (150 MW)

compared with other methods for the detailed analysis of PSO. It can be observed that from Table 7 that for 700 MW power, the value of cost obtained from PSO gives the minimum fuel cost in comparison to other methods. Comparison results for 3-unit cost for the load demand of 700 MW, 800 MW, and 900 MW is tabulated in Tables 7, 8 and 9, respectively, as shown below.

The convergence of cost curve of PSO for case-1 is shown in Figs. 7, 8 and 9. It can be observed that from the above figures that PSO has fast convergence characteristics.

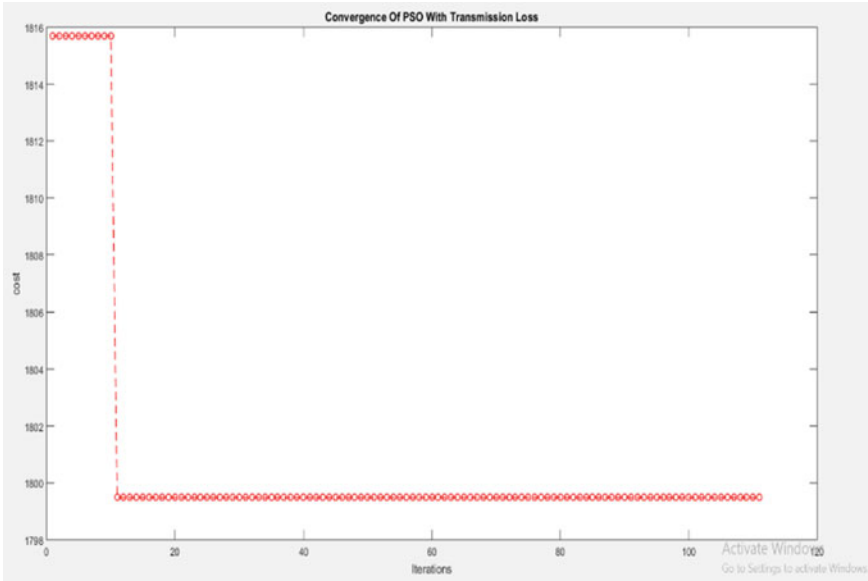


Fig. 5 Convergence of 3-unit system (180 MW)

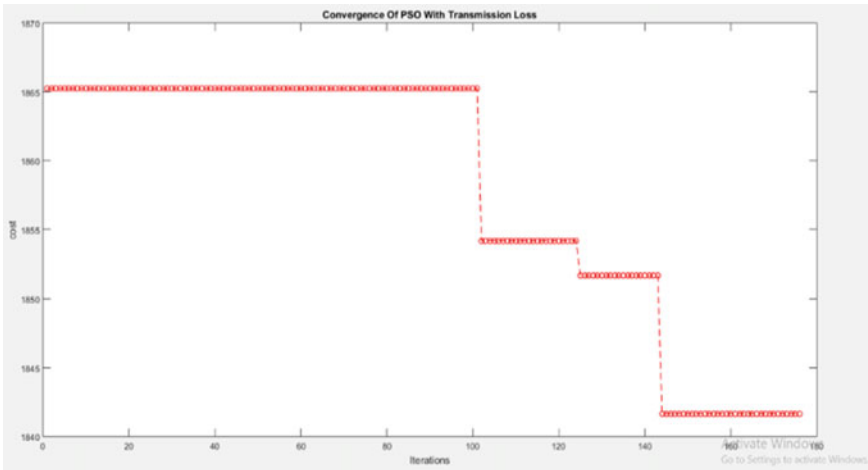


Fig. 6 Convergence of 3-unit system (200 MW)

## 7 Conclusion

In this paper, the economic load dispatch using evolutionary techniques has been implemented. The results of PSO have been compared with other evolutionary techniques and it is found that the cost obtained from PSO is much lesser than the other

**Table 7** Comparison of PSO with other method (for 700 MW)

Load	PSO	CSO	GA
$P_1$	15.462	36.801	17.450
$P_2$	40.236	10.0	20.087
$P_3$	153.573	129.56	108.29
$P_4$	155.135	120.209	132.40
$P_5$	203.200	209.989	225.683
$P_6$	124.236	212.294	216.931
$P_{loss}$	8.15	18.01	20.796
Cost (\$/h)	36557.5	36930.7	37069.8

**Table 8** Comparison of PSO with other method (for 800 MW)

Load	PSO	CSO	GA
$P_1$	21.531	25.173	23.890
$P_2$	124.241	33.496	10.249
$P_3$	57.350	128.330	153.321
$P_4$	120.247	134.645	122.194
$P_5$	270.108	274.875	275.576
$P_6$	195.255	228.368	242.35
$P_{loss}$	11.268	24.889	27.518
Cost (\$/h)	39909.6	42000.33	27069.8

**Table 9** Comparison of PSO with other method (for 900 MW)

Load	PSO	CSO	GA
$P_1$	37.73	28.17	37.09
$P_2$	128.11	40.47	18.83
$P_3$	181.94	176.83	153.42
$P_4$	206.00	138.64	156.77
$P_5$	172.1509	279.85	290.90
$P_6$	161.2072	267.09	280.86
$P_{loss}$	12.86	35.46	37.81
Cost (\$/h)	46538.2	47107.29	47363.66

evolutionary optimization techniques such as cuckoo search and GA. The convergence of PSO has been shown in the results section. In this paper, two test systems of 3-unit and 6-unit have been taken into account and for different load demands, best optimized cost has been estimated for the economic operation in power system. It is also found that by using evolutionary techniques, reliability of the system increases as well as consumption of time is much less than the other conventional methods.

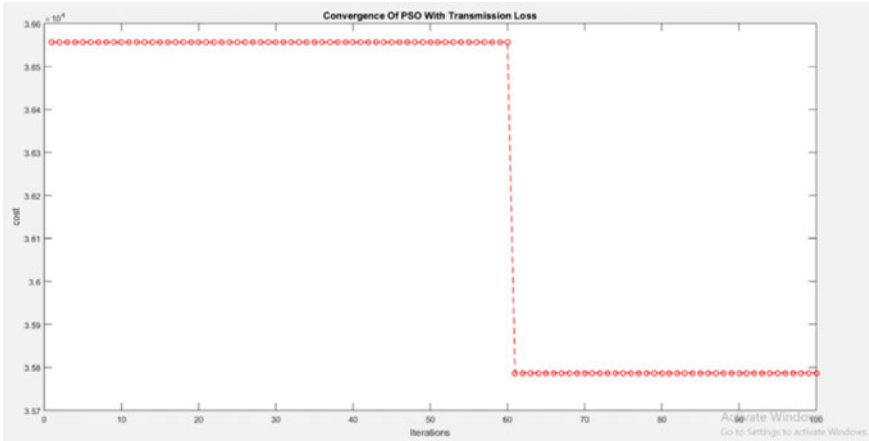


Fig. 7 Convergence of 6-unit system (800 MW)

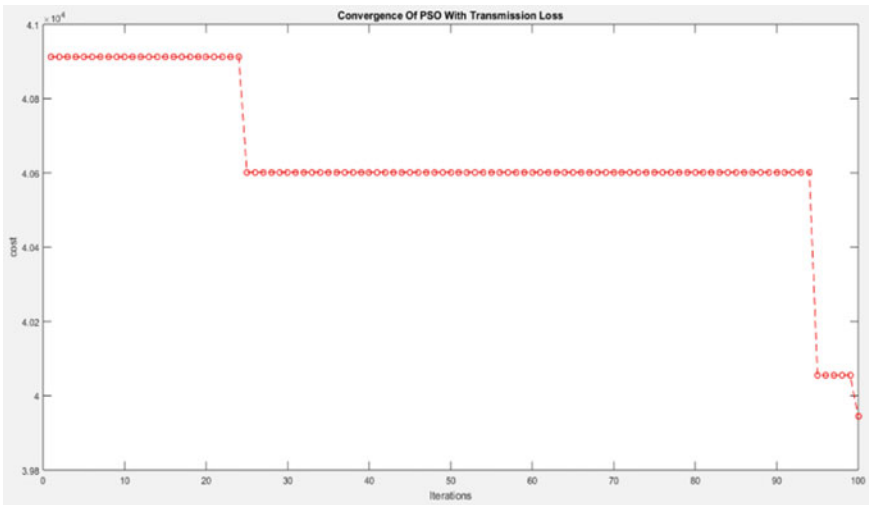


Fig. 8 Convergence of 6-unit system (800 MW)

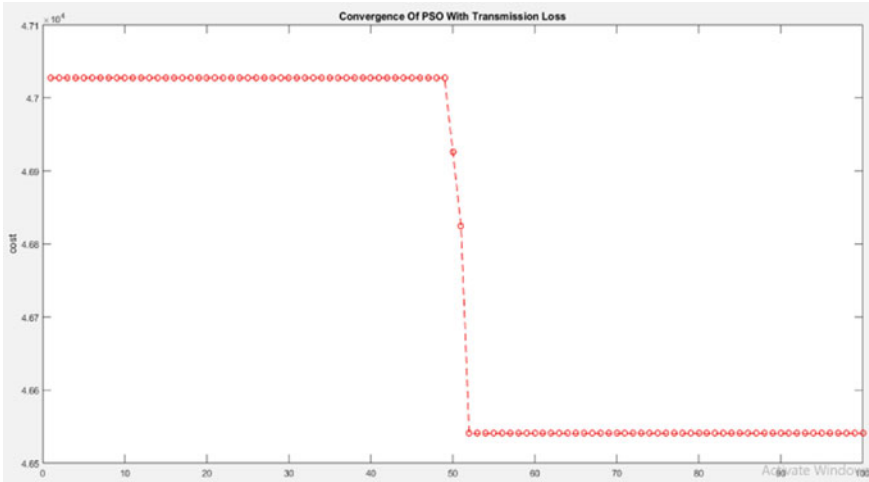


Fig. 9 Convergence of 6-unit system (900 MW)

## References

1. Kumar S, Singh HD (2018) Economic load dispatch problem of power plant in thermal electric generation system using swarm optimization technique. *Int Res J Eng Technol IRJET* 05(02)
2. Maharana HS, Nayak BTM (2018) Solution of combined emission economic load dispatch in thermal power plant by using particle swarm optimization. *Int Res J Eng Technol IRJET* 05(9)
3. Mishraa SK, Mishrab SK (2015) A Comparative study of solution of economic load dispatch problem in power systems in the environmental perspective. In: *International conference on intelligent computing, communication & convergence*. Elsevier, Bhubaneswar, Odisha, India
4. Raza R, Ansari MA, Pachauri RK (2016) Short-term load forecasting using ANN technique. In: *Proceedings of international conference on communication and computing systems (ICCCS-2016)*. India, pp 481–485
5. Sahoo S, Dash KM, Barisal AK (2014) Solution of economic load dispatch by evolutionary optimization algorithms—a comparative study. *Int J Appl Innovation Eng Manag* 3(3):240–246
6. Farheen, Ansari MA, Kardam N (2013) Implementation of particle swarm optimization for dynamic economic load dispatch problem. In: *Proceedings of the IEEE international conference on energy efficient technologies for sustainability (ICEETS' 13)*. India, pp 1273–1278
7. Sharma A, Vadhera S (2017) Comparative analysis of economic load dispatch using evolutionary and nature based algorithms. In: *International conference on power and embedded drive control (ICPEDC)*
8. Kaur A, Singh HP, Bhardwaj A (2014) Analysis of economic load dispatch using genetic algorithm. *Int J Appl Innovation Eng Manage* 3(3):240–246
9. Kardam N, Ansari MA, Farheen (2013) Communication and load balancing using SCADA model based integrated substation. In: *Proceedings of the IEEE international conference on energy efficient technologies for sustainability (ICEETS' 13)*. India, pp 1256–1261
10. Li P, Xu D, Zhou Z, Lee W-J, Zhao B (2016) Stochastic optimal operation of microgrid based on chaotic binary particle swarm optimization. *IEEE Trans Smart Grid* 7(1):66–73
11. Dasgupta K, Banerjee S (2014) An analysis of economic load dispatch using different algorithms. In: *Proceedings of 1st international conference on non-conventional energy (ICONCE 2014)*



12. Tiwari S, Ansari MA, Kumar K, Chaturvedi S (2018) Load flow analysis of IEEE 14 bus system using ANN technique. In: IEEE international conference on sustainable energy, electronics, and computing systems (SEEMS-2018). India, pp 1–5
13. Hosseinnzhad V, Rafiee M, Ahmadian M, Ameli MT (2014) Species-based quantum particle swarm optimization for economic load dispatch. *Electr Power Energy Syst* 63:311–322

# Effective Vibration Damping Using Self-tuning Smart Material



Gayatri R. More and Sharada N. Ohatkar

**Abstract** In order to avoid damage to mechanical structure, the vibration damping is necessary. The conventional methods of damping are less effective as well as tedious. This paper deals with passive techniques of vibration damping using piezoelectric material. Piezoelectric materials called as smart materials are bonded to structure and can be used as actuator as well as sensor. By rigorous variations of  $R$  and  $L$  values, a RLC circuit is designed, which nullifies the vibrations at the frequency of resonance. It results in 80% of vibration damping as compared to conventional methods.

**Keywords** Shunt damping · Vibration control · Piezoelectric material · Resonance · Sensor · Actuator

## 1 Introduction

The most effective way of reducing vibrations and noises is shunt damping. One of the methods of passive vibration damping is the use of piezoelectric material. Piezoelectric material is key element, which has two-way effect. It can be used as sensor and actuator. Here we are using ‘Lead (Pb) Zirconate Titanate (PZT)’ [1], piezoelectric patches with an appropriate circuitry to create a phenomenon called resonance. If the system consists of series inductor—resistor circuit which is connected to the mechanical structure having PZT capacitor with piezoelectric patch on it, which creates damping vibrations when a mechanical structure is hit. Since we are using passive elements to reduce the vibrations, this technique is called as ‘Passive Shunt Damping’. This technique is simple, lightweight, easy to implement and low in cost. Unlike all active techniques, this technique remains a controlled system that is always stable in the presence of any mechanical structural vibrations.

---

G. R. More (✉) · S. N. Ohatkar  
MKSSS's Cummins College of Engineering, S. P. Pune University, Pune, India  
e-mail: [gayatri.r.more@cumminscollege.in](mailto:gayatri.r.more@cumminscollege.in)

S. N. Ohatkar  
e-mail: [sharada.ohatkar@cumminscollege.in](mailto:sharada.ohatkar@cumminscollege.in)

The term vibration damping means reducing the energy which is lost in the form of oscillations this is the main objective of the paper. Second objective is to avoid probability of resonance which is result of vibration. To reduce vibration, isolators and absorbers are most widely used, but proposed design reduce the design complexity. Another objective is to reduce noises due to vibrations. Here contribution has made to meet the desired objectives.

There is one problem associated with PZT passive shunt damping technique. Typically this requires large inductance, so that design of high capacity inductor is troublesome. Physically designing of high capacity inductor using ferrite or iron core and copper coil will lead to clumsy, heavy and space-consuming physical attribute. Therefore, synthetic inductor design should come into consideration [7].

This paper will give you an introduction of piezoelectric materials and their properties, single-mode shunt damping and that of  $R-L$  circuit designing [2]. And let us have a look at the idea of synthetic inductor. Circuit diagrams and formulae are given to calculate the values of resistor and inductor. The points below contain some practical readings of PZT piezoelectric shunt damping circuit on two different tuning forks. We have taken varies inductor and resistor values into consideration. Results conclude that 80% of vibrations are damped. This method effectively reduces the vibrations of any mechanical structure. We can experience this effect just by giving ears to the sounds of hitting the tuning fork when connected and disconnected with the shunt circuit. This explains passive noise cancellation.

## 2 Methodology

### 2.1 *Piezoelectric Materials*

The piezoelectric materials are known as ‘Smart Materials’. This is because the phenomenon of piezoelectricity. Piezoelectricity occurs when non-centrosymmetrical crystal generates electric dipole moments when subjected to mechanical deformation. The converse is also possible, when electric field is applied, it deforms mechanically. In piezoelectric material theory, these two effects are called direct effect and indirect effect. One is when stress as a stimulus is applied and gets electric charge in response and second, is when electric charge is applied and gets strain produced in response. So that it is called as two-way effect, i.e., sensor and actuator [1].

### 2.2 *Capacitance of Piezoelectric Materials*

The capacitance of piezoelectric mainly depends on three parameters, area, thickness and its material properties. Let us see, for example, piezo-stacks. Capacitance of piezo-stacks depends on the number of layers in a stack, distance between two layers and thickness of the stack as well as material properties.

Following formula defines the capacitance of piezo-stack;

$$C = N * \varepsilon_{33}^T * \frac{A}{d_s} \quad (1)$$

where,  $C$  = capacitance (F),  $\varepsilon_{33}^T$  = dielectric constant,  $A$  = surface of electrode and

$$N = \frac{L_a}{d}, \quad (2)$$

where,  $N$  = number of layers,  $L_a$  = length of piezo and  $d$  = distance between electrodes.

From the above equations, relations of the capacitance of piezoelectric material can be given as, if the square of  $N$  that is number of layers increases the capacitance also increases. Therefore, when we compare the capacitance of a piezo-stack manufactured with layers of 100  $\mu\text{m}$  thickness and the capacitance of a piezo-stack with layers of 1 mm thickness, then 1 mm piezo-stack will conserve 100 times more power as well as capacitance of 100  $\mu\text{m}$  piezo-stack is 100 times the other one. Here power given to both the capacitors is same.

### ***2.3 Natural Frequency***

Every structure vibrates when force is applied, for example, hitting a tuning fork. The fact is that each time the tuning fork is hit with mallet; it will vibrate with the same frequency. This frequency is called as natural frequency of the tuning fork.

The periodic frequency with which force is applied is called as forced frequency. When the natural frequency and the forced frequency are equal, then the amplitude of oscillations increases continuously so as structural vibrations increase. This phenomenon is called as resonance (Fig. 1).

### ***2.4 Single-Mode Passive Shunt Damping Using Piezoelectric Materials***

In shunt damping, passive elements are mainly used as resistors, capacitors and inductors. Vibrational modes damping of flexible mechanical structure and also its efficiency very much depends on accurate selection of passive components. Shunt damping resonant circuit can be categorized into two types [4]:

1. Single-mode shunt damping resonant circuit.
2. Multimode shunt damping resonant circuit.

**Fig. 1** Tuning fork hitting with a mallet



If the shunt circuit contains only resistor, then that resistive shunt circuit will become a single-order electric circuit and result will be the dissipation of energy in the resistor; in the form of heat and that will give less damping of vibrations. Therefore, use of resistor and inductor as a shunt circuit is used. This becomes a two-order shunt circuit and gives impressive mechanical damping of vibrations. The idea behind this is resonance, if the natural frequency of shunt circuit and mechanical system matches, it gives effective cancellation of vibrations. This design is also called as vibration absorber [2].

## 2.5 R–L Shunt Series Circuit Design

In series  $R$ – $L$  shunt circuit,  $w_o$  and  $w_s$  are the natural frequencies of piezoelectric path terminals when open and short circuited, respectively [3].

$K_{31}^2$  is generalized electromechanical coupling coefficient.

$$K_{31}^2 = \frac{\omega_o^2 - \omega_s^2}{\omega_s^2} \quad (3)$$

Electromechanical coupling coefficient:

$$k_{31} = \frac{d_{31}}{\sqrt{\epsilon_0 k_3^T s_{11}}} \quad (4)$$

Electrical resonance frequency:

$$\omega_e = \delta \omega_o \tag{5}$$

And,

$$\delta = \sqrt{1 + K_{31}^2} \tag{6}$$

Capacitance of piezo element at constant strain:

$$C_p^S = C_p^T(1 - k_{31}^2) \tag{7}$$

Inductance required:

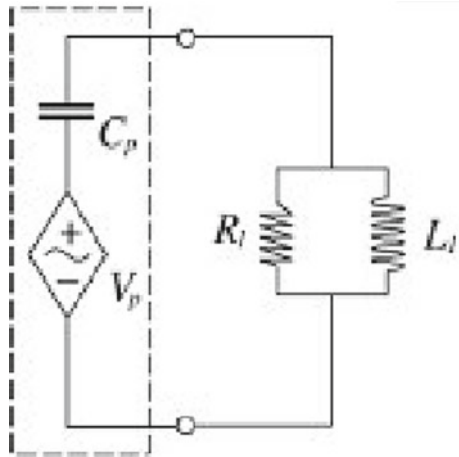
$$L = \frac{1}{\omega_e^2 C_p^S} \tag{8}$$

Resistance required:

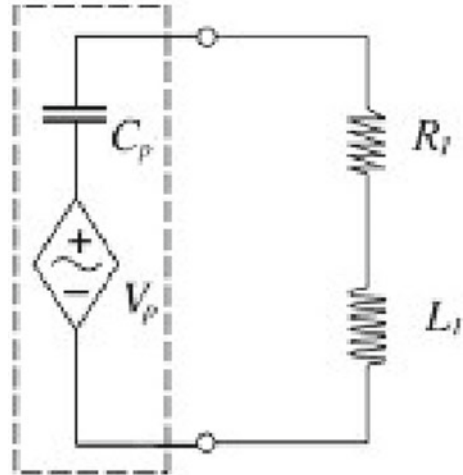
$$R = \frac{\sqrt{2}K_{31}}{C_p^S \omega_o(1 + K_{31}^2)} \tag{9}$$

The circuit diagram for parallel shunt is (Figs. 2 and 3),

**Fig. 2** Circuit diagram of R-L shunt circuit in series fashion



**Fig. 3** Circuit diagram of  $R$ - $L$  shunt circuit in parallel fashion



### 3 Experimental Configuration

Implementation experiments of single-mode  $R$ - $L$  shunt circuit are performed to validate the theory. Passive shunt damping circuit system is performed on tuning fork of 512 Hz.

The tuning fork has dimensions of length 15 mm, height of 85 mm and thickness of 5 mm. The piezoelectric patch of PZT material is bonded to the either arms of tuning fork. PZT patch has dimensions, thickness 0.4 mm, height 50 mm and length 20 mm. PZT patch is bonded with very thin epoxy layer.

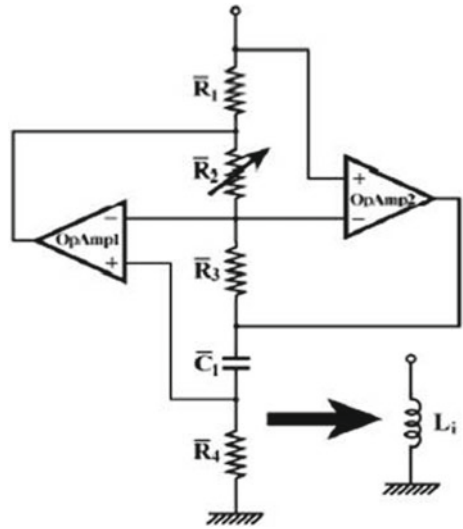
The capacitance of PZT piezoelectric is measured using impedance analyzer. The PZT piezoelectric patch is operating in transverse mode that is in positive  $z$  direction, along its thickness. All the parameters relating transverse mode are given in the datasheet in the respective PZT. The tuning fork is set in motion using a mallet. By using the cathode-ray oscilloscope (CRO), damping vibration readings are taken. The difference between with circuit response and without circuit response is observed.

#### 3.1 Design of a Synthetic Inductor

Following are the steps involved to design the synthetic inductor:

1. The natural frequencies of tuning fork with PZT piezoelectric patch.
2. Calculating the theoretical and practical capacitances of system and the PZT patch.
3. Calculating the values of inductor  $L$  and resistor  $R$ .

**Fig. 4** Circuit diagram of synthetic inductor



But for this to make happen, the value of Inductor is in few Henries which is very high. Designing such an inductor using the ferrite or iron core and copper wire windings, the size of the inductor will increase and design will get clumsy and space-consuming, to implement practically. So it is desirable to design the inductor using passive components [5].

Synthetic inductor is implemented using op-amps, resistors and potentiometer. The circuit is lightweight and easily uses to vary large inductance value using the potentiometer in the synthetic inductor circuit.

There are drawbacks of this circuit designing, the power source to the circuit and it contains several resistors, which are proportional to the inductance, this may have some effect on the shunt circuit [6] (Fig. 4).

### 4 Experimental Results

When calculated the values of inductor and resistor for two tuning forks with different natural frequencies and different PZT piezoelectric patches, here observed the waveform on CRO hence  $\Delta t$  is calculated,

$$\Delta t = t_1(\text{un-damped}) - t_2(\text{damped}). \tag{13}$$

In the following experiment, inductor values are set fixed and resistor values are varying. Then the damped and undamped waveforms are observed on CRO, by marking the cursors, observe the time readings. The readings with greater delta time are the more appropriate reading.

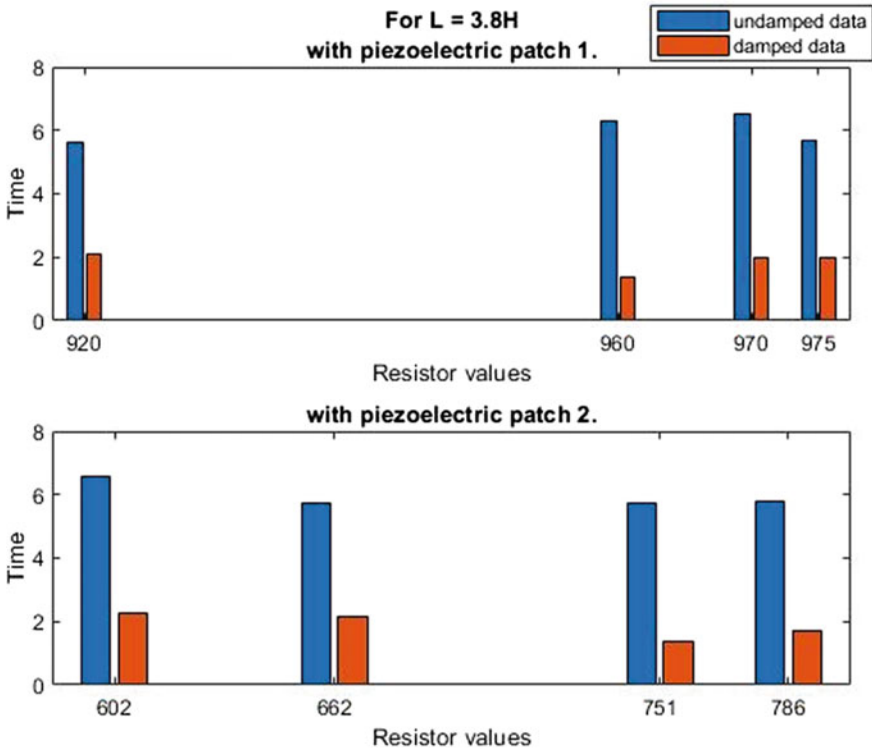


**Table 1** Readings of  $t_1$  and  $t_2$  for tuning fork 2

Tuning fork 1 (512 Hz)			
PZT piezo-patch 2			
	$t_1$	$t_2$ (ms)	$\Delta t$
$L = 3.2$ H			
$R = 608$	2.920	600	2.32
642	2.920	650	2.27
662	2.920	662	2.258

Here, In Table 1, for tuning fork of natural frequency 512 Hz is used with piezo-patch 1 of same thickness above mentioned. According to the table, these all are very close values and give very effective damping. So resistor of 638  $\Omega$  with inductor of 3.2 H will give you the best damping of vibrations and noise.

Following bargraphs are prepared using the reading over second tuning fork of 480 Hz and with another piezo-patch of same dimensions (Fig. 5).



**Fig. 5** Bar graph of damped and undamped time

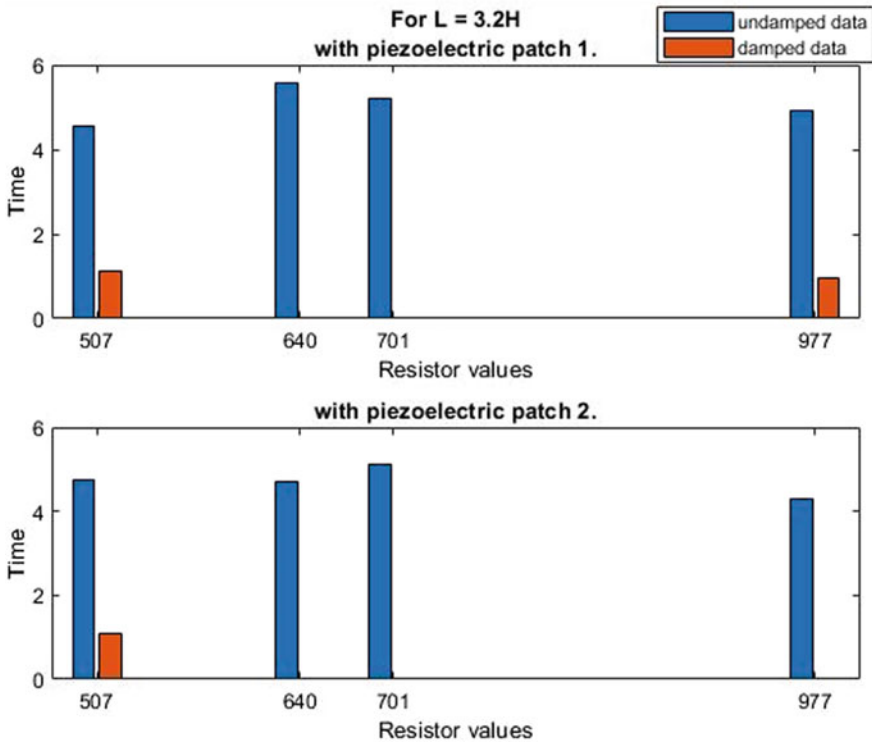


Fig. 6 Bar graph of damped and undamped time

From the above results, for (tuning fork 1, piezo-patch 2) inductor value of 3.2 H and resistor value of 702 Ω, best results are observed and for (tuning fork 2, piezo-patch 2) inductor value of 3.2 H and resistor value of 608 Ω, best results are observed. At these values, natural frequencies of tuning fork and RLC circuit get close, and vibrations are drastically damped (Fig. 6).

## 5 Conclusion

Passive shunt damping using piezoelectric material results in effective damping of first structural mode of tuning fork. When the designed system can generate the equivalent natural frequency similar to original system, then resonance phenomenon happens. If we plot the FFT of our results, we get highest peak at the resonant frequency. By observing the graphs, 80% of vibrations are getting reduced. By hearing the sound of tuning fork when hitting with a mallet, effective damping is easily distinguished.

## References

1. Fleming AJ, Behrens S, Moheimani SOR (2003) Reducing the inductance requirements of piezoelectric shunt damping systems. *Smart Mater Struct* 12(1):57
2. Fleming AJ, Behrens S, Moheimani SOR (2003) A new approach to piezoelectric shunt damping. Department of Electrical and Computer Engineering, The University of Newcastle
3. Høgsberg JB, Krenk S (2015) Piezoelectric RL shunt damping of flexible structures. In: Arauo AL, Mota Soares CA (eds) *Proceeding of 7th ECCOMAS thematic conference on smart structures and materials IDMEC*
4. Jeon JY (2009) Passive vibration damping enhancement of piezoelectric shunt damping system using optimization approach. *J Mech Sci Technol* 23(5):1435–1445
5. Wu SY, Bicos AS (1997) Structure vibration damping experiments using improved piezo-electric shunts. In: *Proceedings of the international society for optical engineering*
6. Wu SY (1996) Piezoelectric shunts with a parallel R-L circuit for structural damping and vibration control. In: *Proceeding in SPIE smart structures and materials, passive damping and isolation*
7. Yan B, Wang K, Hu Z, Wu C, Zhang X (2017) Shunt damping vibration control technology: a review. *Appl Sci* 7(5):494

# Performance Analysis of BIPV Solar Panel Under the Effect of External Conditions



Ravi Sagar, Nidhi Singh Pal, M. A. Ansari, Nivedita Singh and Dilip Yadav

**Abstract** In the present scenario, solar energy contributes maximum to the pool of renewable energy. The rate of consumption of energy is increasing day by day; therefore, the demand of energy is also increasing. To satisfy this increase in demand of energy, building integrated photovoltaic module (BIPV) system is one of the ways to satisfy the consumer demand by providing the energy according to the consumption, since the BIPV system is mainly affected by the outermost conditions the most, and therefore, in the present work, its performance is analyzed under the effect of external conditions. In this paper, the observations under different external conditions are taken which follow the effect of coal, dust, and shading as well as for the normal conditions. The results are then compared and are observed that the performance of BIPV solar panel is low under the effect of external conditions as compared to normal conditions.

**Keywords** BIPV · Solar energy · Consumption · PV · External conditions

## 1 Introduction

Solar energy, which is an important source of renewable energy, is basically radiant light and heat from the Sun that is harnessed using a range of ever-evolving technologies such as solar heating, photovoltaics, and solar thermal energy. It refers to capturing the energy from the Sun and subsequently converting it into electricity with the help of PV solar panel, which is an electrical device [1–3]. Basically, there are two types of solar panels [4]:

---

R. Sagar · N. S. Pal · M. A. Ansari · N. Singh (✉) · D. Yadav  
Department of Electrical Engineering, Gautam Buddha University, Greater Noida, U.P., India  
e-mail: [nivedita.singh10@gmail.com](mailto:nivedita.singh10@gmail.com)

R. Sagar  
e-mail: [ravisagarravi9958@gmail.com](mailto:ravisagarravi9958@gmail.com)

M. A. Ansari  
e-mail: [ma.ansari@ieee.org](mailto:ma.ansari@ieee.org)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_8](https://doi.org/10.1007/978-981-15-2329-8_8)

- (i) Single-sided solar panel: Solar cells are installed on one side of the panel.
- (ii) Double-sided solar panel: Solar cells are installed on both sides of the panel.

The performance of PV panel depends upon the intensity of sunlight falling onto its surface. Many PV panel manufacturers are ‘vertically integrated’ which means that a company supplies and manufactures all the important ingredients including silicon and wafers which are used to make solar PV cells [5]. However, the manufacturers combine solar panels when they include sales, polymer back sheet and encapsulated EVA material in outer parts, so that they can be more selective about what they have to procure, though they do not always control product quality [6–9].

The PV panel is consisting of several solar cells, and each cell contains the 0.5 V. The main solar panel components are—frame, glass, encapsulant, solar cell, back sheet, junction box, etc. [10].

Glass is attached with a frame which makes the panel stable, and it also provides the strength to the panel. Encapsulant is used for the protection of solar panel. There are two types of encapsulants; the solar cells are arranged in between the encapsulant. The back sheet has the junction box, where are basically two diodes and are used to prevent the reverse current flow through the panel [11].

In BIPV solar panel, the solar cell is installed at both sides. Silicon photovoltaic cells or PV cells convert sunlight directly into DC electrical energy. The solar panel performance is determined by the type and characteristics of silicon, in which two of the main type are monocrystalline. The construction made thus converted into the generator based on a durable power source on the development of the construction pipe in integrated photovoltaics (BIPV) [12]. The purpose of BIPV is to give life with general plan objectives, for example, feeling and natural control. Apart from this, the test is that at the time of stress planning during the time of planning, flag management, for example, manages to gain the greatest power of the ideology and sunlight. In creative matters, characteristics of the common BIPV have been altered by experts, so that these types of conflicts can be resolved in the BIPV Fox Plan as an increasing format [13].

There are extensive audits on app viding types, planning collections, reception barriers, and process of implementation; BIPV customization has been investigated. BIPV has not been diagnosed as a specific response. This paper tries to find the BIPV customization capability as a method of increasing the BIPV selection. The current paper separates custom building capabilities parameters with customization rating, level, and technique, and evaluation of their effects [14]. A building integrated photovoltaics (BIPV) framework comprises incorporating photovoltaics modules into the building envelope, for example, the rooftop or the exterior. By the same time filling in as building envelope material and power generator, BIPV frameworks can give investment funds in materials and power costs, diminish the utilization of non-renewable energy sources and emanation of ozone exhausting gases, and add compositional enthusiasm to the building (Fig. 1).



**Fig. 1** BIPV solar panel (mono)

A few board producers are ‘vertically coordinated’ which implies that the one organization supplies and fabricates all the principle segments including the silicon ingots and wafers which are used to make the Sun-powered PV cells. However, many board makers assemble Sun-oriented boards utilizing remotely sourced parts including cells, polymer back sheet, and EVA material. These makers can be increasingly particular about which segments they pick up; however, they do not generally have authority over the nature of the items so they ought to make certain that they utilize the best providers accessible [15].

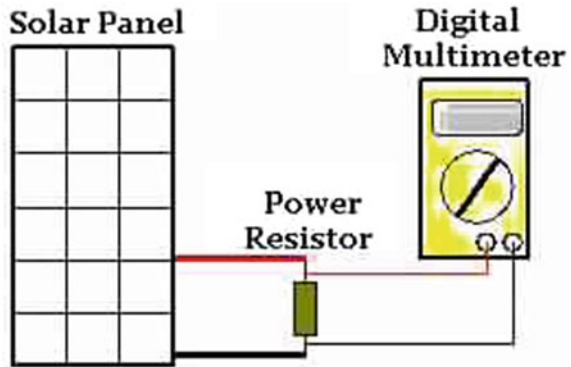
## 2 Experimental Setup

The following components are required for the experimental setup to demonstrate the observations under various working conditions:

### 1. *BIPV Panel*

Basically, building integrated photovoltaic (BIPV) alludes to photovoltaic cells and modules which can be coordinated into the building envelope as a feature of the building structure, and in this way it can supply ordinary building materials, as opposed to being introduced a short time later.

Fig. 2 Connection diagram



## 2. Multimeter

The multimeter is utilized for the estimation of voltage and current of PV board display. It is chiefly used to gauge the short out current, yield voltage, and load opposition.

## 3. Rheostat

It is a movable or variable resistor. It is utilized to control the electrical obstruction of a circuit without intruding on the stream of flow.

## 4. Temperature Meter

The temperature meter is furnished with an estimating tip and an estimating clasp. Since the temperature meter utilizes infrared radiation for temperature estimation, no immediate contact with the estimating object is required. It measures the diverse dimension of temperature of the board at various distinctive points.

## 5. Connecting Wires

Wires are utilized to make the association of ventures. The associations can be in arrangement or parallel as per the requirement and are utilized to tolerate mechanical burdens or power.

In this arrangement, the output terminal of solar panel is connected with the variable resistor, and both ends of the variable resistor are connected with the multimeter (Figs. 2 and 3).

# 3 Results and Discussion

The experimental observations are taken under various working conditions such as—normal conditions, effect of coal, effect of dust, and effect of shading. The performance under these conditions ( $I-V$  and  $P-V$  characteristics) is analyzed and



**Fig. 3** Actual demonstration experimental setup

is given through Tables 1, 2, 3, and 4 under different irradiances (478–705 W/m<sup>2</sup>). The performance characteristics are represented through Figs. 4, 5, 6, 7, 8, 9, 10, and 11.

- (i) *Case-(A): Normal Conditions—at irradiance of 705 W/m<sup>2</sup>*
- (ii) *Case-(B): Effect of Coal—at irradiance of 694 W/m<sup>2</sup>*
- (iii) *Case-(C): Effect of Dust—at irradiance of 704 W/m<sup>2</sup>*
- (iv) *Case-(D): Effect of shading at irradiance of 478 W/m<sup>2</sup>*

**Table 1** Parameters of BIPV panel at irradiance 705 W/m<sup>2</sup>

Voltage	Current	Power	R (Ω)
0	0.8	0	62.5
11	0.74	8.14	125
19.37	0.7	13.55	187.5
23.87	0.68	16.23	250
29.5	0.62	18.29	312.5
35.1	0.42	14.74	375
36.3	0.1	3.64	437.5
36.1	0	0	498



**Table 2** Parameters of BIPV cell at irradiance at 694 W/m<sup>2</sup>

V (V)	I (A)	P (W)	R ( $\Omega$ )
0	0.67	0	62.5
8.81	0.63	5.55	125
15	0.54	8.1	187.5
18.8	0.45	8.46	250
23.2	0.31	7.19	312.5
29.1	0.1	2.91	375
32.3	0.04	1.29	437.5
33.2	0	0	498

**Table 3** Parameters of BIPV cell at irradiance of 704 W/m<sup>2</sup>

V (V)	I (A)	P (W)	R ( $\Omega$ )
0	0.76	0	62.5
10.44	0.68	7.09	125.3
18.32	0.63	11.54	187.3
22.7	0.6	13.62	249.8
31.8	0.41	13.03	309.5
33.4	0.2	6.68	372
34.9	0.07	2.44	437
35.7	0	0	500

**Table 4** Parameters of BIPV cell at irradiance of 478 W/m<sup>2</sup>

V (V)	I (A)	P (W)	R ( $\Omega$ )
0	0.39	0	32
6	0.37	2.28	62.5
9.68	0.35	3.38	93.75
11.93	0.34	4.05	125
14.75	0.31	4.57	156.2
17.55	0.21	3.68	187.5
18.15	0.05	0.91	218.75
18.05	0	0	250

After comparing all the above cases, we have concluded that the maximum output of power occurred in the normal condition and less power output occurred in the effect of shading.

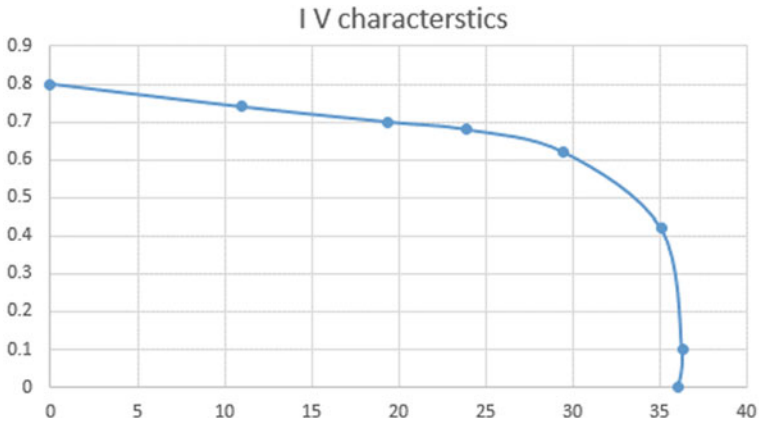


Fig. 4 I-V characteristics at irradiance 705 W/m<sup>2</sup>

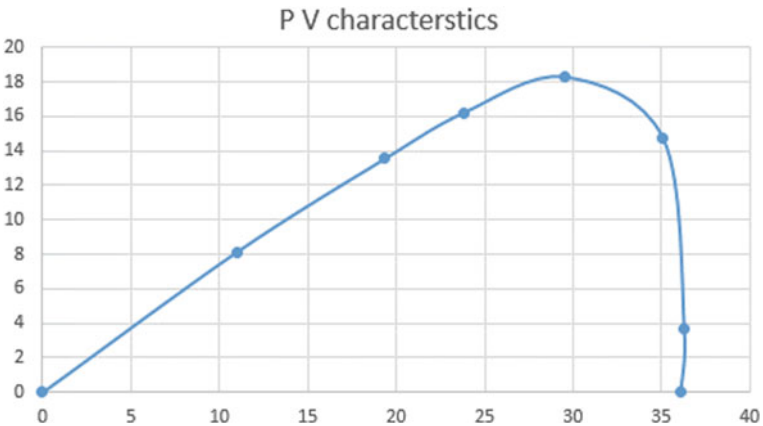


Fig. 5 P-V characteristics at irradiance 705 W/m<sup>2</sup>

### 4 Conclusion

The impact of external conditions on Sun-based board is negative because of its hard molecular structure. Under the impact of coal, the yield control of the Sun-oriented board declines. The most extreme yield of the voltage reported is 33.5 V, and the short out current reported is 0.82 A in coal condition; whereas, the highest yield voltage was 36.10 V in the ordinary conditions. In other external conditions, the voltage and the current parameters have decreased and have also given the less output power. The efficiency of solar panel is also reduced. Overall, the observations under different

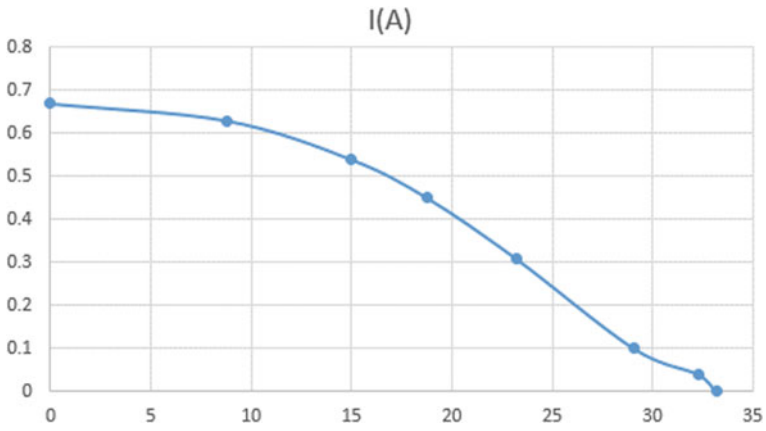


Fig. 6  $I-V$  characteristics at irradiance at  $694 \text{ W/m}^2$

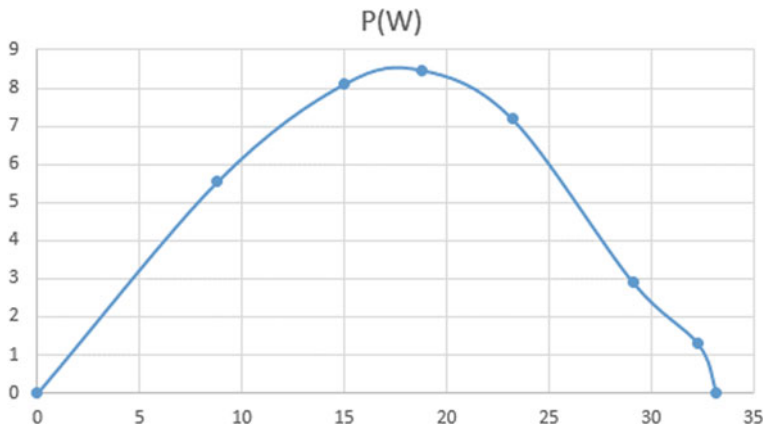


Fig. 7  $P-V$  characteristics at irradiance at  $694 \text{ W/m}^2$

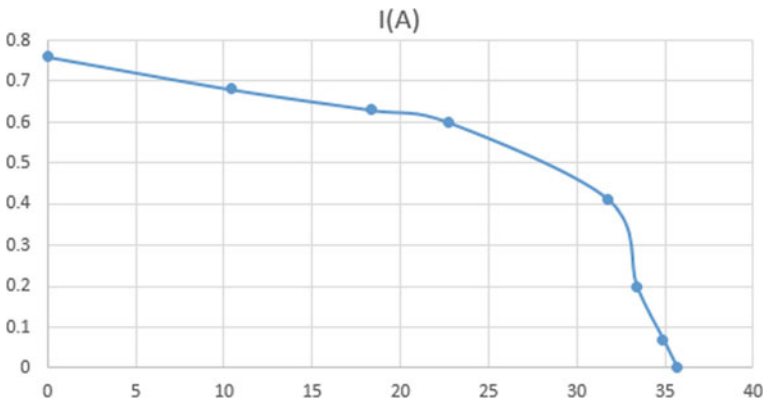


Fig. 8  $I-V$  characteristics at irradiance at  $704 \text{ W/m}^2$

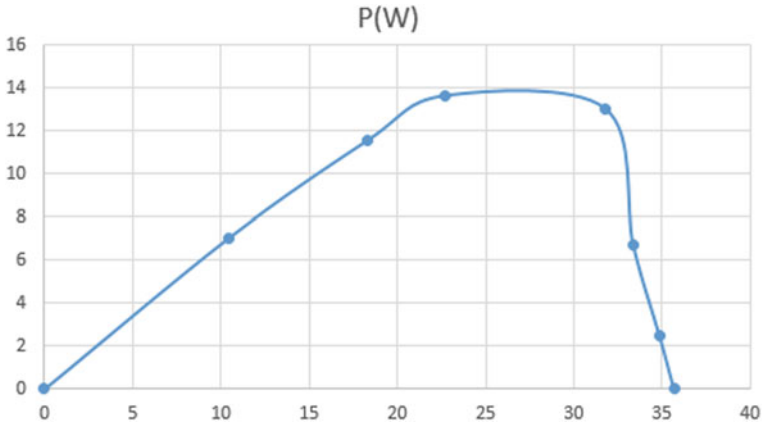


Fig. 9 P–V characteristics at irradiance at 704 W/m<sup>2</sup>

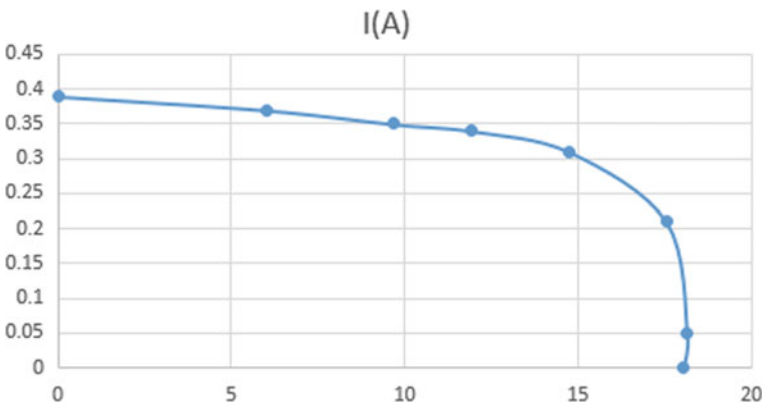


Fig. 10 I–V characteristics at irradiance of 478 W/m<sup>2</sup>

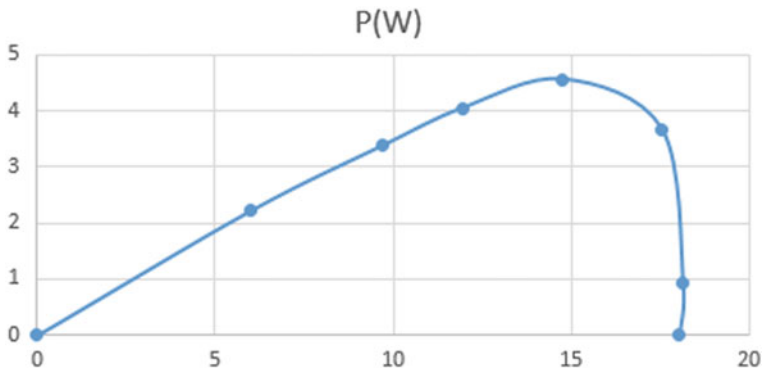


Fig. 11 P–V characteristics at irradiance of 478 W/m<sup>2</sup>

external conditions are taken which follow the effect of coal, dust, and shading as well as for the normal conditions and observed that the best performance comes out in normal conditions. However, in the future, more external conditions may be incorporated to improve the efficiency of the BIPV panels.

## References

1. Tacluc M (2011) PV cells IV characteristics explicit equations with three parameters and its simplified forms. In: International symposium on advanced topics in electrical engineering, pp 709–714
2. Stellari F, Steen SE et al (2010) PV cells characteristics using advance optical tools. IEEE international reliability physics symposium, pp 508–515
3. Izquierdo S, Rodrigues M, Fueyo N (2008) A method for estimating the geographical distribution of the available roof surface area for large-scale photovoltaic energy-potential evaluations. *J Solar Energy* 82(1):929–939
4. Alonso-Garcia MC, Ruiz JM (2006) Experimental study of mismatch and shading effects in the I-V characteristic of a photovoltaic module. *J Solar Energy Mater Solar Cells* 90(1):329–340
5. Skoplaki E, Boudouvis AG, Palyvos JA (2008) A simple correlation for operating temperature of photovoltaic modules of arbitrary mounting. *J Solar Energy Mater Solar Cells* 92(1):1393–1402
6. Krishan K, Ansari MA, Shreshth KV, Vinay R, Arjun T (2018) An efficient technique for power management in hybrid solar PV and fuel cell system. *Smart Sci* 6(3):234–244
7. Biyik E et al (2017) A key review of building integrated photovoltaic (BIPV) system. *Int J Eng Sci Technol* 20(3):833–858
8. Djamila H, Chu C-M, Kumaresan S (2014) Effect of humidity on thermal comfort in the humid tropics. *J Build Constr Plann Res* 2:109–117
9. Akata AMEA, Njomo D, Agrawal B (2017) Assessment of building integrated photovoltaic (BIPV) for sustainable energy performance in tropical regions of Cameroon. *Renew Sustain Energy Rev* 80(1):1138–1152
10. Krishan K, Ansari MA, Vinay R, Royal S (2018) Simulation based solar PV system: a cost-effective study. *Int J Appl Eng Res* 13(11):8894–8898
11. Cyrille Vincelas FF, Ghislain T, Robert T (2017) Influence of the types of fuel and building material on energy savings into building in tropical region of Cameroon. *Appl Therm Eng* 122:806–819
12. Shukla AK, Sudhakar K, Baredar P (2016) A comprehensive review on design of building integrated photovoltaic system. *Energy Build* 128:99–110
13. Quesada G, Rousse D, Dutil Y, Badache M, Hallé S (2012) A comprehensive review of solar facades. Transparent and translucent solar facades. *Renew Sustain Energy Rev* 16:643–2651
14. Quesada G, Rousse D, Dutil Y, Badache M, Hallé S (2012) A comprehensive review of solar facades. Opaque solar facades. *Renew Sustain Energy Rev* 16:2820–2832
15. Krishan K, Ansari MA, Shreya S (2019) Analysis and control of solar-wind power generation system using MATLAB software. *Int J Comput Appl* 182(34):1–5

# Classification of Prediabetes and Healthy Subjects in Plantar Infrared Thermal Imaging Using Various Machine Learning Algorithms



Usharani Thirunavukkarasu  and Snehalatha Umapathy 

**Abstract** In the course of recent years, the size of individuals with diabetes mellitus has been dramatically increased than before. There is a need for screening and interventions which could prevent the individuals from the serious diabetic complications. Prediabetes may be a forerunner of type two diabetes mellitus, as well as a risk factor for heart illness. The body temperature is an essential parameter used for indicating the abnormal activity of human tissues. The thermal imaging primarily uses the infrared radiation emitted from the body naturally. The aim of this study was to evaluate the potential of thermography in screening the prediabetes. Sixty subjects were recruited for this study. Group I: HbA<sub>1c</sub> is <5.7%, Group II: HbA<sub>1c</sub> is 5.7–6.4%, Group III: HbA<sub>1c</sub> is >6.5%. The plantar thermograms were captured, and the temperature was measured at toe, metatarsal 1, metatarsal 3, metatarsal 5, instep and heel, respectively. The HbA<sub>1c</sub> was measured using the standard biochemical method. Three groups were categorized based on the accuracy rate obtained by five different machine learning algorithms (support vector machine, random forest, Naïve Bayes, multilayer perceptron and k-nearest neighbour). In prediabetes group, HbA<sub>1c</sub> exhibited positive correlation with measured temperature at toe region ( $r = 0.917$ ,  $p < 0.01$ ) and the negative relationship with measured temperature at metatarsal 1 ( $r = -0.474$ ,  $p < 0.05$ ), metatarsal 3 and heel regions ( $r = -0.895$ ,  $-0.901$ ,  $p < 0.01$ ). The support vector machine has outperformed the other classifiers with good accuracy rate as 81.6%. The findings from this preliminary study indicate that measured temperature from plantar thermograms may be useful in screening the population for prediabetes.

**Keywords** Prediabetes · Thermal imaging · Plantar region · Machine learning algorithms

---

U. Thirunavukkarasu · S. Umapathy (✉)

Department of Biomedical Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Chennai 603203, India  
e-mail: [sneha\\_samuma@yahoo.co.in](mailto:sneha_samuma@yahoo.co.in)

U. Thirunavukkarasu

e-mail: [ushabme@gmail.com](mailto:ushabme@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_9](https://doi.org/10.1007/978-981-15-2329-8_9)

## 1 Introduction

People at the danger zone of the future diseases have to be recognized at the earliest to reduce the risk of development to disease is an essential aim of decreasing the burden of conditions such as type two diabetes mellitus (DM). Furthermore, the counteractive action of type 2 DM also offers the chance to decrease the danger of cardiovascular diseases that are the most important reason for the premature death [1]. Over the decades ago, the worldwide epidemic of type 2 DM was estimated by the epidemiologists who observed a rapid growth in the predominance of type 2 DM and prediabetes amongst the indigenous people living in the western lifestyle [2]. The prediabetes implies that our body is showing signs and symptoms that could lead to type 2 DM in the forthcoming years [3]. Globally, the prevalence of prediabetes is expanding, and it is anticipated that more than 470 million people can have prediabetes in 2030. Around 5–10% of individuals with prediabetes are ended up in diabetic, even though the conversion rate fluctuates by population characteristics and long-term impact of lifestyle interventions [4]. The overall prevalence of prediabetes in India was around 10.3%, and especially, the diabetes prevalence was observed higher in some affluent states like Chandigarh, Maharashtra and Tamil Nadu [5]. Overall 80 million people in India and around 3.9 million people in Tamil Nadu are suffering from prediabetes [6].

According to the World Health Organization (WHO), the prediabetes accommodates the range of in-between hyperglycemia primarily based at the fasting blood glucose (FBG) or postprandial plasma glucose (PPG) or HbA<sub>1c</sub> levels [7]. Generally, people with prediabetes do not have any signs and symptoms. The biochemical tests such as FBG, 2 h plasma glucose and HbA<sub>1c</sub> can be used to examine the prediabetes [8, 9]. The American Diabetes Association (ADA) sets the threshold value for FBG (5.6–6.9 mmol/L), PPG (7.8–11.0 mmol/L) and HbA<sub>1c</sub> (5.7–6.4% or 39–46 mmol/mol) for the definition of prediabetes [10, 11]. Like diabetes, prediabetes also builds the threat of developing macrovascular problems like cardiovascular complications and stroke. The prediabetes increases with the certain hazard factors, for example, obesity, history of diabetes and gestational diabetes [12]. A common indication for DM is the diabetic foot ulcer [13]. Though the biochemical tests for diagnosing the type 2 DM and prediabetes were considered as a standard gold method. There are some non-invasive methods still in research for assessing the type 2 DM from the plantar region such as Ipswich Touch Test [14], 10-g Semmes–Weinstein monofilament test [15], laser Doppler flowmetry [16] and biothesiometry test [17]. One such non-contact method is thermal imaging which is used to examine the thermal patterns in the human body. The infrared thermal imaging is a clinical imaging methodology for mapping the skin surface temperature distribution progressively with the human body. Also, thermal imaging has been appeared to be a valuable technique in the clinical administration of the diabetic foot [18]. Thus, the surface skin temperature can be the good pointer for health conditions. To the best of our knowledge, the research study in prediabetic plantar thermograms was found to be scarce. So, our focus of this preliminary study is to evaluate the thermal imaging

in plantar region for screening the prediabetes individuals. And, to distinguish the healthy, prediabetes and diabetes groups based on the classification, accuracy rate attained by the different machine learning algorithms such as support vector machine (SVM), Naïve Bayes (NB), multilayer perceptron (MLP), k-nearest neighbour (k-NN) and random forest (RF) using WEKA, data mining tool. The early detection of prediabetes can secure the individuals from any perilous diabetic stages which may lead to foot amputations.

## 2 Materials and Methodology

### 2.1 Study Population and Measurements

The camp was conducted at SRM Hospital and Research Centre, Tamil Nadu, India. Seventy-five South Indian subjects were voluntarily registered for the study, and it has been approved by an Institutional ethical board with an ethical clearance number (834/IEC/2015). The inclusion criteria for this study were subjects aged between 25 and 45 years, and exclusion criteria were subjects with cardiovascular problems, renal failure, fever, arthritis and hyper and hypothyroidism. Based on this inclusion and exclusion criteria, sixty subjects were recruited for this preliminary screening study, and remaining fifteen subjects were excluded from this study. The informed consent forms were acquired from all the enrolled subjects ( $n = 60$ ). The detailed questionnaire was administered to acquire the information on health status of the subjects participated in the study. For each participant, the baseline parameters such as height (cm), weight (cm), hip circumference (cm), waist circumference (cm), core body temperature ( $^{\circ}\text{C}$ ) and systolic and diastolic blood pressure (mm Hg) were measured. The blood samples were collected from all the recruited participants ( $n = 60$ ), and the glycated haemoglobin ( $\text{HbA}_{1\text{c}}$  in %) was measured using the standard biochemical method. The total study population ( $n = 60$ ), mean  $\pm$  standard deviation of the male subjects ( $n = 32$ ) age was  $34.68 \pm 6.92$  (years) and that of female subjects ( $n = 28$ ) was  $33.28 \pm 7.33$  (years), respectively. Based on the standard  $\text{HbA}_{1\text{c}}$  (%) criteria assigned by ADA [19], the sixty subjects were separated into three groups;

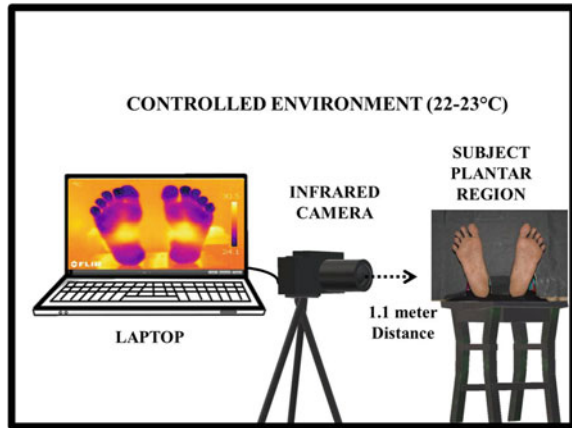
- Group I:  $n = 20$  Healthy subjects ( $\text{HbA}_{1\text{c}}$  is  $<5.7\%$ )
- Group II:  $n = 20$  Prediabetes subjects ( $\text{HbA}_{1\text{c}}$  is  $5.7\text{--}6.4\%$ )
- Group III:  $n = 20$  Type 2 DM subjects ( $\text{HbA}_{1\text{c}}$  is  $>6.5\%$ ).

### 2.2 Experimental Protocol and Thermal Imaging Analysis

As per the standards and guidelines for medical imaging, specified by the International Association of Certified Thermographers from the USA, the infrared thermal



**Fig. 1** Overview of an experimental setup for thermal imaging



imaging has been captured, and the overview of protocol setup has been illustrated in Fig. 1. The participants were guided to be in supine position without shoes or socks for 15 min in the temperature-controlled room of 22 °C which reduces the impact of surrounding temperature on the skin. The subjects were requested to remove the metallic ornaments before imaging. After equilibration, the thermograms were captured at a distance of 1.1 m from the plantar region of the subjects using an infrared thermal camera (FLIR A 305 SC, FLIR systems, USA). The images were saved and analysed later using infrared imaging software, IRT analyser version 6.0 GRAYESS, (Infrared Software and Solutions, Florida, USA). The software contains a variety of analysing tools such as area, histogram, hot detection, cold detection and different palettes. Among these various tools, the iron palette and the area tool have been chosen for analysing the images. The constant area has been fixed to measure the various spots in the plantar region. The six region of interest (ROI) such as toe, metatarsal 1, metatarsal 3, metatarsal 5, instep and heel were chosen plantar sites from the standard template designed for the biothesiometry device. Also, the Semmes–Weinstein monofilament 10g test has been performed in those plantar sites (toe, balls of foot, instep and heel) to detect the protective sensation in the feet of the subjects.

### 2.3 Classifiers and Statistical Analysis

The neural network classifiers such as SVM, NB, MLP, k-NN and RF were chosen from the WEKA data mining tool to perform the multiclass classification. All the measured plantar temperature at various ROIs and baseline variables of the subjects were fed into the classifiers. The ten cross-fold validation was performed to attain better classification accuracy. The random data selection was done by the data mining tool software to train and test the model in order to achieve the classification for each classifier. The true positive rate against the false positive rate plots the receiver

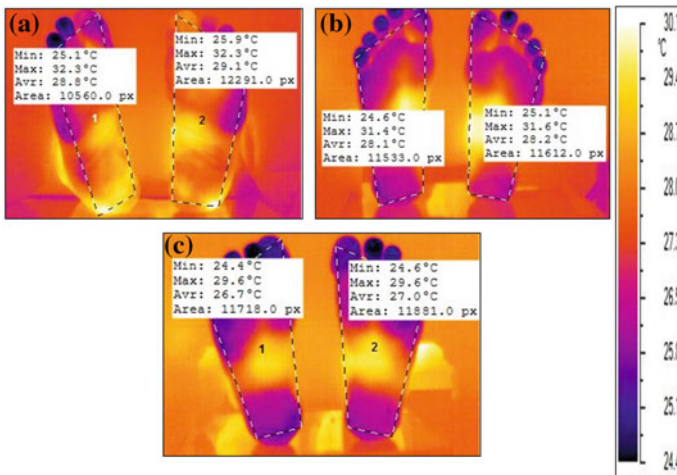
operating characteristic (ROC) curve. The accuracy was obtained based on TP, TN, FP and FN results.

- True Positive (TP): diseased subject predicted as diseased
- False Positive (FP): healthy subject predicted as diseased
- True Negative (TN): healthy subject predicted as healthy
- False Negative (FN): diseased subject predicted as healthy.

The data were expressed in mean  $\pm$  standard deviation (S.D), and the normalization was checked using Shapiro–Wilk test. An independent student t-test was computed for all the parameters. To explore the correlation among the variables, the Pearson correlation analysis was performed using SPSS software package version 21.0, Chicago, IL, USA. The healthy, prediabetes and diabetic groups were classified using WEKA data mining tool (Waikato Environment for Knowledge Analysis), version 3.8.2, New Zealand.

### 3 Results

In the prediabetes group, it was observed that the positive correlation was exhibited between HbA<sub>1c</sub> (%) and measured skin surface temperature at the toe region ( $r$  is 0.917,  $p < 0.01$ ) and the negative association was found with metatarsal 1 ( $r$  is  $-0.474$ ,  $p < 0.05$ ), metatarsal 3 and heel ( $r$  values are  $-0.895$  and  $0.901$ ,  $p < 0.01$ ) regions. From Fig. 2, measured temperature difference between the left feet of healthy versus prediabetic was found to be  $0.7^\circ\text{C}$  and prediabetic versus diabetic



**Fig. 2** Measured temperature at the left and right plantar region. **a** Sample healthy subject, **b** sample prediabetic subject, **c** sample type 2 diabetes mellitus subject

was found to be 1.4 °C. Also, the right foot of healthy and prediabetic obtained the temperature difference as 0.9 °C. The prediabetic and diabetic right foot attained the difference in temperature as 1.2 °C. The right and left healthy versus diabetic foot obtained difference in temperature as 2.1 °C, respectively. Figure 2a–c show the plantar thermograms of healthy, prediabetic and type 2 DM sample subject.

Among the groups, the age and blood pressure parameter showed a significant correlation ( $p < 0.01$ ) between the healthy (group I) and type 2 DM (group III). Also, the measured plantar temperature at metatarsal 1 and heel regions exhibited a significant correlation between all the three groups. Whereas, the waist circumference and core body temperature do not exhibit any significant differences are tabulated in Table 1. Figure 3 represents the measured temperature at various ROIs in left and right plantar thermograms. Figure 3a represents the biothesiometry template to measure the plantar temperature at different ROIs. Figure 3b–d portrays the ROIs of left and right plantar regions of healthy, prediabetic and type 2 DM sample subjects. The ROIs numbered 1 and 7 indicate the toe region, 2 and 8 indicate the metatarsal 1 region, 3 and 9 indicate the metatarsal 3 region, 4 and 10 indicate the metatarsal 5 region, 5 and 11 indicate the instep region, and 6 and 12 indicate the heel regions of sample left and right foot.

The correlation matrix for prediabetes ( $n = 20$ ) is tabulated in Table 2. The HbA<sub>1c</sub> (%) exhibited positive correlation with measured temperature at the toe region ( $r$  value is 0.917,  $p < 0.01$ ) and the negative correlation with measured temperature at metatarsal 1 ( $r$  value is  $-0.474$ ,  $p < 0.05$ ), metatarsal 3 and heel regions ( $r$  values are  $-0.895$  and  $-0.901$ ,  $p < 0.01$ ). Figure 4 portrays the area under receiver operating characteristic plot of true positive rate against the false positive rate of various classifiers. Figure 4a–e is ROC of SVM, NB, MLP, k-NN and RF classifiers. The obtained area under the curve for SVM, NB, MLP, k-NN and RF classifiers was 0.8812, 0.84, 0.8825, 0.75 and 0.8706, respectively. The SVM classifier obtained the classification accuracy rate as 81.6% and outperformed the other classifiers, such that NB yielded 76.6%, MLP achieved 73.3%, k-NN attained 66.6%, and RF produced 71.6% as tabulated in Table 3.

## 4 Discussions

The main purpose of this pilot study was to evaluate the capability of thermography in various plantar sites for screening prediabetes individuals. Chatchawan et al. [20] observed the decreased average plantar temperature at the right side was found as 1.87 °C between healthy and DM subjects. And the left side average temperature difference between the healthy and DM subjects was found to be 2.65 °C. Madaras-ingha et al. [21] has developed the system to measure the foot temperatures and found greater than or equal to 2.2 °C temperature difference in plantar regions between the normal and diabetic groups. In our proposed study, the moderate difference in measured temperature at right side plantar was found to be 0.7 °C and left side was 0.9 °C among healthy and prediabetes groups. The higher temperature difference at both

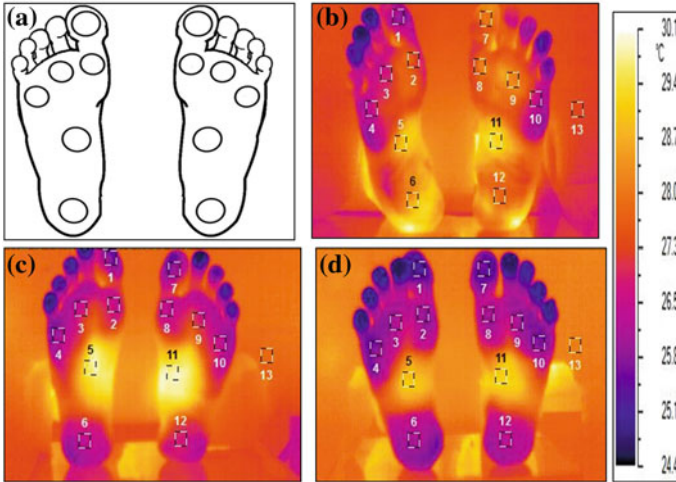
**Table 1** Characteristics and comparisons of the study groups

Parameters	Group I ( <i>N</i> = 20)	Group II ( <i>N</i> = 20)	Group III ( <i>N</i> = 20)	<i>p</i> -value		
				I versus II	II versus III	I versus III
Age (years)	30.55 ± 6.62	33.35 ± 6.26	35.5 ± 6.95	NS	NS	S
BMI (kg/m <sup>2</sup> )	24.46 ± 4.76	28.18 ± 3.48	28.58 ± 5.08	S	NS	S
Waist circumference (cm)	91.8 ± 8.71	93.55 ± 10.74	94 ± 9.21	NS	NS	NS
Hip circumference (cm)	96.85 ± 12.22	103.15 ± 9.63	103.95 ± 9.00	NS	NS	S
Systolic blood pressure (mm/Hg)	117.5 ± 11.64	134.5 ± 14.68	131.5 ± 15.65	S	NS	S
Diastolic blood pressure (mm/Hg)	76.8 ± 7.26	85 ± 8.27	86.8 ± 10.98	S	NS	S
HbA <sub>1c</sub> (%)	5.34 ± 0.18	6.06 ± 0.19	8.67 ± 1.76	S	S	S
Core body temperature (°C)	36.91 ± 0.09	36.84 ± 0.22	36.90 ± 0.08	NS	NS	NS
Toe (°C)	27.35 ± 0.39	27.77 ± 0.30	27.73 ± 0.32	S	NS	S
Metatarsal 1 (°C)	28.38 ± 0.37	27.62 ± 0.35	27.22 ± 0.23	S	S	S
Metatarsal 3 (°C)	28.80 ± 0.37	28.66 ± 0.21	27.38 ± 0.37	NS	S	S
Metatarsal 5 (°C)	28.70 ± 0.17	28.46 ± 0.55	26.52 ± 4.47	NS	NS	S
Instep (°C)	30.65 ± 1.55	31.09 ± 2.00	30.27 ± 1.55	NS	NS	NS
Heel (°C)	29.62 ± 0.36	28.40 ± 0.22	27.21 ± 0.33	S	S	NS

*S* Significant (*p* < 0.01); *NS* Non Significant

left and right foot temperature was found to be 2.1 °C between the healthy and type 2 DM groups. The abatement in the plantar temperature at the right and left side among the groups could reflect the poor blood supply in the lower limbs, probably due to the stiffness in the posterior tibial artery. This might lead to atherosclerosis for prediabetes and diabetes group in future when there is an increase in the HbA<sub>1c</sub> (%) values.

Smieja et al. [22] discussed about the possible sites in the plantar region for distinguishing the diabetic foot was the great toe, first, third and fifth metatarsals which are observed to be insensate during the monofilament test. Likewise, in our study, the monofilament test has been performed, and it was observed that healthy



**Fig. 3** Measured temperature at various region of interest in plantar thermogram. **a** Biothesiometry study template of various ROIs, **b** sample healthy subject, **c** sample prediabetic subject and **d** sample type 2 diabetes mellitus subject

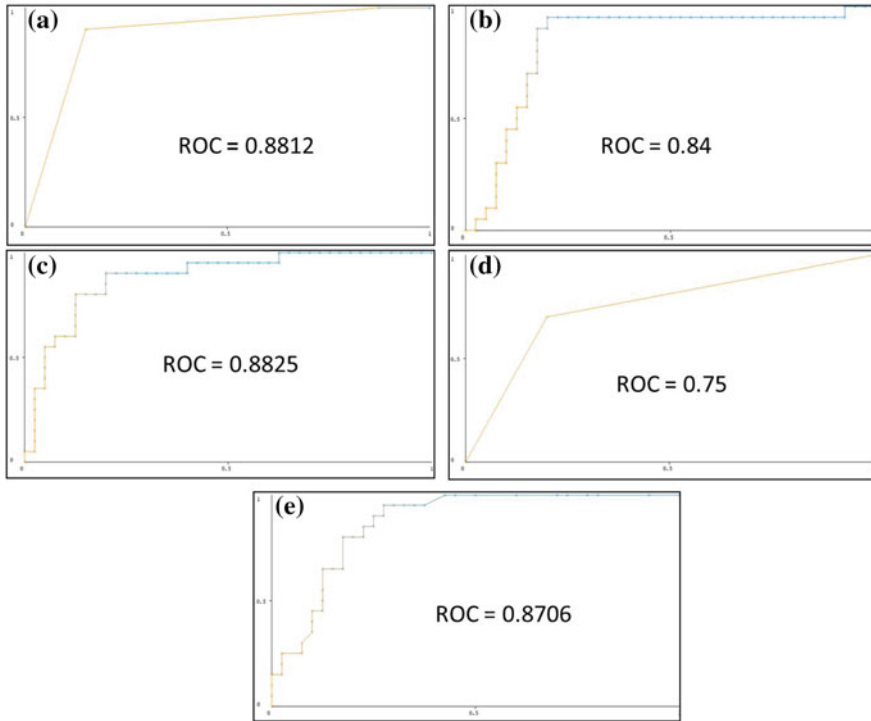
**Table 2** Correlation matrix for prediabetes ( $n = 20$ )

Plantar sites (°C)	HbA <sub>1c</sub> (%)
Toe region	0.917**
Metatarsal 1 region	-0.474*
Metatarsal 3 region	-0.895**
Metatarsal 5 region	-0.192
Instep region	0.295
Heel region	-0.901**

\*\* $p < 0.01$  (significant correlation); \* $p < 0.05$  (significant correlation)

subjects have greater sensation in their feet compared to diabetic subjects. Also, the moderate sensation has been observed in the plantar sites of prediabetes subjects. And, the type 2 DM groups are found to have insensate feet. The loss of sensation in the type 2 DM group could lead to the foot ulcer, and lower sensation in prediabetes group might lead to the progress of type 2 DM in future.

Bagavathiappan et al. [23] have observed the higher temperature and loss of sensation in the great toe region (32–35 °C) of the plantar surface in diabetic neuropathy patients compared with the subjects without neuropathy (27–30 °C). The total loss of sensation in the subjects can incline the patient to foot ulceration in the great toe region. In our study, the instep region does not exhibit any significant correlation among the groups, and also, the measured plantar temperature at instep region was found to be higher between the groups indicates that the instep region might not be



**Fig. 4** Area under receiver operating characteristic plot of various classifiers. **a** Support vector machine, **b** Naïve Bayes, **c** multilayer perceptron, **d** K-nearest neighbour and **e** random forest

**Table 3** Performances of the different classifiers

Classifiers	ROC area	Sensitivity %	Specificity %	Accuracy %
Support vector machine (SVM)	0.8812	85.66	91.6	81.6
Naïve Bayes (NB)	0.84	82	88	76.6
Multilayer perceptron (MLP)	0.8825	79.5	97	73.3
k-nearest neighbour (k-NN)	0.75	74	87	66.6
Random forest (RF)	0.8706	76	89.6	71.66

prone to a diabetic foot ulcer. The increase in temperature at the instep region could be due to the proper blood supply without any stiffness in the tibial artery. So, the most prone region for diabetic foot ulcer could be the great toes and the balls of the foot (metatarsals) and heel region due to the increased glucose levels in the blood. There is a decrease in temperature because of the poor metabolism and blood supply in the plantar sites. The decreased mean plantar temperature in prediabetes group could lead to diabetic neuropathy problems in future.

An abrupt change in the temperature at specific plantar sites of diabetic subjects was observed by the research study conducted by Renero [24]. The temperature differences in the plantar site might correspond to the ulceration or inflammation complications for the diabetic subjects in the particular foot region. In our present study, it was observed that among the diabetic group, the decreased temperature was found at metatarsal 5, and increased temperature was found in the instep region. This lower and higher temperature differences in the diabetic group indicate the development of ulceration or ischemia in increased temperature regions in future.

Choi et al. [25] have developed the machine learning algorithms such as an artificial neural network (ANN) and SVM for pre-screening the prediabetes from the datasets provided by the Korean National Health and Nutrition Examination Survey. The SVM and ANN model have obtained 0.731, 0.729 as the area under the ROC curve and attained the classification accuracy rate as 66.1 and 60.7%. Likewise, in our study, we have used five different classifiers, the SVM classifier has been trained with sequential minimization algorithm (SMO) has outperformed the other classifiers with an accuracy rate as 81.6%, and area under the curve was found to be 0.8812, respectively. The strength of infrared imaging is safe, risk-free, no harmful radiations, fast processing and non-invasive technique which can be used repeatedly. The limitations of the study are that thermal infrared camera is expensive than other cameras, and the sample size for this study was too small.

## 5 Conclusion

The mean average temperature of various ROIs in the plantar region for the study population was measured using GRAYESS thermal software. The plantar thermograms were acquired from the six plantar sites such as toe, metatarsal 1, metatarsal 3, metatarsal 5, instep and heel regions. The HbA<sub>1c</sub> (%) was correlated positively with the toe region ( $r$  value is 0.917,  $p < 0.01$ ) and negatively with metatarsal 1 ( $r$  value is  $-0.474$ ,  $p < 0.05$ ), metatarsal 5 and heel regions ( $r = -0.895$  and  $-0.901$ ,  $p < 0.01$ ). The SVM classifier has outperformed the other classifiers with good accuracy rate as 81.6%, sensitivity as 85.66%, specificity as 91.6%, and area under ROC curve was 0.8812, respectively. The findings from this study show that screening is more important in identifying the prediabetes earlier and enabling prior mediations. The pre-screening of prediabetes helps to reduce the risk of developing the type 2 DM and ulcerations in future.

## References

1. Colagiuri S (2011) Epidemiology of prediabetes. *Med Clin North Am* 95:299–307
2. Chen L, Magliano DJ, Zimmet PZ (2011) The worldwide epidemiology of type 2 diabetes mellitus—present and future perspectives. *Nat Rev Endocrinol* 8:228–236

3. American Diabetes Association (2018) Classification and diagnosis of diabetes: standards of medical care in diabetes—2018. *Diab Care* 41:S13–S27
4. Li G, Zhang P, Wang J, Gregg EW, Yang W, Gong Q, Li H, Jiang Y, An Y, Shuai Y, Zhang B, Zhang J, Thompson TJ, Gerzoff RB, Roglic G, Hu Y, Bennett PH (2008) The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing diabetes prevention study: a 20-year follow-up study. *Lancet* 24:1783–1789
5. Anjana RM, Deepa M, Pradeepa R, Mahanta J, Narain K, Das HK, Adhikari P, Rao PV, Saboo B, Kumar A, Bhansali A, John M, Luaia R, Reang T, Ningombam S, Jampa L, Budnah RO, Elangovan N, Subashini R, Venkatesan U, Unnikrishnan R, Das AK, Madhu SV, Ali MK, Pandey A, Dhaliwal RS, Kaur T, Swaminathan S, Mohan V (2017) Prevalence of diabetes and prediabetes in 15 states of India: results from the ICMR–INDIAB population-based cross-sectional study. *Lancet Diab Endocrinol* 5:585–596
6. Anjana RM, Pradeepa R, Deepa M, Datta M, Sudha V, Unnikrishnan R, Bhansali A, Joshi SR, Joshi PP, Yajnik CS, Dhandhanika VK, Nath LM, Das AK, Rao PV, Madhu SV, Shukla DK, Kaur T, Priya M, Nirmal E, Parvathi SJ, Subhashini S, Subashini R, Ali MK, Mohan V (2011) Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: phase I results of the Indian Council of Medical Research–INDIA DIABetes (ICMR–INDIAB) study. *Diabetologia* 54:3022–3027
7. World Health Organization and International Diabetes Federation (2006) Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia. Report of a WHO/IDF consultation 1–50
8. American Diabetes Association (2013) What is prediabetes? Understanding the warning signs—and how to stay healthy. *Clin Diab* 31:95
9. Yudkin JS (2016) “Prediabetes”: are there problems with this label? Yes, the label creates further problems! *Diab Care* 39:1468–1471
10. Bansal N (2015) Prediabetes diagnosis and treatment: a review. *World J Diab* 6:296–303
11. American Diabetes Association (2019) Classification and diagnosis of diabetes: standards of medical care in diabetes—2019. *Diab Care* 42:S13–S28
12. American Diabetes Association (2016) Classification and diagnosis of diabetes. *Diab Care* 39:S13–S22
13. Viswanathan V (2010) Epidemiology of diabetic foot and management of foot problems in India. *Int J Low Extremity Wounds* 9:122–126
14. Rayman G, Vas PR, Baker N, Taylor CG, Gooday C, Alder AI, Donohoe M (2011) The Ipswich touch test: a simple and novel method to identify in patients with diabetes at risk of foot ulceration. *Diab Care* 34:1517–1518
15. Baraz S, Zarea K, Shahbazian HB, Latifi SM (2014) Comparison of the accuracy of monofilament testing at various points of feet in peripheral diabetic neuropathy screening. *J Diab Metab Disord* 13:19
16. Lal C, Unni SN (2015) Correlation analysis of laser doppler flowmetry signals: a potential non-invasive tool to assess microcirculatory changes in diabetes mellitus. *Med Biol Eng Comput* 53:557–566
17. Viswanathan V, Snehalatha C, Seena R, Ramachandran A (2002) Early recognition of diabetic neuropathy: evaluation of a simple outpatient procedure using thermal perception. *Postgrad Med J* 78:541–542
18. Bharara M, Cobb JE, Claremont DJ (2006) Thermography and thermometry in the assessment of diabetic neuropathic foot: a case for furthering the role of thermal techniques. *Int J Low Extremity Wounds* 5:250–260
19. Nathan DM (2009) International expert committee report on the role of the A1C assay in the diagnosis of diabetes. *Diab Care* 32:1327–1334
20. Chatchawa U, Narkto P, Damri T, Yamauchi J (2018) An exploration of the relationship between foot skin temperature and blood flow in type 2 diabetes mellitus patients: a cross-sectional study. *J Phys Ther Sci* 30:1359–1363
21. Madarasingha KCM, Perera WML, Rathnayaka AJD, Shanuka HPS, Jayasinghe S, Kahaduwa KTD, Silva ACD (2018) Development of a system to profile foot temperatures of the plantar and



- the periphery. In: TENCON 2018—2018 IEEE region 10th conference, Jeju, Korea (South), pp 1928–1932
22. Smieja M, Hunt DL, Edelman D, Etechells E, Cornuz J, Simel DL, International Cooperative Group for Clinical Examination Research (1999) Clinical examination for the detection of protective sensation in the feet of diabetic patients. *J Gen Intern Med* 14:418–424
  23. Bagavathiappan S, Philip J, Jayakumar T, Raj B, Rao PNS, Varalakshmi M, Mohan V (2010) Correlation between plantar foot temperature and diabetic neuropathy: a case study by using an infrared thermal imaging technique. *J Diab Sci Technol* 4:1386–1392
  24. Renero CJF (2018) The abrupt temperature changes in the plantar skin thermogram of the diabetic patient: looking into prevent the insidious ulcers. *Diab Foot Ankle* 1430950
  25. Choi SB, Kim WJ, Yoo TK, Park JS, Chung JW, Lee YH, Kang ES, Kim DW (2014) Screening for prediabetes using machine learning models. *Comput Math Methods Med* 618976

# Fog Computing using Interoperability and IoT Security Issues in Health Care



P. Karthika, R. Ganesh Babu and P. A. Karthik

**Abstract** Nowadays, our day-to-day life starts with smart devices those are fall under the Internet of thing gadgets and these gadgets are becoming a part of our lifestyle of which the health care becomes the most crucial area. Researchers provide a hint of technology, by making health care it through Fog computing in order to make the information transfer through IoT devices simpler. Interoperability and security are particularly major issues raised by such impediments. In this paper, talk about the present issues, including advantages and challenges, ways to deal with, avoid the issues by utilizing, and incorporating gadgets in medicinal services frameworks. Here, present the talk with regards to the Fog networking venture, which focuses on an answer for home consideration, monitoring the patients with ceaseless ailments.

**Keywords** Internet of things · Security · Health care security · Fog networking

## 1 Introduction

Internet of things (IoT) includes a lot of advancements that empower a large scope of apparatuses, gadgets, and objects to associate and convey information surrounded by. People supply the majority of the substance and data found on the Internet by

---

P. Karthika

Department of Computer Applications, Kalasalingam Academy of Research and Education,  
Krishnan Koil, TamilNadu, India  
e-mail: [karthikasivamr@gmail.com](mailto:karthikasivamr@gmail.com)

R. Ganesh Babu (✉)

Department of Computer Science and Engineering, Kalasalingam Academy of Research and  
Education, Krishnan Koil, TamilNadu, India  
e-mail: [ganeshbaburajendran@gmail.com](mailto:ganeshbaburajendran@gmail.com)

P. A. Karthik

Department of Electronics and Communication Engineering, Rajalakshmi Institute of Technology,  
Chennai, TamilNadu, India  
e-mail: [csekarthik7@gmail.com](mailto:csekarthik7@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_10](https://doi.org/10.1007/978-981-15-2329-8_10)

connecting through such devices. There are numerous applications for it, including human services frameworks, which are the principal focal point of this paper. Medicinal services frameworks utilize a lot of interconnected gadgets to make an IoT system available to human services including checking patients and naturally distinguishing circumstances where therapeutic mediations are required. To address the difficulties in verifying IoT gadgets, propose utilizing AI inside an IoT passage to help to secure the framework. AI is a zone of artificial knowledge in which PC programs are empowered to gain as a matter of fact, models, and analogies [1]. As learning happens, the abilities inside the program become progressively clever and the program winds up fit for making educated choices. Inside AI, two of the most famous methodologies are counterfeit neural systems (ANN) and hereditary calculations. ANNs emulate the neurons and neurotransmitters inside the mind to exchange information for correspondence, learning, what's more, basic leadership [2]. ANNs are utilized inside IoT frameworks to screen the territory of IoT gadgets and to make educated choices [3]. In this paper, propose the utilization of ANN to get familiar with the sound condition of a framework and associated gadgets. The rest of this paper is composed as pursues. Area gives a review of the territory of IoT security what's more, the utilization of AI inside security. The portrays our way to deal with calculator learning inside an IoT passage. Segment discusses our experimentation what's more, results incorporating achievement and disappointments in calculator learning inside the gateway.

It is commonly perceived that the constantly sick, for example, individuals with heart disappointment, hypertension, respiratory maladies or diabetes involve restorative, clinic, and crisis benefit more frequently than customary patients [4] in Fig. 1. Data and correspondence advances are among the instruments that could help relieve a portion of the issues related to maturing populaces, expanded rates of constant ailments, and lack of wellbeing experts, and, in the meantime, encourage administration rearrangement. Present-day estimating gadgets, for example, circulating strain, weight, and development sensors, join correspondence abilities. They can make it systems executed for home telemonitoring. These equivalent gadgets are generally making use of by wellbeing laborers to verify the common state of patients with ceaseless diseases.

Fog network is a task which targets home answers for consideration/telemonitoring of the patients with perpetual sicknesses. Various systems and conventions for general information trade among observing gadgets like circulatory strain screens and development sensors are considered. The venture likewise includes the plan and execution of a social insurance gave middleware. The common framework group of data from different detecting gadgets throughout a middleware that provides interoperability and security essential with regards to the Internet of things for medical services. Observing gadgets are associated by means of remote systems administration innovation.

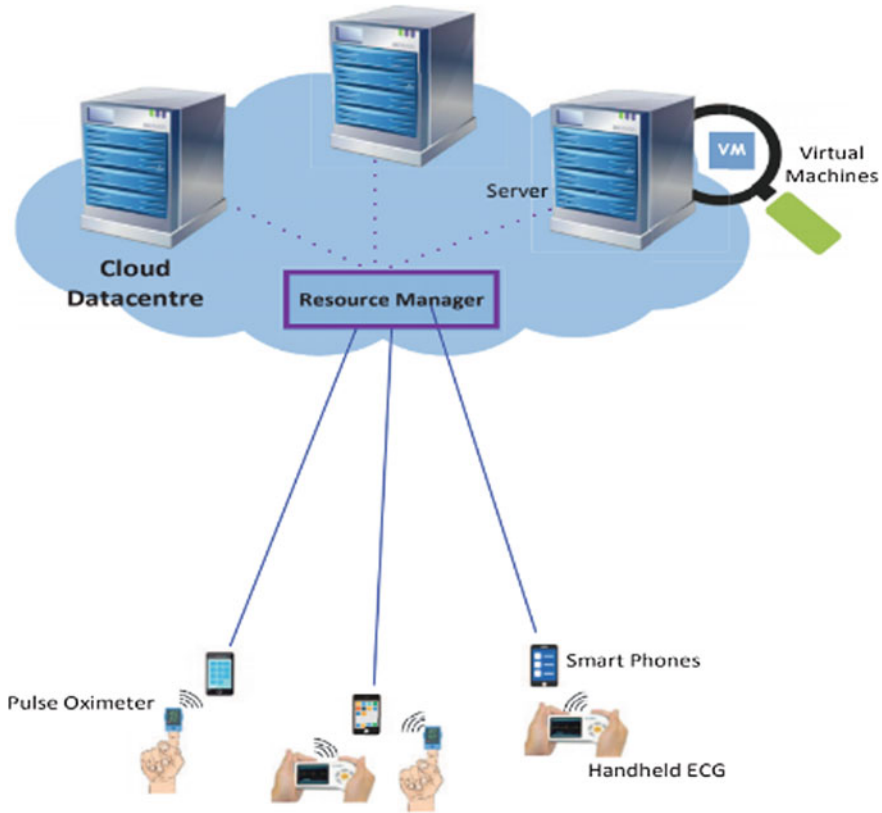
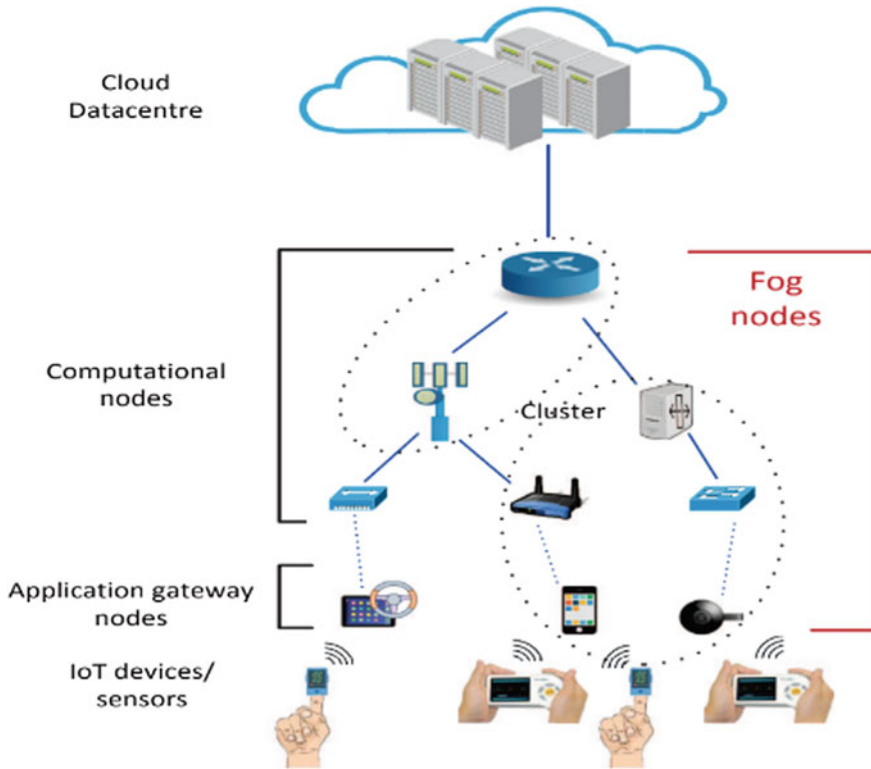


Fig. 1 IoT-based health care system architecture being connected for gadgets

## 2 Devices to Be Connected

In light of necessities of a residence telemonitoring framework for the patients with perpetual illnesses (e.g., circulatory strain, tolerant recuperation of hospitalization), a few gadgets have been chosen (Fig. 2) to be utilized in the Fog network, so as to screen a few parts of the patients well-being. This gadget has Wi-Fi interfaces and highlights that empower interoperability and information transmission. Gadgets chose to use in the task:

- A camera to be recognizing the Panasonic BL-C230A Wi-Fi IP, the developments of observed patients—The absence of development in times of the developments is normal may demonstrate in various basic issues.
- They chose gadget could distinguish developments dependent on three unique wellsprings of information: i. Picture, ii. Sounds, and iii. Body heat.



**Fig. 2** Fog network-based health care telemonitoring devices being connected for IoT sensors

- Each development distinguished sends forward an alarm communication that conveys the pictures which are to be sent to an email address or transferred to FTP (File Transport Protocol (FTP) server).
- A remote body scale fixed with a Wi-Fi interface fabricated by integral with programming which ascertains the patients' level of large, bulky, and weight record.
- The ordinary arrangement of these gadget interfaces the remote system following gauging and send the information to a URL inside to locate the producer's site.
- A circulatory strain gadget with the eccentricity that its activity relies upon the association with an Apple gadget (iPhone or iPod Touch).
- The task is like the level with the distinction that this gadget is working through a request introduced in an Apple gadget.

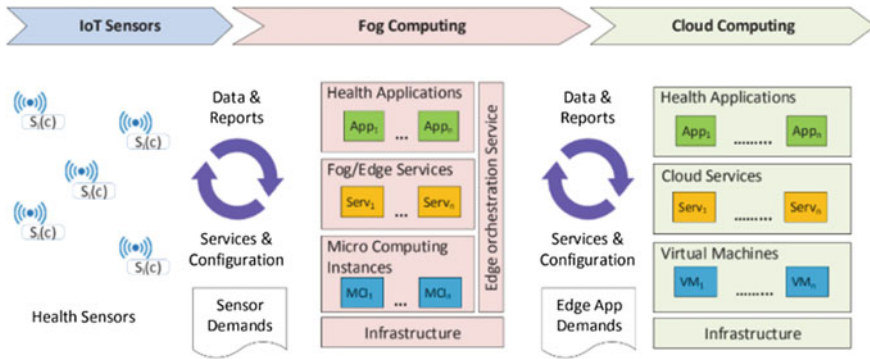


Fig. 3 Architecture of Fog-network integration for interoperable IoT-health care solutions

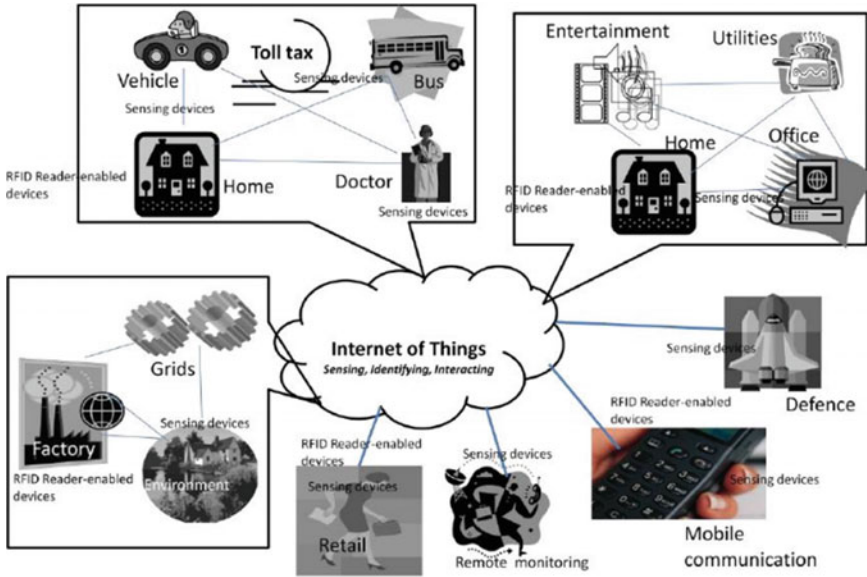
### 3 Interoperability Issues

The Fog-network task was intended to convey an observing situation dependent on Wi-Fi (802.11). First expected that different checking gadgets would most likely speak with remote servers. This methodology, appropriate, has limited the decent variety of medicinal gadgets available, chiefly as a result of propensity to utilize Bluetooth originality in social insurance gadgets. This prototype is the smallest effort and utilize for little power of Bluetooth-empowered gadgets. In spite of real accessibility, it has just discovered a couple of arrangements in the promoter, which empower the Bluetooth (802.14)/ZigBee (802.15.4) systems and Internet connection to avoid the interoperability issues.

In Fig. 3, the interoperability issue may be settled through the Wi-Fi/Bluetooth passage. It is giving all administration abilities wanted, also by utilizing the Simple Network Management Protocol (SNMP) or net administrations. It may, even human services, gadgets worked with Wi-Fi interfaces entirely reasonable to utilize in a flexible observing condition. It establishes that the greater part of the gadgets work by speaking servers and restrictive framework, presently like the illustration of the body scale.

### 4 Security Issues

The way of individual secret information will be gathered through telemonitoring assume the requirement of the systems, its components for the security protection. As featured in [5], having each ‘thing’ associated, new security and protection issues emerge, e.g., secrecy, genuineness, and honesty of information detected and traded by ‘things.’ This creator records the standard security prerequisites shown in Fig. 4.



**Fig. 4** Integration of security IoT-health care varying number of sensors

- Flexibility to assaults—The framework requirements to keep away from the particular purposes of disappointment and ought to change itself to hub disappointments;
- Data confirmation—As a standard, improved locations and article data should be validated;
- Access control—Information suppliers must almost certainly actualize access control on the information gave;
- Client protection—Measures must be taken that solitary the data supplier can construe from watching the utilization of the query framework.

The arranged verification administrations utilize Shibboleth, created for the middleware layer for confirmation and access control an Internet2 venture [6]. By utilizing Shibboleth, it is conceivable to exploit an associate authentication administration officially embedded by the RNP. An organization of the Brazilian research and the training foundations going about as personality and specialist. That is substantial for all administrations offered by the organization, in this manner, wiping out of the required different passwords and its various enrollment forms. In case the IoT citing, further security issues emerge to information uprightness and client protection, while all phones are potentially focus on noxious assaults, which require further headways as far studies and security countermeasures. On medicinal services condition, keep-

ing gone from effective assaults to the framework. If nothing improves them are necessary objectives as disappointments or data outlet can speak to harms to the patients. Rather than different spaces can retain for few expenses of framework misuse, social insurance frameworks can't. When touchy about data a person's medical issue is revealed and public harm is finished, it is difficult to repudiate the data [7].

An assailant has all around characterized objectives when concentrating on cell phones. For the most part, assaults go for taking clients data, assault gadgets, or notwithstanding closing behind certain applications [8]. Numerous dangers are surrounding cell phones; most of the dangers are obtain from traditional figuring frameworks and familiar for phones. Nonetheless, a small amount of dangers get further consideration in view of the possible issues that can cause to the frameworks. Instances of these strings are i. Man-in-the-middle; ii. Routing preoccupation attack; and iii. Rejections of administration expect to cause battery exhaustion in the gadget. The assaults are connected only on cell phones, speed up known factors, for example,

- Distinguishing utilization of communicating for correspondence;
- Deficiency of confirmation sources;
- Deployment of batteries as a power source;
- Versatility. Consequently, to supervise such a huge scope of the cell phones, map and alleviate the majority of important strings in a known setting is a basic assignment.

The observing application for which is practically like a system chief. In charge of putting away, conglomerating, merging, and contrasting the gathered data against chronicled arrangement. The goal limits are crossed the front-end framework issue alarms or executed explicit strategies. If the point of breaking is crossed, upon the contingent edge strategies, alerts are activated to empower wellbeing laborers to quickly reply to wellbeing-connected occasions.

A wellbeing laborer ought to have the option to get to information as of a patient utilizing his/her cell phone though the patient's home visit. Be that as it may, get to is not allowed from different spots or during times of inertia explicitly connected to the patient so as to ensure understanding protection and considering moral contemplation as examined in 2013 to 2022. These techniques plan to reply to security necessities of information verification, get to control, and customer protection [10].

The social insurance application keeps running on a remote server and oversees telemonitoring applying channels. To avert gathering of information from some wellbeing checking gadget that might be as yet dynamic of the subsequent period. Gadgets were moved from a house to somewhere without appropriate enlistment can send information to the framework, however, this information ought not be considered on the off chance that they never again connect with a patient under is observing [11]. Numerous gadgets that include the scene of Internet of things do not have a one of a kind, distinguishing proof and approval of, information originating from these gadgets must be made by applying suitable principles explicit to every unique situation.



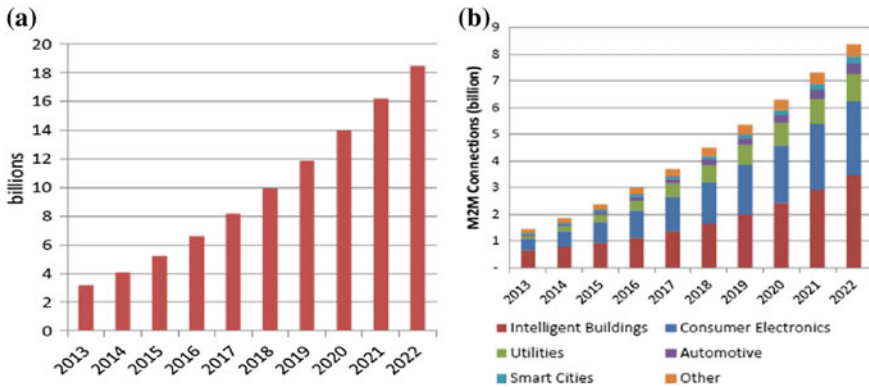


Fig. 5 a Network connections, b machine-to-machine connections in 2013–2022 [9]

In Fig. 5b, IoT Fog network connection machine-to-machine connection in 2013–2022. The outcomes creature produced are talking about the flow passages with their showing. The underlying advance in the usage was to catch the casements that are available in each framework comprising of the data about every single occasion happening on the framework [12]. They are fundamentally of three sort’s applications of the framework and security. Investigation is performed on application security, and who work as pursues: Application catch those occasions which are the application explicit and occasions are delegated mistake, cautioning whats more and data. Then again the security logs identified with the legions of clients onto the windows were effective or not.

## 5 Conclusion

This paper introduced a review of the extraordinary AI calculations used to verify the IoT. The diverse AI approaches are used to verify IoT and ML. While numerous approaches result in high exactness, SVM is less mind boggling than KNN in order. There are as yet numerous difficulties in verifying IoT and ML in light of the fact that AI ought to bargain between an abnormal state of security and a low computational multifaceted nature to be appropriate for the resource limited IoT and ML gadgets. The future work to breaking down biometric-based security systems (BSS) that apply increasingly complex classifiers such as help vector machines or profound neural systems. This methodology can likewise help us to plan new sorts of highlight extraction calculations or classifiers for improving execution and security quality together.

## References

1. Babar S, Mahalle P, Stango A, Prasad N, Prasad R (2010) Proposed security model and threat taxonomy for the internet of things (IoT). In: Third international conference on network security and applications, Chennai, India, pp 420–429
2. Ganesh Babu R, Amudha V (2016) Resource allocation in QoS scheduling for IEEE 802.16 systems. *Int J Sci Innovative Eng Technol* 1(5):50–55
3. Sedlmayr M, Prokosch H, Munch U (2011) Towards smart environments using smart objects. *Stud Health Technol Informatics* 169:315–319
4. Ganesh Babu R, Amudha V (2016) Dynamic spectrum access techniques in cognitive radio networks. *Int J Emerg Technol Comput Sci Electron* 22(2):508–512
5. Karthika P, Vidhya Saraswathi P (2017) A survey of content based video copy detection using big data. *Int J Sci Res Sci Technol* 3(5):114–118
6. Ganesh Babu R, Amudha V (2019) Distributed cooperative ai techniques for cognitive radio networks. *Int J Recent Technol Eng* 8(1S2):198–203
7. Riedel T, Fantana N, Genaid A, Yordanov D, Schmidtke HR, Beigl M (2010) Using web service gateways and code generation for sustainable IoT system development. In: IEEE international conference on internet of things (IoT), Tokyo, Japan, pp 1–8
8. Karthika P, Vidhya Saraswathi P (2019) Digital video copy detection using steganography frame based fusion techniques. In: International conference on ISMAC in computational vision and bio-engineering, Springer lecture notes in computational vision and biomechanics, Palladam, Tamilnadu, India, pp 61–68
9. Chavan SS, Madanagopal R (2009) Generic SNMP proxy agent framework for management of heterogeneous network elements. In: Proceedings of the first international conference on communication systems and networks (COMSNETS'09), Bangalore, India, pp 1–6
10. Ganesh Babu R, Amudha V (2018) Comparative analysis of distributive firefly optimized spectrum sensing clustering techniques in cognitive radio networks. *J Adv Res Dyn Control Syst* 10(9):1364–1373
11. Salceda J, Diaz I, Tourino J, Doallo R (2004) A middleware architecture for distributed systems management. *J Parallel Distrib Comput* 64(6):759–766
12. Karthika P, Ganesh Babu R, Nedumaran A (2019) Machine learning security allocation in IoT. In: IEEE international conference on intelligent computing and control systems, Madurai, India, in press

# Comparison of Manual and Semi-automated Method in Measurement of Joint Space Width Measurement in Feet Region of RA Patients



Meghna Sampath, Snehalatha Umopathy , Sakshi Srivastava and Nelufer Shamsudim

**Abstract** Rheumatoid arthritis (RA) is a systemic life-threatening disorder and affects the body's self immune system and invades the smaller and larger joints. There are different modern imaging techniques used such as magnetic resonance imaging (MRI), positron emission tomography (PET), single photon emission computed tomography (SPECT) and colour Doppler ultrasound were used to evaluate the inflammatory conditions in RA. Still radiography (X-ray) was investigated as the best method in diagnosis of RA. X-ray imaging has been the easiest and most used method of diagnosis of any abnormalities in the hard tissues of the human body. The study aimed as follows (i) to perform a manual and semi-automated segmentation of feet region in the total population studied. (ii) To compute the features extracted from the joint space width, joint space narrowing and erosion in the feet. The analysis of X-ray images was performed using manual method (DICOM viewer) and semi-automated method using Mimics software. The SPSS statistical analysis software was used to calculate mean standard deviation and standard deviation error and further plotted accordingly. There was a significant differences found between the manual and semi-automated methods. Furthermore, it was inferred that the semi-automated method showed more precise readings of the joint space narrowing in RA and normal feet and should be used more extensively for further study.

**Keywords** Rheumatoid arthritis · Joint space width · Semi-automated method · X-ray

## 1 Introduction

Rheumatoid arthritis (RA) is a systemic life-threatening disorder and affects the body's self immune system and invades the smaller and larger joints. This creates the inflammation that leads to joint thickness, swelling and joint pains in smaller

---

M. Sampath · S. Umopathy (✉) · S. Srivastava · N. Shamsudim  
Department of Biomedical Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Chennai 603203, India  
e-mail: [sneha\\_samura@yahoo.co.in](mailto:sneha_samura@yahoo.co.in)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_11](https://doi.org/10.1007/978-981-15-2329-8_11)

and larger joints. The joint spacing between bones can become smaller, and the joint effect is usually symmetrical in RA. In India, about 0.92% population is affected by RA [1]. Each year about 20–40 new patients per lakh population were affected, and the females are likely to get affected frequently than males [2]. The prevalence rate is approximately 1–2% in western part of the World [3]. The developing countries have the prevalence rate estimated to be 1% Worldwide [4].

The different types of diagnostic methods are physical exam, blood tests and radiographic imaging [5]. Foot radiograph is the most extensively used to visualize the damages in small MTP joints and bone erosion in RA patients [6]. The parameters such as joint space narrowing width and joint erosion help to evaluate the condition of RA better and give more precise values for every subject.

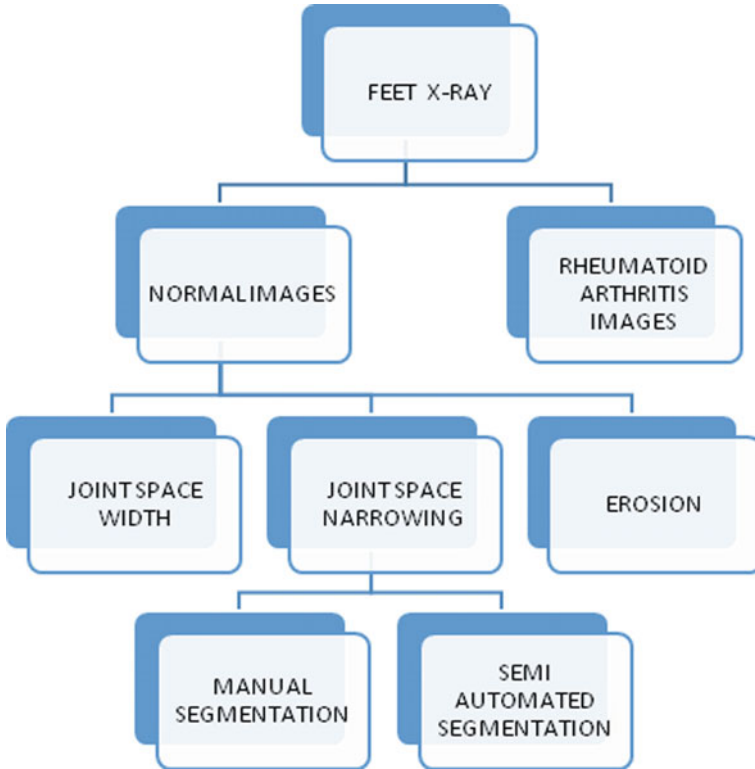
Several researchers have conducted the study in hand radiograph of RA subjects using semi-automated and automated methods [7–10]. Snehalatha et al. [11] performed an automated hand X-ray segmentation using dual-tree complex wavelet transform (DTCWT)-based watershed algorithm. They extracted the GLCM features from the joint space region of the hand for the entire population [11]. Bojny et al. [12] used X-ray scoring methods such as Sharp, Larsen and van der Heijde/Sharp as an established manual scoring method for the estimation of disease progression in RA [12]. Heidari [13] predicted the prognosis of RA based on the clinical and laboratory findings. The new classification criteria of RA provide a scope for prior treatment [13].

The study aimed as follows: (i) to implement a manual and semi-automated segmentation of feet region in RA patients and normal subjects; (ii) to correlate the GLCM features of the joint space width, joint space narrowing and erosion in the feet for both normal and RA.

## 2 Materials and Methodology

### 2.1 Study Population and Measurements

This study was conducted by using the data collected from the Agarwal orthopaedic hospital, Uttar Pradesh. The patients who had persistent inflammation, joint pain in smaller joints, swelling in smaller and larger joints, arthritis and deformation in the feet were considered as inclusion criteria for RA. The patients who had joint surgery in the feet region or previously had a bone fracture were excluded from the study. The study populations were categorized as follows namely: group I: Normal subjects and group II: RA patients.



**Fig. 1** Block diagram of methodology of the proposed work

A standard posterior–anterior view of feet radiograph was obtained in each subject. The geometric measurements for the joint space width were implemented using two different approaches such as manual measurement using DICOM software and semi-automated method segmentation and analysis using Mimics software. The overall methodology was illustrated in Fig. 1.

## 2.2 Manual Measurements

The images (normal and rheumatoid) were taken in the DICOM format, and using a DICOM viewer, the distance between the metatarsal and interphalangeal joints was measured manually using the measurement tool available. This operation was done for a set of 12 ft (each of RA and normal feet) approximately, and the results were tabulated.

### 2.3 *Semi-automated Measurements*

The same set of 12 images (each of RA and normal feet) were taken in the DICOM format again and loaded in the Mimics software, and the measurements were done in a semi-automated manner. Mimics are software that mostly processes only 3D images such as CT and MRI. X-ray images being 2D had to be converted to 3D to be processed in this software. In the Mimics software, a measured mask was applied on the hard tissues (bone) of the feet to get more accurate measurements. The measurement tool was then used between two joints to measure the joint space width and joint space narrowing. The measurements were made in millimetre (mm).

The steps involved are as follows:

1. Initially, 2D radiograph of feet images was loaded in Mimics software.
2. The bone region was segmented from the tissue region using region-thresholding method.
3. Then, measurements of interphalangeal joints (1st–5th joint) and metatarsal joint (1st–5th) width were made using available tools in Mimics.

### 2.4 *Statistical Analysis*

The Student's t-test was performed between the normal and RA patients using the measured parameters. The mean, standard deviation and standard mean error were calculated for the entire population.

## 3 **Results and Discussion**

The joint space width was measured at metatarsal and interphalangeal joints using manual and semi-automated method was illustrated in Tables 1 and 2. Figure 2 represents the JSW measured at normal feet using manual method. Figure 3 depicts the JSW measured at RA feet using manual method. Figure 4 represents the JSW measured at normal feet using semi-automated method. Figure 5 depicts the JSW

**Table 1** Joints space width measured at metatarsal joints using manual and semi-automated method

Manual method (Dicom)			Semi-automated (Mimics)	
Metatarsal joints	Normal	RA	Normal	RA
MTJ 1	27.11 ± 12.96	92.97 ± 27.31	14.37 ± 14.64	42.64 ± 10.39
MTJ 2	15.64 ± 8.14	54.11 ± 18.79	9.96 ± 3.03	22.88 ± 5.26
MTJ 3	19.34 ± 9.57	66.85 ± 24.80	11.20 ± 3.88	31.09 ± 6.45
MTJ 4	33.54 ± 18.45	105.54 ± 30.98	18.16 ± 8.29	43.51 ± 7.47

**Table 2** Joints space width measured at interphalangeal joints using manual and semi-automated method

Manual method (Dicom)			Semi-automated (Mimics)	
Interphalangeal joints	Normal	RA	Normal	RA
IPJ 1	32 ± 12.96	132.21 ± 39.03	16.64 ± 2.91	53.59 ± 14.45
IPJ 2	25.21 ± 12	95.44 ± 40.31	12.28 ± 3.29	32.49 ± 9.64
IPJ 3	25.73 ± 11.91	88.64 ± 47.88	14.43 ± 7.35	32.98 ± 12.12
IPJ 4	27.11 ± 12.96	92.97 ± 27.31	20.11 ± 9.15	46.98 ± 12.41

**Fig. 2** JSW measured at normal feet using manual method



measured at RA feet using semi-automated method. Figure 6a indicates the comparison of manual and semi-automated method for RA and normal subjects in metatarsal Joints. Figure 6b represents the comparison of manual and semi-automated method for RA and normal subjects in interphalangeal joints



Fig. 3 JSW measured at RA using manual method

Upon evaluating the measured distance between the interphalangeal and metatarsal joints of the feet, by performing Student's t-test and finding the mean and median, we could infer the following:

- (1) The distance between both the joints significantly increases from normal when the patient is suffering from rheumatoid arthritis in the feet region. This is because the skin inflates, and erosion occurs thus increasing the distance between two consecutive joints next to each other.
- (2) Normal Feet: Metatarsal—8.1–60 mm; Interphalangeal—10.8–60 mm;  
RA: Metatarsal—28–129 mm; Interphalangeal—23–205 mm;

Borman et al. evaluated the foot pain and foot involvement through radiological changes in RA patients [14]. They observed that the major cause of foot pain in RA was due to various factors such as morning stiffness, high body mass index and prolonged disease duration. They found decreased foot symptoms in 24% of RA patient who have undergone medical treatment. Simon detected marginal bone erosion at metatarsal region in tomosynthesis method and compared with radiography [15]. The researcher observed that tomosynthesis was found to be superior to radiography in detecting the bone erosions in feet region. The results of our research described the significant change in specifically the joint space dimensions as a measurement of



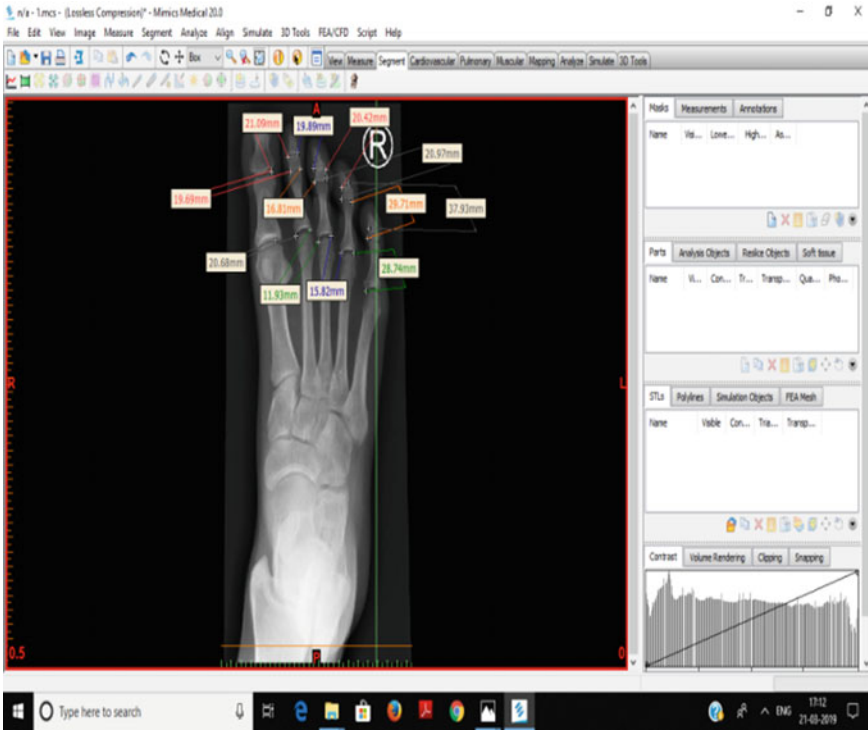


Fig. 4 JSW measured at normal feet using semi-automated method

RA. This helps identify the intensity of the condition better because joint space narrowing progressively increases as the severity of RA increases. This gives an upper hand at identifying the important aspects of RA especially in the feet region.

The limitation of our study includes (i) negligence of the medications factor (ii) Focuses only on the front feet region. A clear comparison between a manual method of measurement and semi-automated method of measurement had been implemented in the proposed approach. The deviations when calculated shown that the semi-automated method is more accurate. This is because of the masking feature of the Mimics software that allows the measurement of the hard tissue region accurately. The variation in the readings between the two methods provides a range of space for further research using automated methods.

X-ray imaging remains one of the most efficient methods to diagnose RA in the feet region. The semi-automated and manual methods used in the proposed approach provided an almost accurate mathematical consensus of the RA condition. This research can further be extended using an automated segmentation method. The manual and semi-automated method results can then be compared to find out which of them give a better and more precise evaluation of the condition. Furthermore, these techniques

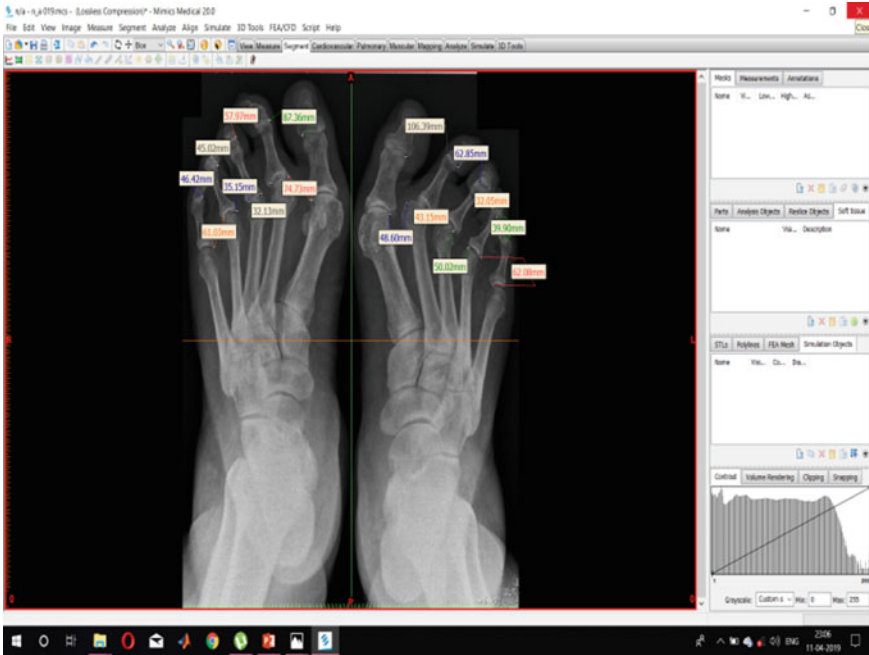
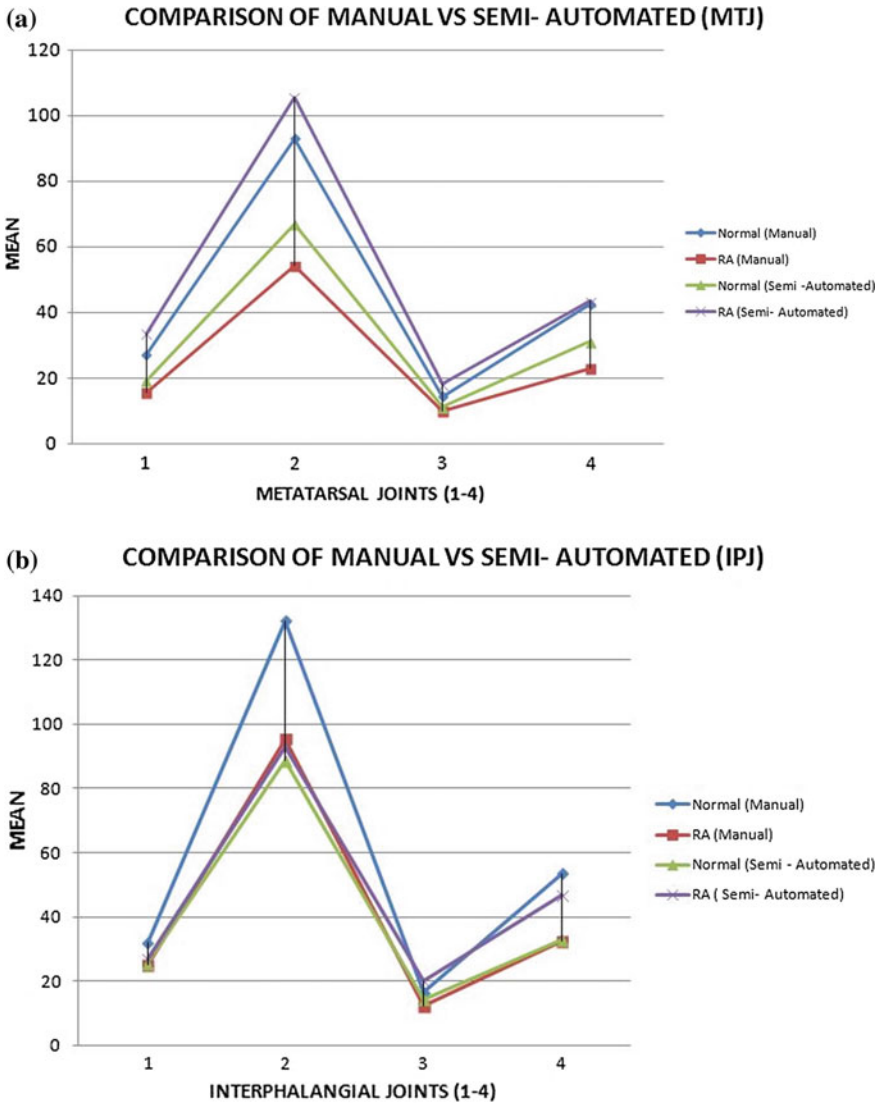


Fig. 5 JSW measured at RA patients’ feet using semi-automated method

can be used on other joint regions of the body, and scoring can also be done based on a standard methods.

### 4 Conclusion

The manual and semi-automatic methods were used to measure the joint space width distance more accurately using Mimics (semi-automatic) compared to DICOM viewer (manual). This is because the bone region of the feet was distinctly segmented in Mimics compared to DICOM viewer. At every stage, there have been various methods used to characterize RA to get a better picture of its actual trigger and better treatment methods henceforth. This method is one such approach and a step towards making diagnosis and hence understanding of this medical condition better.



**Fig. 6** **a** Comparison of manual and semi-automated method for RA and normal subjects in metatarsal joints. **b** Comparison of manual and semi-automated method for RA and normal subjects in interphalangeal joints

**Acknowledgements** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee. The study is approved by Ethical Clearance Committee of SRM Hospital and Research Centre. The informed consent form was obtained from all the participants for the study.

## References

1. Gupta B, Ritu G, Ashok K, Madhu B. Rheumatoid arthritis research in India: a scientometric assessment of publications during 2007–2016. *Ortho Res Online J* 3(1). <https://doi.org/10.31031/oproj.2018.03.000552>
2. <http://www.arthritis-india.com/rheumatoid-arthritis.html>
3. Alamanos Y, Voulgari PV, Drosos AA (2006) Incidence and prevalence of rheumatoid arthritis based on the 1987 American College of Rheumatology criteria: a systematic review. *Semin Arthritis Rheum* 36:182–188. <https://doi.org/10.1016/j.semarthrit.2006.08.006>
4. Chopra A, Abdel-Nasser A (2008) Epidemiology of rheumatic musculoskeletal disorders in the developing world. *Best Pract Res Clin Rheumatol* 22:583–604. <https://doi.org/10.1016/j.berh.2008.07.001>
5. Van der Heijde DM (2000) Radiographic imaging: the ‘gold standard’ for assessment of disease progression in rheumatoid arthritis. *Rheumatology (Oxford, England)* 39(Suppl 1):9–16
6. Salaffi F, Carotti M, Carlo M (2016) Conventional radiography in rheumatoid arthritis: new Scientific insights and practical application. *Int J Clin Exp Med* 9(9):17012–17027
7. Platten M, Kisten Y, Kälvesten J et al (2017) Fully automated joint space width measurement and digital X-ray radiogrammetry in early RA. *RMD Open* 3:e000369. <https://doi.org/10.1136/rmdopen-2016-000369>
8. Plant MJ, Saklatvala J, Jones PW et al (1994) *Clin Rheumatol* 13:487. <https://doi.org/10.1007/BF02242948>
9. Bakker KWD, Kroon HM, Zwinderman AH, Breedveld FC, Hazes JMW (2000) Radiographic damage of large joints in long-term rheumatoid arthritis and its relation to function. *Rheumatology* 39(9):998–1003
10. Kuper HH, van Leeuwen MA, van Riel PL et al (1997) Radiographic damage in large joints in early rheumatoid arthritis: relationship with radiographic damage in hands and feet, disease activity, and physical disability. *Br J Rheumatol* 36:855–860
11. Snehalatha U, Anburajan M (2012) Dual tree wavelet transform based watershed algorithm for image segmentation in hand radiographs of arthritis patients and classification using BPN neural network. In: *World congress in information and communication technologies (WICT 2012)*, pp 448–452
12. Boiny S, Guillemin F (2001) Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. *Ann Rheum Dis* 60(9):817–827
13. Heidari B (2011) Rheumatoid arthritis: early diagnosis and treatment outcomes. *Caspian J Intern Med* 2(1):161–170
14. Borman P, Ayhan F, Tuncay F, Sahin M (2012) Foot problems in a group of patients with rheumatoid arthritis: an unmet need for foot care. *Open Rheumatol J* 6:290–295
15. Simoni P (2018) Optimisation of X-rays imaging techniques for the assessment of joint space. *J Belg Soc Radiol* 102(1):23. <https://doi.org/10.5334/jbsr.1447>

# Translation into Pali Language from Brahmi Script



Neha Gautam, Soo See Chai and Megha Gautam

**Abstract** The Brahmi script is widely used in ancient time for writing the various types of religious books, information about the rules and regulation of the kingdom of king Ashok and Pali language is used to read the Brahmi script. Pali to English dictionary is available to understand the Pali language. However, any dictionary or translation system related to Brahmi script to Pali script is not available. In order to understand the content in Brahmi script, an automated system for facilitating machine translation of Brahmi text to Pali language is suggested in this study. Detail research on Brahmi script to Pali language has been accomplished, and necessary rules are analyzed. Based on the research and analysis, this study illustrates one-to-one and many-to-one mapping system for Brahmi text to Pali language translation.

**Keywords** Brahmi script · Pali language · Conversion rules · Mapping method

## 1 Introduction

A machine-translation system is used to convert a target language output sentence from a source language sentence without changing the meaning of the source language sentence. The translator system analysis on the grammatical structure of the input sentence and transfers it to the target language structure [1]. Various types of translation system are available related to the modern script like English, Devanagari, Bangla, etc., and few are available in the ancient scripts, such as Sanskrit.

---

N. Gautam (✉) · S. S. Chai  
University of Malaysia Sarawak, Kota Samarahan, Malaysia  
e-mail: [nehagautam1208@gmail.com](mailto:nehagautam1208@gmail.com)

S. S. Chai  
e-mail: [sschai@unimas.my](mailto:sschai@unimas.my)

M. Gautam  
National Institute of Pharmaceutical Education and Research, Sahibzada Ajit Singh Nagar, India  
e-mail: [meghadops11@gmail.com](mailto:meghadops11@gmail.com)

Kunchukuttan et al. [2] suggested a Brahmi-Net system; it is an online system for transliteration and script conversion for all major Indian language pairs (306 pairs). The system covers 13 Indo-Aryan languages, 4 Dravidian languages, and English. Apart from the English language, all languages and scripts are driven from the Brahmi script, thus Brahmi-Net name is used for the system. However, this system and any other translation system did not focus on translating the Brahmi script. Brahmi script is an ancient script, and it has a very rich background in terms of culture and religion, such as Lalitavistara (Buddhist), Samavayanga Sutra, and Pannavana Sutra (Jain) were written in Brahmi script. So, the translator of Brahmi script to another language can help to understand it.

Translator is used for converting a language from another language and Brahmi is not a language. So, firstly, Brahmi script should be converted into a language then a translation can be performed easily. Pali language was used to read the Brahmi script [3], and Pali to English dictionary is available to understand Pali language [4, 5]. So, translation to Pali language from Brahmi script can help to understand the Brahmi content in the modern world. Brahmi script is a way to write Pali language, so translating process is not depended on any type of grammar, verb, etc. Thus, one-to-one mapping is followed to build the transition system where each character of Brahmi text is separately converted into Pali language.

Details of the previous works related to this study are given in part II. Information of the characteristic and properties of Brahmi script is described in Sect. 3. Information of the Pali language and relationship of the Pali language of the Brahmi script is discussed in Sect. 4. The methodology of translation process including input, mapping, and output steps is presented in Sect. 5. Results of the proposed system are presented in Sect. 6, and concluding remarks are mentioned in Sect. 7.

## 2 Literature

The translation process helps the researcher to translate one content into another content such as translating one language to another language, one script to another language, one language to another script, and one script to another script. The translation of one language to another is a tough task because languages are based on rules of grammar, and each language has its own grammar and sentence structure. However, various studies have worked on translating one language to another language, while other studies are still focusing on improving translation system performance.

Various translation systems have been built using major modern popular languages such as the translation of Hindi to English [6], English to Hindi [7], Telugu to English [8], English to Telugu [9, 10], Telugu to English [11], English to Malayalam [12], Malayalam to English [13], English to Sanskrit [14], English to Bangla [15], Bangla to English [16], Telugu to Sanskrit [17], Malayalam-Tamil Machine Translation [18], Kannada to Telugu [19], Telugu to Marathi [20], etc. Sata-Anuvadak is a translation system in which 11 Indian languages (Hindi, Urdu, Punjabi, Bengali, Gujarati, Marathi, Konkani, Tamil, Telugu, Malayalam, and English) were translated

using an English–Indian language translator and Indian-Indian language translation [21].

As can be seen, not a single study investigated Brahmi script translation. However, Brahmi script has been used for a long time [third BCE to the sixth century (CE)]. Brahmi script is an ancient script, and it has a very rich background. So, the translator of the Brahmi script to another language could help the general population understand religious books, the culture of this religion, and the rules of King Ashoka’s Empire. However, a translator is used to convert one language into another language, and Brahmi is not a language. Therefore, firstly, the Brahmi script should be converted into a language so that its translation can be performed easily. Pali language was used to decipher or read the Brahmi script [22], as a Pali to English dictionary is available for understanding the Pali language [4, 5]. Hence, the translation of Brahmi script to Pali language will fill this gap and help the researcher and general population to better understand Brahmi content in the modern world.

The script-to-language translation process was performed prior to translating the Brahmi script to Pali language. A few studies have been done in the field of script-to-language translation, and script-to-script translation. Bangla text to Braille translation was developed to help blind people understand the Bangla text. This system used a mapping process to translate the Bangla text to Braille and was able to convert all Bangla characters. The Brahmi-Net system was also developed involving the translation of 18 languages (13 Indo-Aryan languages, 4 Dravidian languages, and English) and the conversion of 10 scripts (Bengali, Gujarati, Kannada, Malayalam, Odia, Punjabi, Devanagari, Sinhala, Tamil, and Telugu) [2]. Brahmi-Net is considered to be a form of language-to-language translation and script-to-script translation. Brahmi-Net also uses the mapping process to perform script-to-script translation because it is easy to use and yielded the best performance.

Brahmi script is a way to write Pali language, so the translation process does not depend on the type of grammar, verb, etc. Instead, the one-to-one mapping must be performed to build the transition system in which each character of the Brahmi text is separately converted into Pali language. Hence, mapping is a better option for translating Brahmi script into Pali language.

### 3 Brahmi Script

Brahmi script is an ancient script. Brahmi script was used from third BCE to the sixth century (CE) [23, 24]. This script was used to write various religious books on Buddhism and Jainism culture [25, 26]. At one time, the Brahmi script was referred to as the “pin-man” script, that is, the “Stick Figure” script in English [26, 27]. In other words, Brahmi characters have geometric features such as lines, curves, corners, and dots [28] and these features can help identify the Brahmi script from other various scripts.

### ***3.1 Properties of Brahmi Script***

- (a) A vowel that follows a consonant may take a modified shape and may be placed on the left, right, top, or bottom of the consonant depending on the vowel. Such characters are called compound characters [29]. Sometimes, two vowels follow a consonant and become a new character called complex compound characters. Both properties can be seen in the Brahmi script such as the Devanagari script and the Bangla script.
- (b) Brahmi script has 368 characters in total, including 33 consonants, 10 vowels, and (325) compound characters making up the rest [30, 31].
- (c) Brahmi script is written from left to right.

### ***3.2 Characteristic***

The shape of consonant will change according to the vowels to build the compound characters. The features add in the consonants to create the compound characters called “Matra.” Mostly, these “Matra” can be seen on the outer edge of the consonants (but not always because it also depends on the shape on the consonants) or add a dot feature (.) after the consonant.

## **4 Pali Language**

Pali is a middle Indo-Aryan language that is in the Prakrit language group and was indigenous to the Indian subcontinent [32]. But no one is using this language in modern time for the communication; thus, this is called a dead language [32]. Thus, the first Pali Dictionary to English published in 1869 and the first Pali translated text to English released in 1872 to understand the text in Pali language [5].

### ***4.1 Relationship Between Characters of Brahmi Script and Pali Language***

Vowel and consonants of Brahmi script and pronunciation of Brahmi script in Pali language are also mentioned in Table 1. However, Brahmi script has the property of compound characters where a consonant is followed by one vowel (Table 1). So, the way of pronunciation in Pali language is also changed according to the compound characters of Brahmi script. Pali language of the Brahmi consonant and Pali language of the vowel are used as prefix and suffix of the compound character, respectively (Table 1).



**Table 1** Translation of Pali language from Brahmi script

Consonants		Vowels		Compound characters
Pali (Brahmi)		Pali (Brahmi)		Pali (Brahmi)
Ka (𑀓)	(+)	a (𑀅)	=	ka (𑀓𑀅)
Ka (𑀓)	(+)	ā (𑀅̄)	=	kā (𑀓𑀅̄)
Ka (𑀓)	(+)	i (𑀇)	=	ki (𑀓𑀇)
Ka (𑀓)	(+)	ī (𑀇̄)	=	kī (𑀓𑀇̄)
Ka (𑀓)	(+)	u (𑀆)	=	ku (𑀓𑀆)

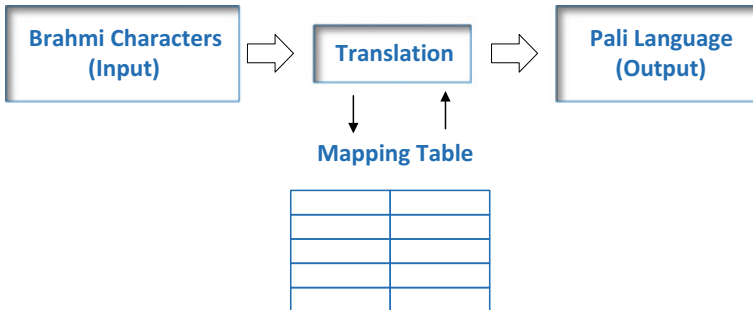
Like: Ka (Pali consonant) + ai (Pali vowel) = kai (Pali compound character) of  
 𑀓 (Brahmi consonant) + 𑀅 (Brahmi vowel) = 𑀓𑀅 (Brahmi compound character)

## 5 Methodology

Translation process is divided into three steps: input of the system, mapping, and output of the system (Fig. 1), where mapping is the main step for the translation.

### 5.1 Input

Brahmi word is used as the input of the translation system. Length of the words and characters in the text are not fixed. All characters of each line are read from the left side to the right side, and all lines of the text are read from top to bottom.



**Fig. 1** Computational model for Brahmi text to Pali language translation

**Table 2** Translation table of Pali language from Brahmi script

Brahmi script	Pali language	Brahmi script	Pali language
𑀓	ka	𑀓	kū
𑀣	kā	𑀣	ke
𑀤	ki	𑀤	kai
𑀥	kī	𑀥	ko
𑀦	ku	𑀦	kaṃ

### 5.2 Mapping

The association of each element of a given set with an element of another set is called mapping. In simple words, the mapping process replaces one set of words by another set of words. The first set of elements is Brahmi characters, and the second set of the element is the Pali language. According to Table 2, the relationship of all 368 types of Brahmi characters with Pali characters was identified, and the Brahmi characters were replaced with Pali characters in the mapping step.

### 5.3 Output Text

After the mapping step, all Brahmi characters were translated into Pali characters. However, all characters were isolated and translated into Pali language, so the translated Pali characters were arranged according to the place of the Brahmi characters in the word and text. Figure 2 shows the translation process as an example. Each character of a Brahmi text was converted into Pali language using the mapping step, and all isolated characters of the Pali language were rearranged according to the presence of the characters in the Brahmi text. As already mentioned, that Pali is a language, so, English phonetic characters were used to present the Pali text.

**Fig. 2** Translation system



## 6 Experimental Result

All characters can be correctly translated by this system because of the mapping step. The recognition system is capable of reading the Brahmi characters so, the translation system showed 100% accuracy to translate the Brahmi text into Pali text.

## 7 Conclusion

Understanding the Brahmi text can help to understand the various religious books. So, translation can be an useful manner to understand the Brahmi text. Translation system from Brahmi script to Pali language is suggested by this study. Brahmi word is used as input of the translation system. Mapping process is used to translate the Brahmi word and mapping process just replace the Brahmi words to Pali words and translation system is 100% capable to translate the text into Pali language by Brahmi text.

## References

1. Ashrafi SS, Kabir MH, Anwar MM, Noman A (2013) English to Bangla machine translation system using context-free grammars. *Int J Comput Sci Issues (IJCSI)* 10:144
2. Kunchukuttan A, Puduppully R, Bhattacharyya P (2015) Brahmi-net: a transliteration and script conversion system for languages of the indian subcontinent. In: *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: demonstrations*, pp 81–85
3. Trautmann TR (2006) *Languages and nations: the Dravidian proof in colonial Madras*. Yoda Press, New Delhi
4. Davids TWR, Stede W (1993) *Pali-English dictionary*. Motilal Banarsidass Publications, New Delhi
5. Society PT, Davids TWR, Stede WB (2016) *The Pali text society's Pali-English dictionary*. Wentworth Press, Australia
6. Gehlot A, Sharma V, Singh SP, Kumar A (2015) Hindi to English transfer based machine translation system. *arXiv preprint [arXiv:1507.02012](https://arxiv.org/abs/1507.02012)*
7. Dwivedi SK, Sukhadeve PP (2015) Translation rules for English to Hindi machine translation system: homoeopathy domain. *Int Arab J Inf Technol* 12:791–796
8. Padhya D, Sheth J (2019) A review of machine translation systems for Indian languages and their approaches. In: *Emerging trends in expert applications and security*. Springer, Berlin, pp 103–110
9. Lingam K, Lakshmi ER, Theja LR (2014) Rule-based machine translation from English to Telugu with emphasis on prepositions. In: *2014 first international conference on networks & soft computing*, pp 183–187
10. Suryakanthi T, Sharma K (2015) Discourse translation from English to Telugu. In: *Proceedings of the third international symposium on women in computing and informatics*, pp 222–227
11. Prasad TV, Muthukumaran GM (2013) Telugu to English translation using direct machine translation approach. *Int J Sci Eng Inv* 2:25–32
12. Nithya B, Joseph S (2013) A hybrid approach to English to Malayalam machine translation. *Int J Comput Appl* 81:11–15

13. Haroon RP, Shaharban T (2016) Malayalam machine translation using hybrid approach. In: 2016 international conference on electrical, electronics, and optimization techniques, pp 1013–1017
14. Rathod SG (2014) Machine translation of natural language using different approaches. *Int J Comput Appl* 102:26–31
15. Rabbani M, Alam KMR, Islam M (2014) A new verb based approach for English to Bangla machine translation. In: 2014 international conference on informatics, electronics & vision, pp 1–6
16. Rhaman MK, Tarannum N (2012) A rule based approach for implementation of Bangla to English translation. In: 2012 international conference on advanced computer science applications and technologies, Kuala Lumpur, Malaysia, pp 13–18
17. Rao TK, Prasad T (2016) Machine translation of Telugu plural pronoun declensions to Sanskrit. In: 2016 2nd international conference on applied and theoretical computing and communication technology, Bangalore, India, pp 298–304
18. Jayan JP, Rajeev R, Rajendran S (2011) Morphological analyser and morphological generator for Malayalam-Tamil machine translation. *Int J Comput Appl* 13:15–18
19. Sindhu D, Sagar B (2017) Dictionary based machine translation from Kannada to Telugu. In: IOP conference series: materials science and engineering, p 012182
20. Ghosh S, Thamke S (2014) Translation of Telugu-Marathi and vice-versa using rule based machine translation. arXiv preprint [arXiv:1406.3969](https://arxiv.org/abs/1406.3969)
21. Kunchukuttan A, Mishra A, Chatterjee R, Shah R, Bhattacharyya P (2014) Sata-anuvadak: tackling multiway translation of Indian languages. Presented at the proceedings of the ninth international conference on language resources and evaluation, Reykjavik, Iceland
22. Steever SB (2015) *The Dravidian languages*. Routledge, London
23. Seland EH (2014) Archaeology of trade in the western Indian Ocean, 300 BC–AD 700. *J Archaeol Res* 22:367–402
24. Singh U (2008) *A history of ancient and early medieval India: from the stone age to the 12th century*. Pearson Education, India
25. Subba JR (2008) *Evaluation of man and the modern society in Sikkim*. Gyan Publishing House, New Delhi
26. Salomon R (1998) *Indian epigraphy: a guide to the study of inscriptions in Sanskrit, Prakrit, and the other Indo-Aryan languages*. Oxford University Press, Oxford
27. Keay J (2008) *India: a history*. Paw Prints, USA
28. Gautam N, Chai SS (2017) Optical character recognition for Brahmi script using geometric method. *J Telecommun Electron Comput Eng* 9:131–136
29. Pal U, Chaudhuri B (2004) Indian script character recognition: a survey. *Pattern Recogn* 37:1887–1899
30. Siromoney G, Chandrasekaran R, Chandrasekaran M (1983) Machine recognition of Brahmi script. *IEEE Trans Syst Man Cybern SMC-13*:648–654
31. Roy A, Mandal M (2016) *Brahmi: rediscovering the lost script*
32. Noss DS, Grangaard B (2016) *A history of the world's religions*. Routledge, London

# The Dataset for Printed Brahmi Word Recognition



Neha Gautam, Soo See Chai and Megha Gautam

**Abstract** Publicly available dataset is important for character, word or document recognition. The use of a standardized dataset will provide a fair or reliable comparison between the performances of the underlying recognition algorithms. Research on Brahmi words recognition had achieved encouraging results. However, there is no publicly available standardized Brahmi dataset. In this paper, the steps in producing a publicly available Brahmi dataset are presented. These steps include data collection, segmentation, storage, labeling, and statistical distribution. A total of 7,011 images of Brahmi characters were collected. The collected dataset is divided into three classes: vowel, consonants, and compound characters. In total, there are 170 classes with 4 of these classes belong to vowels, 27 classes of consonants, and 139 classes of compound characters. The 170 classes of characters are further divided into training and testing sets; 6,475 images in the training set while 536 images in the testing set.

**Keywords** Brahmi script · Segmentation · Writing system · Dataset · Brahmi word recognition

## 1 Introduction

Automatic word recognition remains a challenging task even though the latest works have shown significant improvements in the performance of the word recognition systems. Optical character recognition (OCR) technology has enabled us to convert scanned paper documents, pdf files or images captured by a digital camera

---

N. Gautam (✉) · S. S. Chai  
University of Malaysia Sarawak, Kota Samarahan, Malaysia  
e-mail: [nehagautam1208@gmail.com](mailto:nehagautam1208@gmail.com)

S. S. Chai  
e-mail: [sschai@unimas.my](mailto:sschai@unimas.my)

M. Gautam  
National Institute of Pharmaceutical Education and Research, Mohali, India  
e-mail: [meghadops11@gmail.com](mailto:meghadops11@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_13](https://doi.org/10.1007/978-981-15-2329-8_13)

into editable, automatic readable and searchable data. In order to benchmark the recognition algorithms, a standardized dataset is required. Standardized datasets are available for some common languages like English [1], Devanagari [2], and Chinese [3]. Brahmi script has a wealthy archeological background and had been used in ancient time. However, research on Brahmi character and word recognition was not based on a standardized dataset, causing the results obtained to be questionable.

In the work of Brahmi scripts recognition, the suitable features of the Brahmi scripts are first extracted, and next using the suitable classifier, the extracted features will be classified. For instance, geometric features [4] of the Brahmi scripts had been used as the suitable features for Brahmi scripts recognition, zoning method [5, 6] was applied to obtain the features of the Brahmi scripts and template matching [6], and artificial neural network (ANN) [5] were used as classifier to classify the extracted features. In 2016, Vellingiriraj et al. [5], who worked with Brahmi word recognition, notified that an official dataset is required for Brahmi scripts in order to evaluate the performance of the Brahmi scripts recognition algorithms. Currently, researchers in Brahmi scripts recognition research are using their own datasets with limited samples. A publicly available standardized dataset will serve as a transparent platform for comparison of the different recognition algorithms.

The process of developing the standardized Brahmi dataset involves data collection, segmentation, storage, labeling, and statistical distribution. During the data collection, various free online resources and personal Web sites were utilized. The collected images were of various sizes and resolutions. These images were next segmented, labeled, and stored. There was no suitable technique that was found from the literature to segment the characters from the Brahmi words and text automatically. Therefore, in this study, all the characters are manually segmented from the Brahmi word and text. The isolated characters were next labeled and stored. Similar characters were grouped with a unique name.

The remainder work of the article is organized: the origin, the writing style of various scripts of the world writing systems, and the details of previous studies based on the dataset of different scripts are shown in Sects. 2 and 3, respectively. Methodology to build the dataset including the details of the training and testing dataset described in Sect. 4 and the conclusion of this study are presented in Sect. 5.

## 2 Writing Systems and Scripts of the World

In character recognition, the main distinguishing features are the characteristics of the different writing systems and the structural properties of the characters. According to Ghosh et al. [7], there are six prominent writing systems, as shown in Fig. 1. The detail of these six writing systems is further explained.

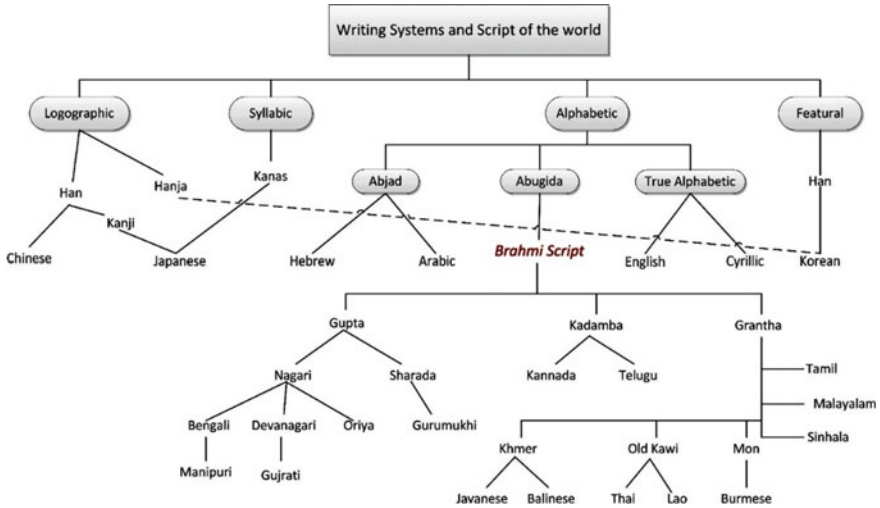


Fig. 1 World writing system [7]

### 2.1 *Logographic System*

Sub-headings should be typeset in boldface italic and capitalize the first letter of the first word only. Section number to be in boldface roman.

### 2.2 *Syllabic System*

The logographic system uses a single symbol to represent the entire word. The syllabic system, on the other hand, uses a set of written symbols to represent or approximate syllables or phonic sounds, to make up words.

### 2.3 *Abjad System*

Writing system can include logographic system such as Chinese (images that represent words or ideas, sometimes with a phonetic element denoting the pronunciation); alphabetic systems (consonant alphabets such as Arabic, or phonemic alphabets with vowels, such as Latin); and syllabaries (with symbols for consonants and vowels together), such as Ethiopic [8].

## ***2.4 Abugida System***

Nearly, entire scripts in Southeast Asia and India belong to the Abugida system. The alphabetic-like writing system in the Brahmic family of scripts originally acquired from ancient Indian Brahmi script. Figure 1 shows the evaluation of original Brahmic scripts which were popular in India and Southeast Asia.

Various scripts such as Bengali, Gujarati, Manipuri, Gurumukhi, and Devanagari are derived from Brahmi script and some characteristics are also common in these scripts to Brahmi script. Such as, all these have “Matra” or “Shirorekha” features in compound character.

## ***2.5 True Alphabetic***

This refers mainly to the Roman script, with English as the most popular script or language worldwide. The original alphabetic system contains letters for consonants and vowels.

## ***2.6 Featural System***

The featural system includes phonemes composed of symbols or characters which represent the features. Prominently, Korean Hangul is according to the featural system. Mixing logographic Hanja with featural Hangul, the Korean script consists of more Hangul than Hanja, as depicted in Fig. 1. Korean script is comparatively less complicated and less compact while the Japanese and Chinese have more ellipses and circles.

# **3 Literature Review**

Many current scripts have already recognized by automatic recognition systems, and some works also find out the recognition of ancient scripts. Dataset plays an important role to build a recognition system. This literature is covered the available standard datasets related to various popular scripts of only Abugida system because Brahmi script is a part of this system. Abugidas include the extensive Brahmic family of scripts of South and Southeast Asia. Various scripts including Brahmi script are part of the Abugida system such as Devanagari and Bangla. Numerous types of work completed related to the dataset of this system. More information about the datasets are further discussed.



A latest work related to Bangla character dataset [2] collected the 59,892 number of characters as the samples, where 41,536 number of compound characters, 15,000 number of primary characters as well as 3,360 number of vowels and consonant. A benchmark dataset of isolated Bangla handwritten compound characters was introduced [9]. The total number of separated characters was 55,278 including 199 various pattern shapes and 171 types of characters classes in the dataset. In the year 2012, CMATERdb1 dataset [8] developed by Jadavpur University, Kolkata, India. This dataset used 150 page-level documents, among the 150 handwritten document pages, 100 pages written in Bangla script with the remaining pages written in Bangla text mixed with English words.

To introduced a character-level Devanagari script dataset [10], 750 writers contributed to writing the 5,137 and 20,305 number of samples of Devanagari numerals and Devanagari characters, respectively. An another dataset of Devanagari script, CPAR-2012 [11] included 35,000 isolated handwritten numerals, 83,300 characters, 2,000 constrained, and 2,000 unconstrained handwritten pangrams. For Tamil dataset, around 500 different hands contributed to write city name in Tamil script to build the dataset (Tamil-DB). It was primarily developed dataset for the postal automation system [12]. An unconstrained Kannada handwritten text database (KHTD) introduced [13]. KHTD contained 204 handwritten documents produced by 51 native speakers of Kannada in four different categories. This dataset constructed by 4,298 text lines and 26,115 words. A multipurpose dataset of handwritten samples of Tamil and Kannada [14] contained approximately 100,000 words from 600 various classes.

Few studies took multiple scripts to build the datasets, like: a handwritten isolated numeral dataset [15], such as Devanagari, Bangla, and Oriya scripts. This dataset comprises of 23,392 Bangla numerals written by 1,106 individual writers, 22,556 Devanagari numerals written by 1,049 persons, and 5,970 Oriya numerals written by 356 hands. PBOK [16] was a page-level dataset which contained four different scripts as Bangla, Oriya, Persian, and Kannada. PBOK generated two types of ground truths, based on pixel information and content information. PBOK developed with a total number of 707 text pages, 12,565 text lines, 104,541 words, 423,980 characters, and 436 persons were involved to write the content for this dataset. A new study related to the word-level numeral dataset of Bangla, Devanagari, Roman, and Urdu script was suggested [17], and all four scripts are well used in India for the writing purpose and communication. This dataset consists of a total of 5,659 numeral strings comprising of Bangla (1,602), Devanagari (1,139), Roman (1,602), and Urdu (1,316) images. Total 43 writers were involved in the building of this dataset. A new dataset PHDIndic\_11 [2] suggested, which contains 1,458 handwritten page-level images from 11 official scripts of India, namely: Bangla, Devanagari, Urdu, Gurumukhi, Oriya, Roman, Telugu, Gujarati, Malayalam, Tamil, and Kannada. Total 463 individuals from different places of India have written 1,458 pages. Similarly, a dataset finds out related to railway signboards. A 'station signboard' database [18] which includes the railway signboards written in five different Indic scripts: Hindi, Odia, Urdu, Telugu, and English script. A total number of 500 images gathered related to

ground truths which produced in a semiautomatic way. These images pose various challenges such as illumination effect, occlusion, perspective distortion, and so on.

All discussed studies are the evidence that most studies have been conducted on modern scripts like English, Devanagari, etc., and few studied focused on the ancient scripts like Jawi. According to this whole discussion, it can be seen that no work has found on Brahmi dataset. Few works can be seen related to Brahmi character recognition [4, 6] and Brahmi word recognition [5, 19, 20] but the local dataset used to train the system in all these systems. Although Vellingiriraj et al. [5], who worked with Brahmi word recognition, notified that an official dataset is required for Brahmi scripts in order to evaluate the performance of the Brahmi scripts recognition algorithms.

## **4 Methodology**

Ideally, any recognition system requires a dataset to train and test the system. Similarly, the dataset will present an important part to recognize the Brahmi word with reliable accuracy. Methodology to build a dataset discuss in this section. Steps of the method include collect the data, isolate the useful character from the words and texts, divided the isolated characters into group, and execute labeling process to store the data in a proper manner for further use.

### ***4.1 Data Collection***

Finding suitable sources for the samples of Brahmi text was considered as the first step for building the dataset in this study. Various online resources were used to collect the images of Brahmi words and text as samples. However, only images that were visible and noise free were considered for the sample collection. During the collection of the samples, many images of the text and words were collected from various Web sites and other online resources. After this step, the characters were separated from the words and text, an essential task, because these isolated characters were then used to train the system.

### ***4.2 Segmentation***

Segmentation process is used to isolate all the characters from the text. Segmentation process is divided into two parts. However, this study is not found out any suitable method for the segmentation of the characters from the Brahmi text from the literature. Singh and Kushwaha [21] suggested a technique to isolate the characters from the text, but this method is not enough capable to segment the all the characters from

the Brahmi text. So, all characters are manually separated from the words and text and it considered that all features of the characters are placed together in an image.

### ***4.3 Data Storage and Labeling***

All isolate characters should be in a uniform format. So, all extract images of characters change into JPG format. Next to that decided the group and place the characters accordingly. So, all characters that have same structure (shape) are placed together in a folder (directory). Next to that give a unique name of all directories which are the collections of the same types of characters.

All directories are divided into three categories: vowel, consonants, and compound characters. All vowels are denoted by alphabet (a, b,...). Similarly, all consonants are denoted by digits (1, 2,...). Compound characters are the combination of the consonants and vowels so, for the better understanding, every single directory of compound character is represented by two entities: a number along with an alphabet, where numbers represent the consonants, and alphabet (a, b,...) denote vowels.

### ***4.4 Statistical Distribution***

Brahmi dataset is the combination of the various types of samples of characters. In the dataset, 170 classes of a total number of characters where 27 varieties of consonants, 4 types of vowels, and rest (139) belongs to compound characters. As the knowledge, Brahmi script has 368 types of characters [19], but could not find all kind of characters because of the lack of resources. The database consists of printed Brahmi words composed of 6,475 images of 170 classes for the training and 536 isolated characters of 170 classes for the testing. This dataset also has 55 samples of Brahmi text (which are not isolated) to check the performance of segmentation methods in the future.

## **5 Conclusion**

Brahmi word recognition dataset, a word-level samples of Brahmi script, has been suggested. Steps to complete the dataset and converted all row samples into useful manner are discussed to introduce the Brahmi word recognition system. Moreover, Brahmi dataset will also able to provide reliability on the performance of the system. Various types of words and text are collected to build the dataset. However, characters are manually segmented from the words and text. So, the limitation of dataset, if some more word is further added to increase the dataset, so that samples should be isolated. Various types of vowels, consonants, compound characters, and

complex compound characters are not part of the Brahmi dataset because of the lack of resources. Moreover, all steps after data collection, such as segmentation, data storage, and labeling are completed manually by this study, so further work focus to do it automatically. Furthermore, Brahmi word recognition system can be developed by using this dataset.

## References

1. LeCun Y, Cortes C, Burges C (2010) MNIST handwritten digit database. AT&T Labs [Online], vol 2. Available: <http://yann.lecun.com/exdb/mnist>
2. Obaidullah SM, Halder C, Santosh K, Das N, Roy K (2018) PHDIndic\_11: page-level handwritten document image dataset of 11 official Indic scripts for script identification. *Multim Tools Appl* 77:1643–1678
3. Su T, Zhang T, Guan D (2007) Corpus-based HIT-MW database for offline recognition of general-purpose Chinese handwritten text. *Int J Doc Anal Recognit* 10:27–38
4. Gautam N, Chai SS (2017) Optical character recognition for Brahmi script using geometric method. *J Telecommun, Electron Comput Eng* 9:131–136
5. Vellingiriraj E, Balamurugan M, Balasubramanie P (2016) Text analysis and information retrieval of historical tamil ancient documents using machine translation in image zoning. *Int J Lang Lit Linguist* 2:164–168
6. Gautam N, Sharma R, Hazrati G (2016) Handwriting recognition of Brahmi script (an artefact): base of PALI language. In: *Proceedings of first international conference on information and communication technology for intelligent systems*, vol 2. Springer, Cham, pp 519–527
7. Ghosh D, Dube T, Shivaprasad A (2010) Script recognition—a review. *IEEE Trans Pattern Anal Mach Intell* 32:2142–2161
8. Sarkar R, Das N, Basu S, Kundu M, Nasipuri M, Basu DK (2012) CMATERdb1: a database of unconstrained handwritten Bangla and Bangla-English mixed script document image. *Int J Doc Anal Recogn* 15:71–83
9. Das N, Acharya K, Sarkar R, Basu S, Kundu M, Nasipuri M (2014) A benchmark image database of isolated Bangla handwritten compound characters. *Int J Doc Anal Recogn* 17:413–431
10. Dongre VJ, Mankar VH (2012) Development of comprehensive Devanagari numeral and character database for offline handwritten character recognition. *Appl Comput Intell Soft Comput* 2012:1–6
11. Kumar R, Kumar R, Ahmed P (2013) A benchmark dataset for Devnagari document recognition research. In: *6th international conference on visualization, imaging and simulation*, Lemesos: Cyprus, pp 258–263
12. Thadchanamoorthy S, Kodikara N, Premaretne N, Pal U, Kimura F (2013) Tamil handwritten city name database development and recognition for postal automation. In: *12th international conference on document analysis and recognition*, Washington, DC, USA, pp 793–797
13. Alaei A, Nagabhushan P, Pal U (2011) A benchmark Kannada handwritten document dataset and its segmentation. In: *International conference on document analysis and recognition*, Beijing, China, pp 141–145
14. Nethravathi B, Archana C, Shashikiran K, Ramakrishnan AG, Kumar V (2010) Creation of a huge annotated database for Tamil and Kannada OHR. In: *International conference on frontiers in handwriting recognition*, Kolkata, India, pp 415–420
15. Bhattacharya U, Chaudhuri B (2005) Databases for research on recognition of handwritten characters of Indian scripts. In: *Eighth international conference on document analysis and recognition*, Seoul, South Korea, pp 789–793

16. Alaei A, Pal U, Nagabhushan P (2012) Dataset and ground truth for handwritten text in four different scripts. *Int J Pattern Recognit Artif Intell* 26:1–25
17. Obaidullah SM, Halder C, Das N, Roy K (2016) A corpus of word-level offline handwritten numeral images from official Indic scripts. In: *Second international conference on computer and communication technologies*, New Delhi, pp 703–711
18. Verma M, Sood N, Roy PP, Raman B (2017) Script identification in natural scene images: a dataset and texture-feature based performance evaluation. In: *Proceedings of international conference on computer vision and image processing*, Springer, Singapore, pp 309–319
19. Siromoney G, Chandrasekaran R, Chandrasekaran M (1983) Machine recognition of Brahmi script. *IEEE Trans Syst Man Cybern SMC-13*:648–654
20. Soumya A, Kumar GH (2015) Recognition of historical records using Gabor and zonal features. *Signal Image Process: An Int J* 6:57–69
21. Singh AP, Kushwaha AK (2019) Analysis of segmentation methods for Brahmi script. *DESIDOC J Libr Inf Technol* 39:109–116

# A Fractal Boundary Wideband Antenna with DGS for X-Band Application



Shashi Bhushan Kumar and P. K. Singhal

**Abstract** In the present scenario, the requirement of wideband antenna is increasing day by day. Keeping the view, a symmetrical wideband fractal boundary antenna is proposed in this paper. This proposed antenna with DGS structure has wideband with a compact size of 25 mm × 25 mm. Proposed antenna was simulated using IE3d antenna simulator software and got good result from 8.5 to 11.7 GHz. Presented antenna is resonating at 10 GHz, achieved wide bandwidth of 3.2 GHz, which is quite high. To validate the simulated result, the proposed antenna was fabricated, tested, and very good result has been achieved. Simulated and measured result has been compared, and there is a little deviation in the results due to minor changes in the physical dimension. This antenna can be used for X-band application.

**Keywords** Fractal geometry · Wideband · Fractal boundary · X-band · DGS

## 1 Introduction

In the present scenario, wireless communication systems require antennas with multiband, wideband and small in physical size. A green area of researchers is the microstrip antenna, where research has been carrying out in various directions. Microstrip antenna has many advantages like lightweight, low profile, conformability, easy to fabricate, compatible with MMIC and cheap in cost [1, 2]. With this advantage, the microstrip antenna attracts researcher to work on this. Due to its lightweight and low profile, it can be shaped in any form either planar or non-planar

---

S. B. Kumar (✉)

Department of Electronics and Communication, Bharatividyaapeeth's College of Engineering,  
New Delhi, India

P. K. Singhal

Department of Electronics, MITS Gwalior, Gwalior, M.P., India

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication*

*Engineering*, Lecture Notes in Networks and Systems 106,

[https://doi.org/10.1007/978-981-15-2329-8\\_14](https://doi.org/10.1007/978-981-15-2329-8_14)

structure. As it is conformal to irrespective of any surfaces, it can be used on curved and plain surfaces as well. One of the big disadvantages of conventional microstrip antenna is narrowband, and bandwidth is about 2–3% of center frequency. However, the requirement of bandwidth is increasing day by day as per new technologies. To meet the requirement of multiband and wideband, a fractal geometry has been introduced in 1880 by Mandelbrot [3, 4]. Fractal boundary geometry is a novel concept that has been used in very few geometry. Fractal geometry has been more popular due to its excellent feature of self-similarity and self-affinity properties. Fractal geometry miniaturizes antenna size [5–7]. By applying the miniaturization technique, the size of the antenna gets reduced. It has been seen in many published papers, enhancement in the bandwidth was not much introduced and there was much focus on circularly polarized characteristics [8–10]. Bandwidth enhancement has been discussed with the capacitive loading effect and EBG structure [11, 12]. Importance of bandwidth has been a great role in data transmission. This proposed antenna achieves 3200 MHz bandwidth.

## 2 Antenna Design and Analysis

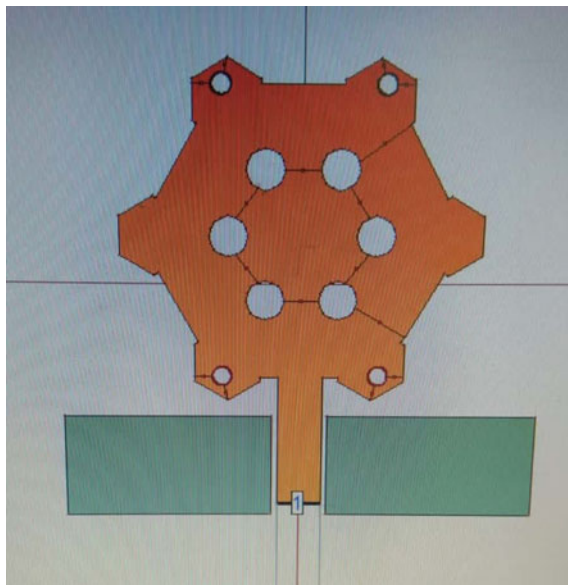
A poly fractal boundary antenna has been introduced. Proposed antenna was designed with hexagon shape on the boundary of the antenna. To get better result, ten circular holes were cut in the antenna DGS structure. There are six holes at the center of antenna and four holes at the center of hexagon polynomials with 2 mm and 1 mm diameter, respectively.

Dielectric constant of substrate: 4.4, width of patch: 25 mm, length of patch: 25 mm.

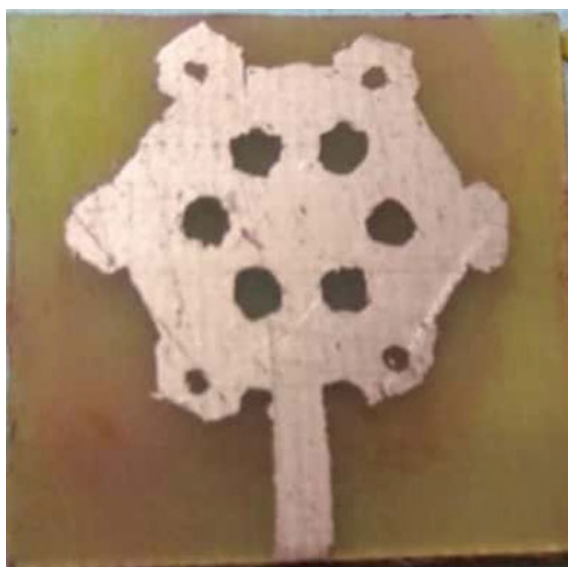
Height of substrate: 1.6 mm, feed width: 2.4 mm (Figs. 1, 2 and 3).

## 3 Simulated and Measured Results

Proposed antenna was simulated using IE3D simulator, and return loss curve is shown in Fig. 4. A wide bandwidth has been achieved. The proposed antenna was fabricated with FR4 material and tested using a vector network analyzer (Fig. 5). Figures 6 and 7 show that simulated 2D radiation pattern and measured radiation pattern, respectively. They are in good agreement in the simulated and measured results.



**Fig. 1** Poly fractal boundary antenna



**Fig. 2** Top view of fabricated antenna



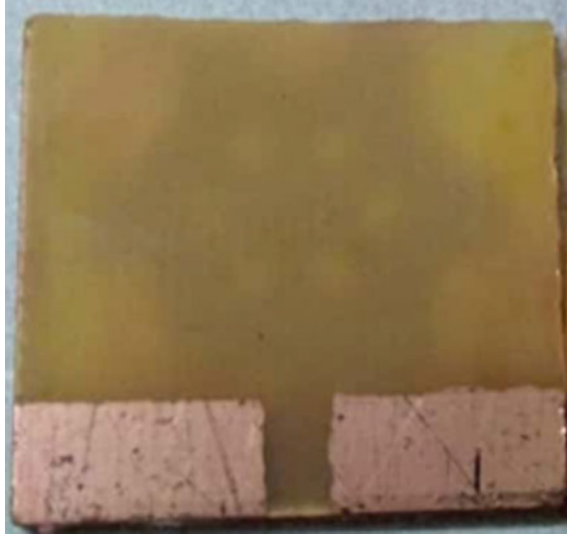


Fig. 3 Bottom view of fabricated antenna

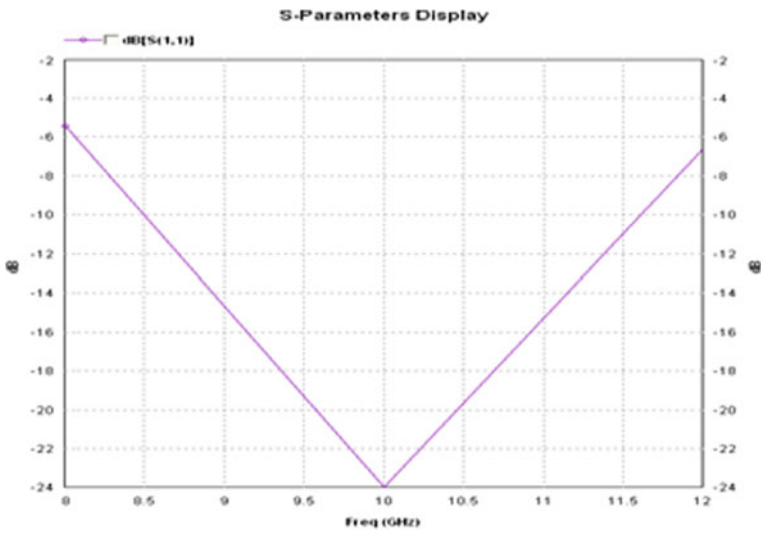


Fig. 4 Return loss versus frequency curve

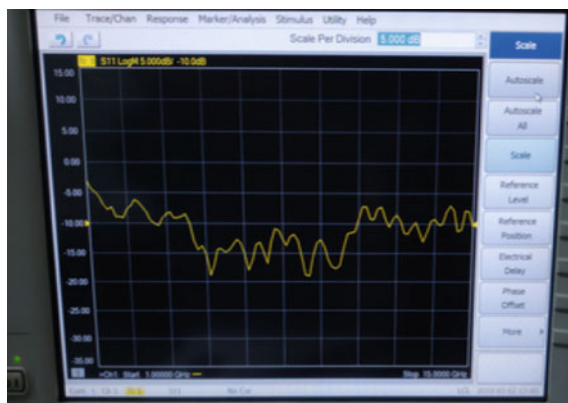


Fig. 5 Measured return loss versus frequency curve

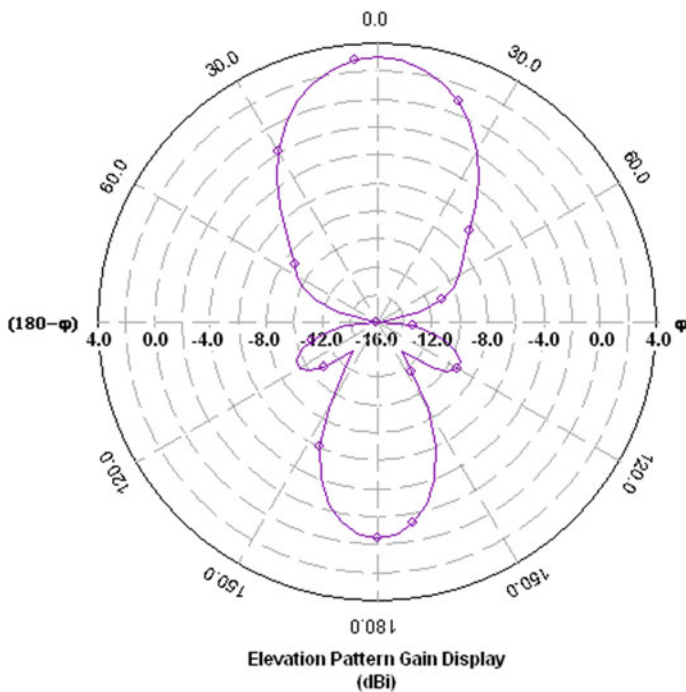


Fig. 6 Radiation pattern of proposed antenna

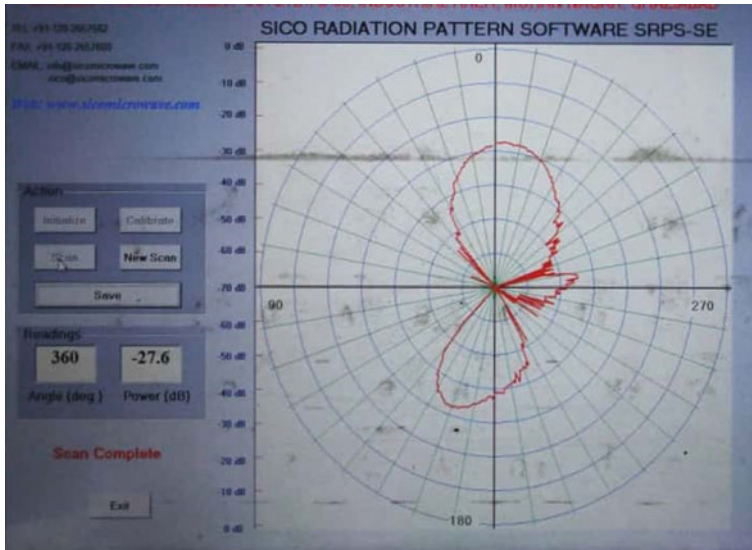


Fig. 7 Measured result of radiation pattern

## 4 Conclusion

A symmetrical fractal boundary antenna was proposed for X-band application. Antenna was designed and simulated. Impedance bandwidth of this antenna is about 3.2 GHz, and the same antenna was fabricated, tested. Measured and simulated result was compared. It was seen there is a minor difference in the results due to changes in physical dimensions while fabricating manually.

## References

1. Garg R, Bahl I (2001) Microstrip antenna design handbook. Arcade House Publication, YOP, pp 1–68
2. Balanis CA (1997) Antenna theory, 2nd Edn, Wiley, YOP, pp 28–79
3. Falconer K (1990) Fractal geometry mathematical foundations and application. Wiley
4. Werner, DH, Ganguly S (2003) An overview of fractal antenna engineering research. IEEE Antennas Propag Mag 45(I, February):38–57
5. Naghshvarian Jahromi M (2008) Novel miniature semi-circular-semi fractal monopole dual band antenna. J Electromagn Waves Appl 22(2–3):227–237
6. Kumar SB, Singhal PK (2012) Performance tested of microstrip fractal antenna for wireless communication. CIIT Int J Wirel Commun Online: ISSN: 0974-9740, DOI: WC12201201
7. Kumar SB, Singhal PK (2012) On the investigation of a Sierpinski's carpet microstrip fractal antenna. Int Eng Sci Adv Technol, India, 2(March-April):200–203
8. Reddy VV, Sarma NVSN (2015) Poly fractal boundary circularly polarized microstrip antenna for WLAN/Wi-MAX wireless application. Def Sci J 65:379384. <https://doi.org/10.14429/dsj.65.8905>

9. Kumar SB, Singhal PK (2018) Investigation of circularly polarized fractal antenna At 2.1 GHz for LTE application. *Int J Eng Res Dev*, 14(January):44–50
10. Reddy VV, Sarma NVSN (2014) Compact circularly polarized asymmetrical fractal boundary microstrip antenna for wireless applications. *IEEE Antennas Wirel Propag Lett* 13:118–120
11. Kumar SB, Singhal PK Fractal geometry with enhanced bandwidth using periodically capacitive loading structure
12. Qu D, Shafai L, Foroozesh A (2006) Improving microstrip patch antenna performance using EBG substrates. *IEEE Proc Microw Antennas Propag* 153(6):558–563

# An Approach to Automated Spam Detection Using Deep Neural Network and Machine Learning Classifiers



Shubham Vashisth, Ishika Dhall and Garima Aggarwal

**Abstract** This paper presents a deep neural network model for performing spam detection. Unlike conventional machine learning models like naïve Bayes, support vector machines, a deep neural network is immune to various fluctuating environments. This paper also proposes the application of CountVectorizer in order to perform feature extraction on the text-based data. In order to increase the accuracy score of the proposed model, hyperparameter tuning has also been done. This paper also compares the accuracy of the proposed deep neural network to various machine learning classifiers like logistic regression, support vector machine,  $k$ -nearest neighbor, Bayes, etc. Experimental results of this paper show that the proposed deep neural network model is able to outclass all other machine learning models in terms of achieved accuracy score, and naïve Bayes classifier is the most efficient model with respect to its computation cost.

**Keywords** Artificial intelligence · Supervised machine learning · Deep neural network · Naïve Bayes classifier · Artificial neural network · Spam detection

## 1 Introduction

A spam message is basically a collection of uninvited bulk messages, mostly used for marketing and promotion directed via the system of electronic messaging. Spam messages are one of the major contributors in making the systems insecure and vulnerable by prompting the users to consume a malware affected entity.

Many solutions are constantly proposed with the intention to make the task of detecting a spam message more precise and efficient as explained in [1]. Even though

---

S. Vashisth (✉) · I. Dhall · G. Aggarwal  
Department of CSE, Amity University, Noida Sector-125, Noida, Uttar Pradesh, India  
e-mail: [shubham.vashisth.delhi@gmail.com](mailto:shubham.vashisth.delhi@gmail.com)

I. Dhall  
e-mail: [ishikadhall11@gmail.com](mailto:ishikadhall11@gmail.com)

G. Aggarwal  
e-mail: [gmehta@amity.edu](mailto:gmehta@amity.edu)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_15](https://doi.org/10.1007/978-981-15-2329-8_15)

a perfect solution to this problem does not exist, diverse approaches and technologies are being applied to this problem, including techniques like blacklisting, content filtration, greylisting, etc., in order to achieve the desired goal. The supreme governing factor that classifies an electronic mail or a message as a spam is based on its solicitation with the user; the messages that are not relatable or associated with the user are the ones that are being placed in the spam category. Other factors that influence this classification are the usage of an IP address that was previously used for transmitting a spam message, having low open rates, having low mailbox usage, deceptive subjects, often used for spam trigger words or phrases, etc. Using the concept of machine learning, it is conceivable to crack the problem of binary classification of a message as spam or ham, i.e., not spam by making the model learn from thousands of examples. This approach enables the model to perform classification tasks in a more efficient and robust scheme without conceding on the accuracy of the application.

The main contribution of this paper is the proposal of deep neural network model which is able to perform the task of spam detection with high accuracy score. This paper also deals with the application of CountVectorizer, based on the process of tokenization which allows to polish up the process of feature extraction on the text-based dataset. Also, hyperparameter tuning has been done for the proposed model in order to increase its accuracy and ultimately reduce the loss of the model. Another contribution of this paper is the related comparison of the proposed deep neural network model to the various machine learning classifiers for the task of detecting spam messages.

The rest of the paper has been organized as given: Sect. 2 presents the preliminaries; Sect. 3 discusses the related works; Sect. 4 is proposing the methodologies and Sect. 5 presents the result. Finally, conclusion is discussed in Sect. 6.

## 2 Preliminaries

### 2.1 *Naïve Bayes Classifier*

Multinomial naive Bayes classifier is a supervised learning algorithm that is based on the notion of prior beliefs and assumes independence among predictors  $X$ , with an aim to estimate the most probable hypothesis. This classifier has shown highly escalating results on large text-based datasets than many other classifiers like logistic regression, SVM, decision tree, etc. This can be explained by the fact that naive Bayes works based on Bayes' theorem which is further based upon the concept of conditional probability as explained by [2]. The algorithm scans the message and searches for the contents, i.e., the tokens and then matches it against its two-word lists to measure the chances of an  $i$ th word contributing toward a message being a spam.

## 2.2 Artificial Neural Network

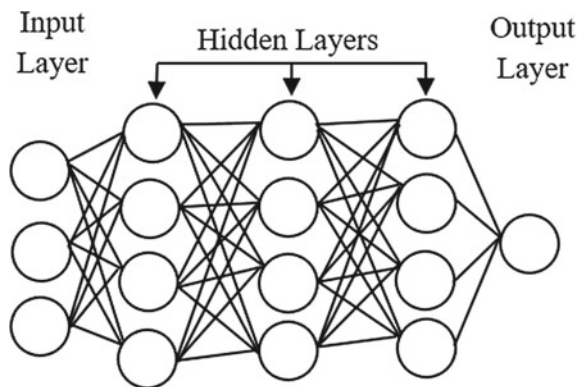
Artificial neural networks have the ability to outclass all the other classifiers. According to [3], massive iteration numbers allow the network to discover the optimum values of weights.

The deep and dense fully connected layers can be appended in an existing network as depicted in Fig. 1 in order to increase the depth of feature extraction and advance the overall accuracy of the model.

## 3 Related Work

Most of the existing work in the area of spam detection prefers Bayesian classification and artificial neural network approach of data mining with pattern classification of e-mail content. Accuracy score of multilayer perceptron and naïve Bayes [4] was not quite impressive, and it could have been improved by tuning parameters and applying data preprocessing techniques. Top reasons to make use of various preprocessing techniques are complexity reduction, faster training of large datasets, easy implementation, and improved accuracy score of the given model. According to the existing work in spam e-mail detection [5], the classification methods like support vector machines and naïve Bayes have been used proving SVM to be inappropriate for binary decision problems. Some works [6] performed an investigation on the efficacy of feature selection methods on a spam dataset and suggested that feature selection approach can be divided into two approaches of feature ranking and subset selection which was a new way of reducing complexity. It was deduced that feature selection removed the unnecessary attributes, hence promoting the improved accuracy score. In order to propose a binary classifier, the technique of learning syntax of each message using word vector training mode has been used too. Some work [7] was focussed on implementation of deep learning techniques due to issues like spam

**Fig. 1** Fully connected deep neural network



drifting and information fabrication and received low accuracy scores. This can be improved by using CountVectorizer which provides a simpler way to both collect the text and tokenize a collection of text document. It encodes the new word using the built vocabulary and improves the score further.

## 4 Methodology

Machine learning algorithms are used to provide the systems with an ability to learn. This ability to enable machines to learn artificially is made possible by using a tremendous amount of data and various statistical techniques in order to extract features from the recorded data. The science behind machine learning is closely related to that of mathematical optimization, computer statistics, and even data mining. The following steps were adopted in order to generate the models to perform binary classification of messages as spam or ham:

(i) Data preparation and preprocessing (ii) training (iii) testing (iv) hyperparameter tuning.

### 4.1 Data Preprocessing and Preparation

A file with 5572 different sample messages along with their corresponding label was selected. As these machine learning algorithms work with numerical data, it is essential to convert or remap the respective value of  $Y$  for each unique value of  $X$ .

Since  $Y$  can only have two binary possibilities of being either a spam or ham, therefore a spam label is mapped as 1 and a ham label is mapped as 0 for all unique messages or  $X$  as shown in Fig. 2.

### 4.2 Training

This step is also considered as the bulk of machine learning. The model is trained using supervised machine learning approach as explained in [8]. The classic machine learning algorithms like logistic regression, naïve Bayes, random forest, etc., were initially designed in order to work with numerical figures. Therefore, there is a need for conversion of initial text data into the numerical data without losing the needed information before feeding it into the classifying algorithms. This problem can be solved by using the CountVectorizer function as presented in [9]; this function is provided by the scikit-learn library which is supported by python programming language.



```
spam_data.head()
```

	Label	Message
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

```
spam_data['Label']=spam_data['Label'].map({'ham':0,'spam':1});  
spam_data.head()
```

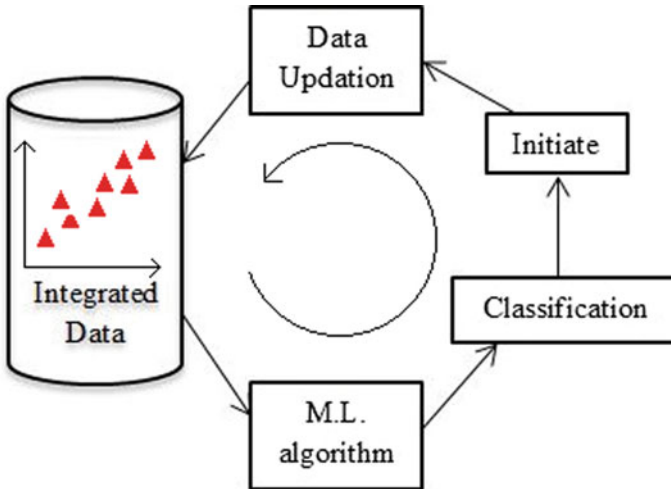
	Label	Message
0	0	Go until jurong point, crazy.. Available only ...
1	0	Ok lar... Joking wif u oni...
2	1	Free entry in 2 a wkly comp to win FA Cup fina...
3	0	U dun say so early hor... U c already then say...
4	0	Nah I don't think he goes to usf, he lives aro...

Fig. 2 Preprocessing data

CountVectorizer enables us to tokenize a pool of text data in order to construct a vocabulary of known words and encode the text data using the same vocabulary as depicted in Fig. 3.

```
C=CountVectorizer();  
  
initial_text='We are fine, how are you all?';  
example=C.fit_transform(['We are fine, how are you all?']);  
  
print('Initial text: ',initial_text);  
print('Tokens generated for the text: ',C.get_feature_names());  
print('Encoded intial text data: ',example.toarray())  
  
Initial text: We are fine, how are you all?  
Tokens generated for the text: ['all', 'are', 'fine', 'how', 'we', 'you']  
Encoded intial text data: [[1 2 1 1 1 1]]
```

Fig. 3 Tokenizing using CountVectorizer



**Fig. 4** Batch or offline learning

After generating the required data, the machine learning model is trained using batch learning where all the available data is fed to the system at once, and the whole process can be automated; this learning mechanism is also referred as offline learning as shown in Fig. 4. This batch learning system, as discussed in [10], is adopted for training by feeding 80% of the whole dataset to each and every classifier that has been selected, i.e., logistic regression, artificial neural network,  $k$ -nearest neighbor, support vector machine, decision trees, random forest, and naïve Bayes classifier.

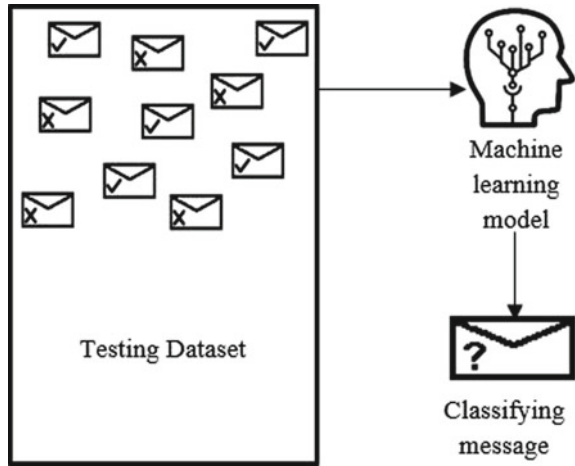
The goal now is to train the chosen models in a way that it improves its prediction with every training step or iteration. The more data the model will be trained upon, the more accurate the prediction will be. The basic explanation of this phase goes by the equation  $y = m * x + c$ , where  $y$  is the output,  $x$  is the input, and these two variables cannot be altered, but  $m$ , the slope, and  $c$ , the  $y$ -intercept, can be altered with every iteration based on every single data from the complete dataset. The collection of  $m$  forms a matrix  $W$  (weight matrix) and the collection of  $c$  forms the  $C$  matrix. Both matrices are expected to be updated for each unique value that is being fed to these models.

### 4.3 Testing

Out of the complete dataset, the left 20% of the dataset is being used to test the accuracy of all the selected classifiers.

The testing phase allows us to test the model against data that has never been used before for training as shown in Fig. 5. This is done by evaluating values of  $Y$ , i.e., output label for corresponding values of  $X$ , i.e., input message and then comparing

Fig. 5 Testing the model



the generated values of  $Y$  to the original message label. This phase is important to check how the designed model will work in the practical world.

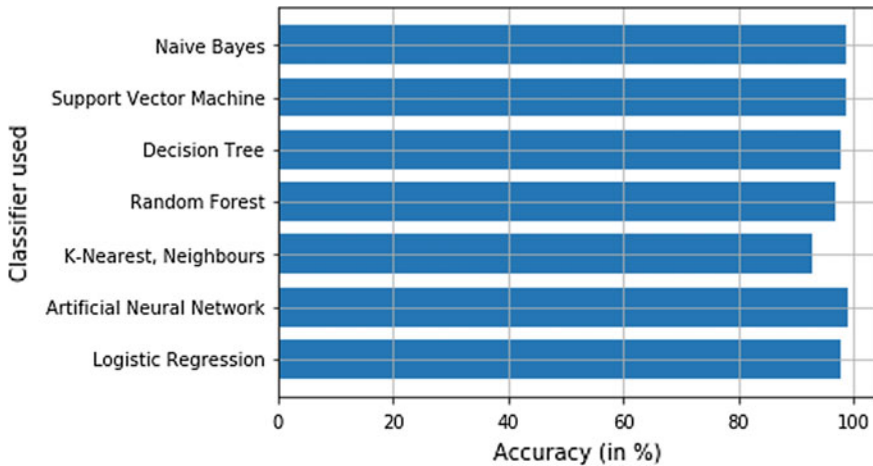
### 4.4 Hyperparameter Tuning

Since the machine learning classifiers are depended on hand-tuned parameters, it is always possible to be able to improve the pre-examined accuracy of the existing models. Various parameters such as learning rate, number of hidden layers, optimizer, epochs in neural networks, the value of  $k$  in  $k$ -nearest neighbors, soft-margin constant  $C$  in support vector machine, etc., can heavily affect the accuracy rate for the machine learning models. Hence, by tuning such parameters, much higher accuracy scores can be obtained for the task of spam detection.

## 5 Result

The proposed deep neural network model was able to achieve a high accuracy score as compared to some of the conventional machine learning models. This accuracy was made achievable by determining the superlative hyperparameters for the selected model.

Figure 6 shows the accuracy scores that were achieved in case of each separate model, according to which artificial neural network was able to achieve the highest score of 0.9892 followed by naïve Bayes classifier with a score of 0.9883. As observed, naïve Bayes classifier is giving an accuracy close to that of artificial neural network, but it is feasible to use naïve Bayes due to its reduced computational cost



**Fig. 6** Accuracy scores

and time of processing. Table 1 shows the values of accuracy scores achieved by the models. Conversion of initial text data into the numerical data which is being encoded as integer or floating values also served as an important step in extracting features without losing the needed using CountVectorizer.

Table 2 shows the value of hyperparameters that were being selected. These hand-tuned features play a vital role in defining the learning capability of the network that varies from problem to problem. The goal is to minimize the loss followed

**Table 1** Accuracy comparison of the models

Classifying model	The accuracy achieved (in %)
<i>K</i> -nearest neighbors	92.73542600896862
Random forest	96.95067264573991
Decision tree	97.75784753363229
Logistic regression	97.9372197309417
Support vector machine	98.7443946188341
Naïve Bayes	98.83408071748879
Artificial neural network	98.9237668161435

**Table 2** Hyperparameters of the deep neural network

Hyperparameter	Configuration
Number of hidden layers	3
Number of nodes per layer	{100, 50, 25}
Activation function	Logistic
Learning rate	0.001
Optimizer	Adam

by the computation cost of generating the model, keeping in mind the problem of overfitting. After trying various values of given parameters, it was observed that the above-mentioned values for given hyperparameters are the most optimal values.

## 6 Conclusion

The generated deep neural network model is able to perform spam detection with high accuracy rate. The features were extracted from the data using concept of CountVectorizer. Hyperparameter tuning allowed the evaluation of the most optimal hand-tuned features. Also, upon comparing the accuracy score of all the models, the proposed deep neural network model was able to give the highest accuracy score in comparison with the conventional machine learning models for the problem of spam detection. As the future work, various data preprocessing strategies can be adopted in order to increase the accuracy score. Also, the implementation of advanced forms of recurrent neural networks like LSTMs can be done which are much more robust for the problems that lie in the domain of natural language processing.

## References

1. Spirin N, Han J (2012) Survey on web spam detection: principles and algorithms. *ACM SIGKDD Explor Newslett* 13(2)
2. Pang, B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: *Proceedings of the ACL-02 conference on empirical methods in natural language processing*, vol 10. Association for Computational Linguistics
3. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *nature* 521(7553)
4. Sharma AK, Prajapat SK, Aslam M (2014) A comparative study between Naive Bayes and neural network (MLP) classifier for spam email detection. In: *IJCA Proceedings on national seminar on recent advances in wireless networks and communications. Foundation of Computer Science (FCS)*, vol 2, pp 12–16
5. Youn S, McLeod D (2007) A comparative study for email classification. In: *Advances and innovations in systems, computing sciences and software engineering*. Springer, Dordrecht, pp 387–391
6. Parimala R, Nallaswamy R (2011) A study of spam email classification using feature selection package. *Glob J Comput Sci Technol*
7. Wu T et al (2017) Twitter spam detection based on deep learning. In: *Proceedings of the Australasian computer science week multiconference ACM*
8. Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: a review of classification techniques. *Emerg Artif Intell Appl Comput Eng* 160:3–24
9. Kulkarni A, Shivananda A (2019) *Converting text to features*. In: *Natural language processing recipes*. Apress, Berkeley, CA, 6796
10. Keskar NS et al (2016) On large-batch training for deep learning: generalization gap and sharp minima. *arXiv preprint [arXiv:1609.04836](https://arxiv.org/abs/1609.04836)*

# Analysis and Implementation of IWT-SVD Scheme for Video Steganography



Urmila Pilonia and Prinima Gupta

**Abstract** In the world of computer, the skill of sending and displaying secret data particularly in public places faced several challenges and had received more attention now. Protecting information on communication media is a vital need in information transmission technology. We can protect information through encryption process over the transmission media but due to advanced computing techniques, encrypted information can easily be detected and decrypted. So, for avoiding the unauthorized user from accessing information, we need some more advanced techniques. For this purpose, we use steganography which is a technique to conceal secret message over communication networks. In this paper, IWT-SVD scheme is proposed to conceal watermark image in video cover file. Concealing watermark in video cover file gives higher concealing capacity. It has been shown that concealing watermark in HH and LL sub-band results in good perceptual quality, more robustness, and less computational cost. Simulation results also show that this new scheme outperforms adaptive steganography based on IWT-SVD in term of PSNR, MSE, and concealing capacity.

**Keywords** Video steganography · Transform domain · Spatial domain · Integer wavelet transform (IWT) · Singular value decomposition (SVD)

## 1 Introduction

With the help of communicating device and Internet, a lot of information can be sent from one location to another location. When secure data is sent through any communication channel, it can be altered by unauthorized persons. The security measures need to get enhanced to tackle interception, i.e., the illegal alteration done by illicit persons. Although there are many technologies exist for the safe transfer

---

U. Pilonia (✉) · P. Gupta  
Department of Computer Science & Technology, Manav Rachna University,  
Faridabad, Haryana, India  
e-mail: [urmila@mru.edu.in](mailto:urmila@mru.edu.in)

P. Gupta  
e-mail: [prinima@mru.edu.in](mailto:prinima@mru.edu.in)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_16](https://doi.org/10.1007/978-981-15-2329-8_16)

**Table 1** Description of watermarking and steganography

Term description	Watermarking	Steganography
Objective	Copyright protection	Hidden communication
Requirements	Robust watermarking	Hiding capacity, perceptual quality, and robustness must be good
Source file	Image/video	Text/audio/image/video
Secret message	Watermark	Text/image
Secret keys	May exist depends on algorithm used	Depends on embedding algorithms
Output file	Watermarked object	Stego file
Security	Depends on watermarking technique	Depends on embedding algorithms, must be high
Perceptual quality	Transparency depends on type of application	Invisible
Robustness	Robust against image processing attacks	Robust against attacks
Attacks	Signal processing operations	File only, compression attacks, structural attacks, visual attacks, statistical attacks
Goal failed	Watermark is erased or exchanged	Communication is detected

of information like cryptography, watermarking, digital signature, fingerprinting, and steganography [1], watermarking protects copyright material from the attackers. Watermarking protects secret data by embedding information in the form of copyright protection while steganography embeds secret information inside a cover file. Steganography is secret communication, which hides even the existence of a message inside cover medium from the outside world [2]. By using steganography algorithm, we remove the unnecessary bits in the cover file and add top secret information into those gaps. A detailed explanation of watermarking and steganography [3] is shown Table 1.

### *1.1 Different Steganography Techniques*

There are many steganography techniques exist like distortion techniques, cover generation technique, spatial domain technique, and transform techniques. In distortion technique, stego object is formed by relating a series of modification to cover file [4]. In cover generation technique, a cover is generated for the purpose of hiding the secret information. Based on the domain, secret information can be concealed inside time domain as well as frequency domain of the signal. In case of spatial domain steganography, we change some bits of the cover file from the bits of secret information. The simplest spatial domain technique is the least significant bit. LSB offers

high embedding capacity and provides simplest way to control the quality of stego file but easy to detect because of its simplicity [5, 6]. After LSB technique, many more spatial domain techniques played a vital role in information hiding like PVD, EBE, RPE, MVD, etc. In transform domain techniques, the cover file goes through various transformations before hiding the secret data inside it. These techniques hide information in the frequency domain rather than time domain. Hiding information in frequency domain is much more complex and provides more security and robustness. These techniques embed data in those portions of the cover file that remain less opened to firmness, cropping, and image processing. A number of transform domain techniques are may be lossless or lossy, and they are independent of image format [2, 7, and 8]. Classification of these techniques includes DFT, DCT, DWT, IWT [9], etc.

## 2 Related Work

Steganography is embedding its existence [1]. Invisible ink is used to mark certain letters in newspaper and also added sub-perceptible echo in audio at certain places then sent it to third party. DWT technique [10] is used to hide secret information inside the cover image. Here, we divide cover file into LL, LH, HL, and HH coefficients, these bands keep up a correspondence to common rough calculation, horizontal, vertical, and diagonal facts of cover file, respectively. Data is embedding into LL and HH coefficient of cover file because covariance of these two bands is small as compared to LH and HL sub-bands. Image steganography using IWT and GA [9] is used for concealing information in integer coefficients with the help of mapping function, hiding information using DWT technique [11], and then optimizing using PCA. PCA removes correlation between the information. In combination with wavelet coefficient technique, we use PCA for efficient hiding of information. Cover message [12] is decomposed by applying DWT technique, and then less significant wavelet band is replaced by secret message. Review of different spatial steganography techniques like LSB, MSB, and PVD [13] is based on different parameters. Transformation domain techniques like DFT, DCT, and DWT are compared by author based on various parameters [14]. Data is embedded in secret image using IWT steganography technique [15]. Instead of using DCT technique, IWT is used to embed secret image which provides lossless compression. Secret data is embedded inside audio cover file [16] using IWT technique which is lossless because of integer nature of audio coefficients of cover file. After literature survey, analysis of implemented steganography techniques is done in given Table 2.



**Table 2** Analysis of already existing steganography techniques

Sr. no.	Title of paper	Technique used	Advantage	Disadvantage
1	A blind CDMA image watermarking scheme in wavelet domain	CDMA and DWT	High resiliency with good visual quality of watermark images	Coefficient in this technique are in float numbers
2	Steganographic method based on IWT and GA	IWT and GA	High embedding capacity	More execution time
3	Video steganography approach for effective	SVD and Advanced LSB	Good robustness and perceptual quality	Computational cost is high
4	Hiding data in video using discrete wavelet transform and principal component analysis	Multilevel DWT and PCA	Good imperceptibility and robustness is achieved	Low embedding capacity
5	Integer wavelet transform based steganographic method using OPA algorithm	IWT and OPA	Reduce error difference among original coefficients values and modified values	Complex because of additional keys
6	Integer wavelet transform for embedded lossy to lossless image compression	IWT and EZW	Secret information embedding with lossless image compression	Computational cost increases
7	A new approach to hide data in color image using LSB steganography technique	LSB and AES	Provide double-layer security	Does work well on compressed images
8	Simulation-based watermarking for confidential data security for video signal	DCT and JPGE image compression	Embed secret information without any loss and with good visual quality	NC varies with quality of secret image
9	Wavelet-based color video steganography	DWT and LSB	Large embedding capacity	Low PSNR value

### 3 Parameters in Video Steganography

As in process of hiding, information inside cover media quality may degrade. So, to measure the quality of stego media, some parameters are used. These parameters are.

#### 3.1 Peak Signal-to-Noise Ratio (PSNR)

PSNR determines quality of stego video. Higher value of PSNR represents better quality of video. PSNR is the ratio of noise in original video to the stego video [16, 5, 17, 18, and 19].

#### 3.2 Mean Square Error (MSE)

MSE is a factor to find the excellence of stego object. This method is inverse of the above method [5, 3, 19].

$$\text{PSNR} = 10 \log_{10} \left( \frac{\text{MAX}_I^2}{\text{MSE}} \right)$$

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2$$

Here,  $M$  and  $N$  are rows and columns of cover image,  $I$  and  $K$  are cover file and stego image, respectively. Where  $\text{MAX}$  is the range of pixel values ( $R = 255$  for RGB images).

## 4 Proposed Work and Discussion

Existing steganography technique such as LSB, EBE, PVD, DFT, DCT, and DWT has some shortcoming in terms of payload, computational cost, security, perceptual quality, robustness, and loss of information during compression [18]. If we try to improve one parameter, then other parameters got affected because of the changes. Embedding a large amount of secret information in cover file results in low perceptual quality of stego video as well as computational cost increases. Existing results of DCT technique achieved PSNR of 38.26 db and that of DWT is 35.17 db [2] which is very low. Low value of PSNR results in less robustness and poor quality of stego file. This poor quality of stego file may also be attacked by hackers. IWT technique

is the one which balances all these parameters, consumes less power, and is more efficient in terms of area. Hardware implementation of IWT is also very easy because of its integer nature. With IWT, cover file is converted into four coefficients LL, HH, LH, and HL. HH coefficients have more noise in it, and LL coefficients have useful information. So, secret information is embedded inside HH coefficients of the cover file. To validate dispersion result, compute covariance ( $\sigma_{X_n X_m}$ ) for all the four sub-bands, by taking two at a time [10]. Covariance is statistically expressed as follows:

$$\sigma_{X_n X_m} = (X_n X_m)^c - (X_n)^c (X_m)^c$$

From the computation, it is proved that covariance for LL and HH sub-bands are two smallest values equated to values for the same in all other combinations of sub-bands.

SVD is a linear algebra transform which factorizes real or complex matrix having applications in numerous fields of image processing [8]. Digital image is characterized in the form of matrix, entries in matrix give intensity value of each pixel in image, and SVD of an image  $M$  with dimensions  $m \times m$  is given below:

$$M = USV^T$$

Main benefits of SVD to employ in steganography are [20]:

- 1 Little changes in singular values do not affect the quality of the image.
- 2 Singular value of image has high stability that is they do not alter after many attacks.

#### ***4.1 Embedding Algorithm***

Following steps given below explain the embedding process:

1. Take video as cover file and divide the video into  $N$  number of frames.
2. Take input as a secret image into  $64 * 64$  blocks at sender side which is to embed in cover medium.
3. Using IWT scheme, find frequency domain representation of frames and convert each frame into four sub-bands LL, HH, LH, and HL that is into low, high, and middle frequency sub-bands (Fig. 1).
4. Conceal the secret image into HH and LL sub-band by applying IWT and SVD, respectively, on the specific frame resulting into stego file.
5. Then apply inverse IWT and SVD on stego file to extract the secret image.
6. To find the quality of scheme, apply performance matrix PSNR and MSE.

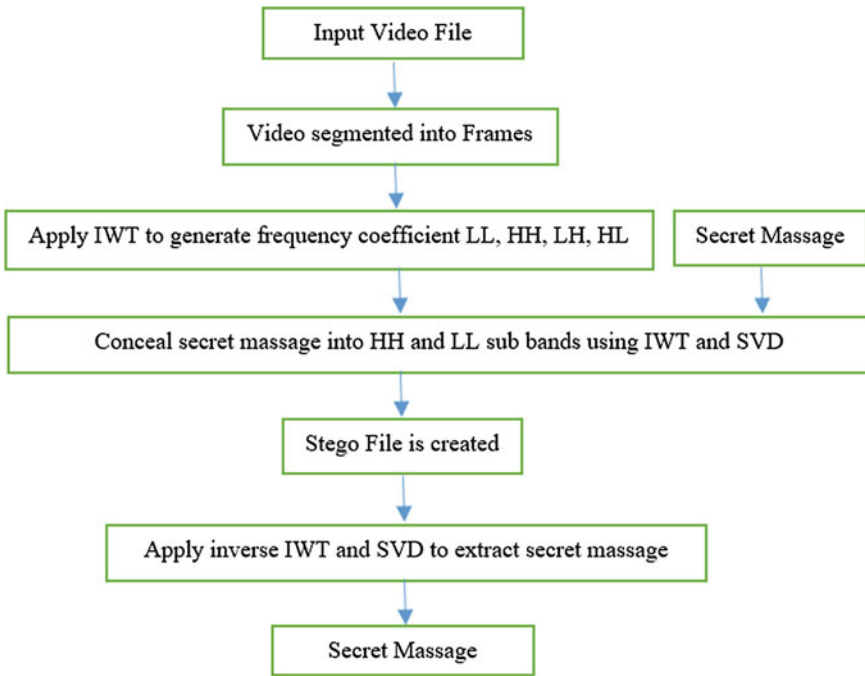
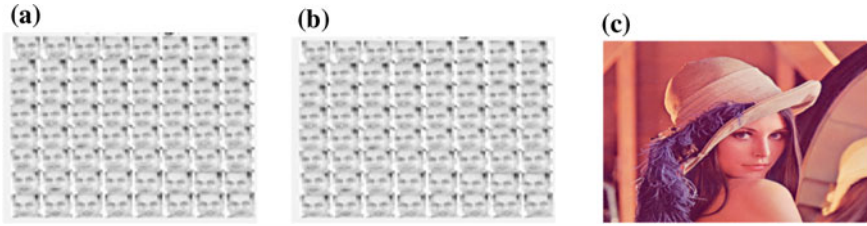


Fig. 1 Embedding flowchart

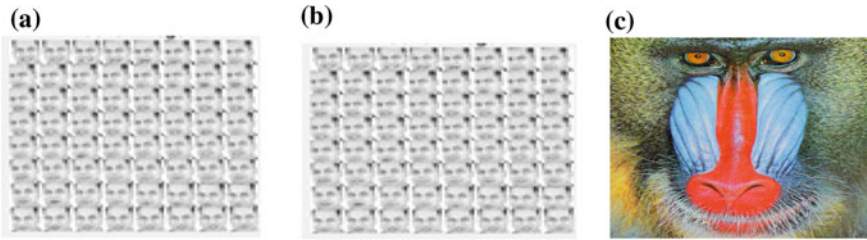
## 5 Experimental Results

Video is taken as the cover file, and the watermark is taken in the form of 64 \* 64 images. Cover video file is divided into n number of frames. Then, concealing of secret information is done inside random frames so that the attacker will not be able to find where the watermark is stored and into how many frames. In this way, hiding capacity as well as robustness of the system is improved. Common landmark images like “Lena,” “Pepper,” “Nadal,” and “Mandrill,” etc., are taken as watermark. These four different watermark images are concealed into the same video cover file into different frames to measure the performance of planned method. For measuring the performance of system concealing capacity, PSNR and MSE parameters are used (Figs. 2, 3, 4, and 5).

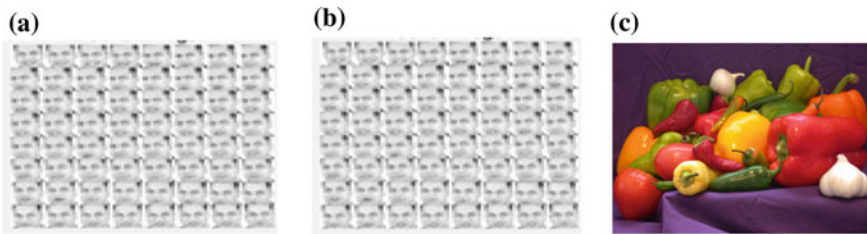
After concealing watermark into a cover object, the following results are obtained in MATLAB. Graph is also created to calculate the variation in the value of PSNR and MSE. PSNR represents the quality of stego object. Higher the value of PSNR good is the quality of stego object. Lower the value of MSE means less is the error present between original and watermarked object. Experimental results show that there is a little variation in the values of performance matrixes when different secret files are concealed in the same cover file which does not affect the quality of our proposed scheme (Figs. 6 and 7; Table 3).



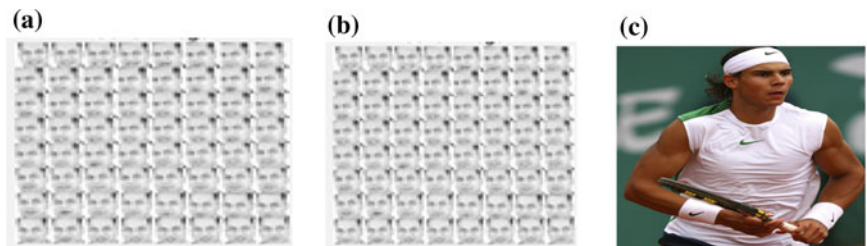
**Fig. 2** a Cover file before embedding, b cover file after embedding, and c extracted watermark



**Fig. 3** a Cover file before embedding, b cover file after embedding, and c extracted watermark



**Fig. 4** a Cover file before embedding, b cover file after embedding, and c extracted watermark



**Fig. 5** a Cover file before embedding, b cover file after embedding, and c extracted watermark

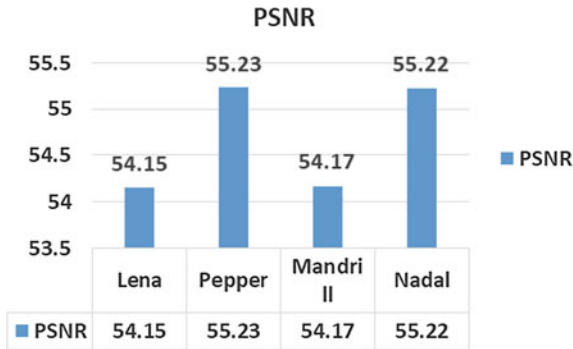


Fig. 6 PSNR graph for watermarked image

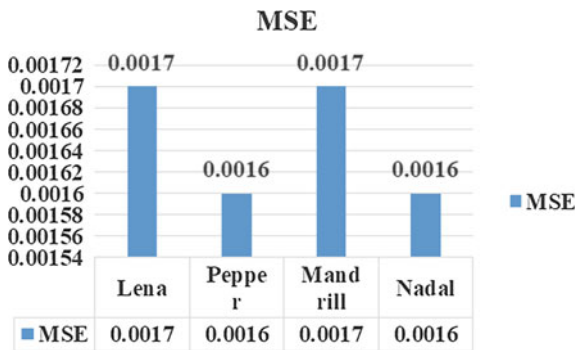


Fig. 7 MSE graph for watermarked image

Table 3 Experimental results

Test image (64 * 64)	Capacity (bytes)	PSNR	MSE
Lena	4,75,136	54.15	0.0017
Pepper	2,94,912	55.23	0.0016
Mandrill	6,38,976	54.17	0.0017
Nadal	4,21,888	55.22	0.0016

## 6 Conclusion

Implementation of the proposed video steganography IWT-SVD scheme has been done in MATLAB. Purpose of the experimental study is to measure the quality of concealed watermarks in video cover file and to conceal watermark inside it without any modification and damage of information. Simulation results show good performance of the proposed scheme in terms of embedding capacity, PSNR, and MSE for different watermark images. IWT technique takes less time and computational

cost for embedding secret image compared to other transform domain techniques. IWT technique consumes less power and is more efficient in terms of area. Hardware implementation of IWT is very easy because of its integer nature. SVD ensures little change in singular values does not affect the quality of watermark and has high stability against different types of attacks.

## References

1. Petitcolas FAP, Anderson RJ, Kuhn MG (1999) Information hiding-a survey. *Proc IEEE* 87(7):1062–1078
2. Chandel B, Jain S (2016) Video steganography: a survey. *IOSR J Comput Eng* 18
3. Cheddad A, Condell J, Curran K (2010) Digital image steganography: survey and analysis of current methods. *Signal Process* 90(3):727–752
4. Maharshi A, Gupta RP, Shrimal VM (2012) Simulation based watermarking for confidential data security for video signal. *Int J* 2(1)
5. Islam MN, Jannat F, Sultana Z (2017) A new approach to hide data in color image using LSB steganography technique. In: 2017 3rd international conference on (EICT), 7–9 Dec 2017, Khulna, Bangladesh
6. Biradar RL, Umashetty A (2016) A survey paper on steganography techniques. *Int J Innov Res Comput Commun Eng (A High Impact Factor, Monthly, Peer Reviewed Journal)* 4
7. Jayasudha S (2013) Integer wavelet transform based steganographic method using OPA algorithm. *Int J Eng Sci* 2(4):31–35
8. Dave M, Saket Swarndeep J (2017) Video steganography approach for effective embedding of secret data. *IJARIE-ISSN(O)-2395-4396* 3(2)
9. Ghasemi E, Shanbehzadeh J, ZahirAzami B (2011) A steganographic method based on integer wavelet transform and genetic algorithm. In: 2011 international conference on communications and signal processing. *IEEE*
10. Maity SP, Kundu MK (2004) A blind CDMA image watermarking scheme in wavelet domain. In: International conference on image processing, 2004, ICIP'04, vol 4. *IEEE*
11. Hood AA, Janwe NJ (2013) Hiding data in video using discrete wavelet transform and principal component analysis. *Int J Eng* 2:1165–1169
12. Kolakalur A, Kagalidis I, Vuksanovic B (2016) Wavelet based color video steganography. *Int J Eng Technol* 8(3):165
13. Rejani R, Murugan D, Krishnan DV (2015) Comparative study of spatial domain image steganography techniques. *Int J Adv Netw Appl* 7(2):2650
14. Thakur A, Singh H, Sharda S Different techniques of image and video steganography: a review. *Int J Electron Electr Eng* 2
15. Reichel J et al (2001) Integer wavelet transform for embedded lossy to lossless image compression. *IEEE Trans Image Process* 10(3):383–392
16. Giurcaneanu CD, Tabus I, Astola J (1999) Integer wavelet transforms based lossless audio compression. *NSIP*
17. Kharrazi M, Sencar HT (2006) Cover selection for steganographic embedding, image processing. In: *IEEE international conference, Oct 2006, Atlanta, GA*
18. Ansari, AS, Mohammadi MS, Parvez MT (2019) A comparative study of recent steganography techniques for multiple image formats. *Int J Comput Netw Inf Secur* 11(1):11
19. Singh N (2019) High PSNR based image steganography. *Int J Adv Eng Res Sci (IAERS)* 6(1, January)
20. [http://en.wikipedia.org/wiki/Singular\\_value\\_decomposition](http://en.wikipedia.org/wiki/Singular_value_decomposition)

# Handling Sparsity in Cross-Domain Recommendation Systems: Review



Nikita Taneja and Hardeo Kumar Thakur

**Abstract** Cross-domain recommendation Systems (CDRS) is a significant research area which has been target of many companies these days. From last few years, there is amount of publications in CDRS domain including recommendation systems which have been rising sharply alongside information retrieval and machine learning. Recommender systems help companies to specifically identify preferences of the user from the collected data from various sources. These data values are then used to identify user preferences also known as recommendations. In cross-domain, we use the data from one domain such as movies to recommend items in other domains, for instance, Books. There are numerous issues which recommender systems suffer together with sparse data, synonymy, data privacy, algorithm scalability, perspective awareness (context) and cold start problems. Data sparsity is a major issue in recommender systems, especially in the presence of novel users or items, or when user drift exists. This paper reviews recent efforts made for CDRS sparsity and user drift which are prevalent in most CDRSs such as user-based, item-based or knowledge transfer. This paper formalizes the CDRS illustrates sparsity related issues which are addressed in prior works and finally proposes for future research trail.

**Keywords** Cross-domain recommendation systems (CDRS) · Collaborative filtering · Knowledge transfer · Sparsity

## 1 Introduction

Recommendation systems have a significant role to play in how we surf the Internet, e.g., the recommendation given by YouTube decides which video we will see next. This could also have some negative consequences, e.g., if we recommend users only one type of content, they can be easily manipulated [1]. In the Web sites related to

---

N. Taneja (✉) · H. K. Thakur  
Manav Rachna University, Faridabad, India  
e-mail: [nikita@mru.edu.in](mailto:nikita@mru.edu.in)



e-commerce, such as Amazon and many more, exploiting the evaluations of the user would be highly significant regarding many kinds of products, so as to produce a model which is general for the preferences of user. As a matter of fact, there could be many correlations as well as dependencies which can exist between different domain preferences. Despite each item treatment independently such as music and electrical products, the knowledge of the user is obtained inside a domain which is able to be shifted and several other domains can be used to exploit it [2].

One of the promising research areas nowadays is cross-domain recommendation (CDR). This CDR method exploits knowledge from secondary domains such that movies having further user preference data to enhance recommendation on a target domain, for example—books. As depending on a large scope of existing data in most of the cases is a key to ease the troubles of such sparse user-item data in the same target domain. CDRS may also in the same way benefit its data owners by enhancing quality service in same or different domains. The main cross-domain recommendation goal furthermore is not this but, however, a join offer could be given by a system, for example, a movie recommendation, CDs music and books, any video game which is related somehow to the movie. In the same way, in the applications of tourism, a cultural event can also be suggested to a client who is interested to get a room booking in a hotel recommended. Joint recommendations are also offered by some systems in different domains for an item, but collectively speaking, in one domain so as to build a recommendation, the user preferences are only exploited by them on the domain which is targeted. The three most important challenges identified for CDRSs are

- (a) Global correlations existence and their verification for the items in various domains and user preferences.
- (b) Model designing uses the user-based preferences on the domain source for the user preference prediction on the domain targeted.
- (c) For CDRs, the development of the appropriate evaluations.

## 2 CDRS Issues and Challenges

Cross-domain recommendation is a tough but also widely under explored research area. Even though CDR has been considered from various angles and the exact definition of the CDR problem has not evolved yet. A few of the existing works have been examined and classified the in-literature cross-domain recommendation techniques. Significant challenges of CDRS have been described in Table 1.

**Table 1** CDRS challenges and its description

Challenge	Description
Synonymy	When an item or product is represented with more than two names. The CF [3] approach which is based on memory, for example, will be able to treat differently the comedy movie or a comedy film
Privacy	Feeding personal information often ends up in services with better recommendation, but at the same time can hamper data security and privacy. Hence, the recommenders of CF are seen to be highly prone to the issues, for example—privacy and data security [4]
Limited content analysis and overspecialization	The data when available in a limited form can cause the over specialization kind of problems
Scalability	For a typical recommender, it becomes difficult to process a data which is of large scale. Amazon.com is an example of it which suggests products over 18 million and their customers are over 20 million [7, 8]
Latency problem	The latency problem is faced by the recommenders of CF when there is frequent addition of new products to the database in a case where the suggestion is made by the recommenders on the items already rated as items which are added newly and are not yet given any rating
Evaluation and the availability of online datasets	The evaluation criteria and its design along with the suitable evaluation selection metrics are regarded as problems in the systems of recommenders
Context awareness	The upcoming systems of recommenders as envisioned are supposed to make use of data which is contextual and is collected via infrastructure of cellular services and user’s history
Cold start problem	When a not existing before customer enters the system this problem takes place and also in the case when into the catalogue more items are added
Sparsity	The data of larger size which is available regarding the catalogue of products and the users disinclination for the item rating often makes up a matrix of dispersed profile and this can lead to recommendations which are lesser accurate

### 3 Causes and Solutions to Data Sparsity

For a recommendation system, its user data is prepared into a database which is utilized via all recommender systems (RSs) to produce recommendations. These are widely applicable in e-commerce business to offer the customers a good quality and customized recommendations from multiple of different choices.

The collaborative filtering (CF) is known as a second-hand skill to create recommendations. The idea behind collaborative filtering is to combine the ratings of like-minded users/customers. But, the generated matrix of user-item ratings, in general, occurs extremely sparse due to their lack of knowledge or incentive to rate items. Also, in the case of fresh users or fresh items, generally, receive only handful or no ratings at all. Both of the mentioned issues will avert the CF from offering effective recommendations because user's preferences are complex to extract. For example, in the given Fig. 1, each row of the matrix corresponds to one user, and each column correlates with one item. If the user has rated the item, the position in the matrix where the row corresponding to the user intersects with the column corresponding to the item represents the user's rating value for the corresponding item. Such user-item matrix is called a scoring matrix.

Figure 1 is an example of a scoring matrix which tells that the user has not evaluated the item yet, and the final goal of the recommendation system is to forecast the score of not rated items for any one user, and recommend the corresponding item to the user in descending order of the score.

A widely used machine learning technique for low-rank matrix factorization is applied to provide recommendations through a set of user ratings for a given item. Such ratings are presented as a user-item utility matrix shown in Fig. 2 [7] and the missing values are calculated by optimizing utility matrix on a low-rank factorization. The idea of the matrix factorization model is that the information encoded is, firstly, not exactly independent for items inside the columns of the utility matrix, and secondly, for users inside the rows of the utility matrix. We need to optimize the objective function given in Eq. 1 [7].

User/Item	Item 1	Item 2	Item 3	Item 4	Item 5	Item M
User1	NIL	4	2	5	....	NIL
User2	2	1	NIL	3.5	....	4
User3	2.5	0.5	NIL	1	....	3.5
User4	3.5	2.5	3	1	....	NIL
User5	1	NIL	4.5	2.5	....	3.5
User6	1	NIL	3	NIL	....	2.5
.....	....	....	....	....	....	1.5
User N	5	2.5	NIL	1.5	....	1

**Fig. 1** Typical user/item matrix of recommender systems

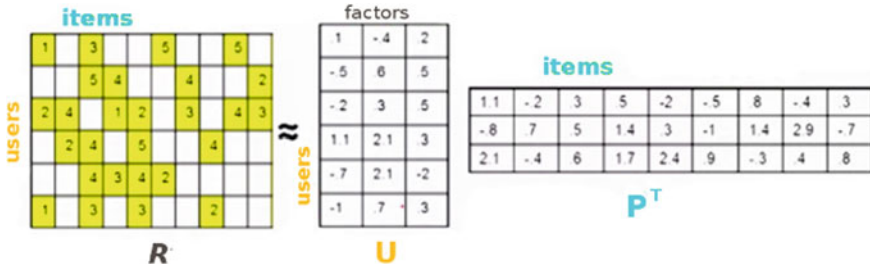


Fig. 2 Low-rank matrix factorization

$$\sum_{(u,i)} (R_{ui} - P_i^T U_u)^2 \tag{1}$$

The benefit of factorization models is the in comparison simplicity of computing recommendations once the model is formed. But, for a huge user and item sets, this would be difficult as it necessitates computation and storage over large number of millions of users and items factor vectors. In addition, one more advantage is that they are likely to provide high performance. Disadvantage to factorization models is that they are in comparison extra difficult to comprehend and infer to nearest-neighbor models and are generally extra computationally intensive throughout the model’s training phase. Although many algorithms such as eigenvector decomposition, SVD, SVD++, TSCD, have been proposed to tackle sparsity. However, these issues are yet to be well addressed.

### 4 Literature Review

The area of cross-domain recommendation is filled with a range of difficulties. So, to resolve the cold start as well as data sparsity troubles that weighed down previous CF-based recommender systems, existing methods for CDRS systems utilize a variety of artifacts that connect diverse domains. Consider example of affinity networks called Mob hinter [8] which epidemically spreads recommendations from spontaneous similarities among different users. In the paper, they paid attention on collaborative filtering techniques where impulsive affinity associations are derived and subjugated to provide customized advices through direct meetings among users. Their major objective was to suggest an approach capable to compute the least number of rating constraints among users and also in-between users and artifacts required to create precise recommendations. But authors do not considered the difficulty in distributed and sparse situations, so the difficulty of predicting recommendations in this area is still pending. In [9], authors give consideration to a new method for easing this sparsity crisis by moving user-item rating patterns from a intense auxiliary rating matrix in further domains such as a famous movie rating Web site to a

sparse rating matrix in a subjected domain such as a novel book ranking Web site. There is no necessity that the users and items in both domains must be same. They initially reduce the ratings in the supplementary rating matrix into useful knowledge. Then, they put forward an effective algorithm to rebuild the “target rating matrix” by increasing the codebook. Even though their projected CBT approach can deliberately break the other contrast approaches on both the target data, it can be seen that there is still scope of performance upgrading. However, following the temporal dynamics [10] of buyer likings to items lift more troubles. Each user and article strictly goes via a various sequences of modification in their properties. Furthermore, they generally require molding all the changes in a atomic model by interconnecting users/products to each other for classify such communal patterns based on behavior.

One more crucial challenge that has acknowledged slight thought till now is how to comprehend what kind of content users are paying attention to. The supposition generally considered is that buyers have a structured and up-to-date profile telling their choices and that data is then structured by means of a shared arrangement. Still, understanding with the work shows that often the assumptions made are not hold at all. They thus suggest an easy method that automatically learns what user’s select from the tags they use when they make and/or use content. To make the most of the chances of giving significant content to attracted users, they perform a tag-expansion method to improve content descriptions further than the folksonomy [11]. Instead of working on centrally managed knowledge about—which users add which tags to which resource in an entirely distributed way can be beneficial to much great level, than utilizing of available local knowledge. In [12], it is observed that, although their selected data is sparse for CF systems but relatively dense and related supplementary data may in the past, present in any other enhanced application domains.

In this paper, they concentrate on the data sparsity trouble in a target domain via transporting knowledge regarding both users and items from auxiliary data sources. They notice that in diverse domains the user responses are generally heterogeneous, for example, ratings versus clicks. Their algorithm first discovers a subspace where coordinate systems are used for knowledge transmission, and then utilize the transmitted knowledge to adapt to the target domain data. They have not yet extended CST in other cross-domain settings such as—for transmuting knowledge between cinemas, music, books and many more. Bring up [13] and look at different item ranking techniques which can generate recommendations that have much higher aggregate diversity among all users along with maintaining as good level of recommendation accuracy. In contrast to individual diversity, they began exploring the effect of recommendation systems on sales range by bearing in mind the aggregate variety of recommendations across all users. It is observed that it is not necessary that high individual diversity of recommendations imply more aggregate diversity. Such as if the system suggests to every user the same top-selling articles which are not similar to one another. In [14], they considered the job of succeeding personalized point-of-interest recommendation in LBSNs. They first examine the spatial-temporal traits of the LBSN datasets. They next put forward a new matrix factorization model, namely FPMC-LR [14], to incorporate together—customized Markov chain and localized regions, for resolving the recommendation job. Their experiment comes out on two

large-scale LBSN datasets, the contextual information of POIs and how to integrate social information to reinforce succeeding personalized POI recommendation was still not performed.

In most approaches, it is supposed that user conduct in every domain is similar. This supposition is not correct for all time because every user could have diverse interest, such as rating items frequently in one domain than in comparison to some other. In this paper [15] using tensor factorization technique, authors presented user-item-domain relation via tensor of order three and used them to factorize users-item-domains into latent feature vectors. The triumph of their model is extremely reliant on efficiently transporting knowledge from auxiliary that could be finely exploited; they also wish for to utilize their model to new complex CDRS scenarios mainly when the source and target domains are extra diverse. They put forward to utilize a rich feature set of their Web browsing history and search queries to characterize users. They employ a “deep learning” method for the applications where the likeness between users and their favorite items is maximized by mapping users and items set to a latent space. They expand the method to collectively learn from features of items from various domains and user features through bring in a multi-view deep learning model [16]. As deep learning has newly been proposed for constructing recommendation systems for both collaborative and content-based approaches incorporating more user features into the user view can help in sparsity reduction.

A good user-based recommendation model CRUS [17] which is based on user likeness, which first of all commences the trust relation among friends into cross-domain recommendation. Although friends generally tend to have alike interests in some domains but they may share variation in some of their interests. Taking this into consideration, they describe all like users with the target user as similar friends. By altering the transfer matrix in the random walk, friends having alike interests are highlighted. However, more solutions on modifying the transfer matrix will be implemented and evaluated. Besides, considering the interests of users can be changed according to the time is needed.

In [18], a hybrid recommendation approach utilizing user preferences with items is used. To solve sparsity problem a multi-criteria collaborative filtering algorithm is used for providing somewhat better results. It also removes the problem of data sparsity to some extent, however, multi-criteria causes each decision criterion to be evaluated every single step causing computation complexity.

Thus, to solve such problems in [19], a preference for each item available is predicted using fuzzy logic by finding conceptual similarity in the item-tree and the user liking tree. However, the problem remains for n items tree matching algorithm and the liking algorithm must run n times making it slower once again.

Also in [20], authors introduced ontology-based models for the effective recommendations generation the prediction models can be induced by correlating user liking and user items they have used a KNN CRDS, however, the problem of KNN being that it can only correlate in neighborhood of K nearby items, thus can be fairly inaccurate. Also, user preferences can change, thus making it inefficient in production environment. Sparsity though cannot be easily handled by KNN; matrix factorization is needed for same.

Inspired by [26] similarities between users, in [21], authors computed user-user similarity by evaluating users friend groups, and they do this by collecting information from neighborhood of friends of larger radii. However, the computation is again dependent on local neighborhoods which may lack in producing global recommendation thus making sparsity in larger dataset with non-correlating users. In [22], authors used rating patterns from the lower-level matrices from source domains which were then transferred to the target domain, thus the knowledge extraction method was used to enrich sparse matrix in CRDS model, their approach was effective for sparsity reduction, however, since then many other approaches for matrix factorization have been proposed which can further improve on their work.

In [24], authors developed a neural model to generate knowledge transfer method for recommending on synthetic dataset, it extracts useful knowledge from unstructured text and transfers the learned knowledge using a transfer network. Their work needs to be implemented on real-world datasets to further check the validity. Improving on the SLIM approach [27] the authors worked on producing recommendation models using regression algorithms [28], they used high-dimensional regression models for handling model sparsity. However, for cross-domains, their dense model cannot be fitted into memory that can produce sparse versions' recommendation models. They further extended their work in [29] by combining the benefits of autoencoders and neighborhood-based approaches for producing accurate item similarity matrix which can be used for neighborhood-based approaches. However, this model will work where similarity of items in the dataset closes to zero, else the hidden layers of autoencoder with limited capacity can produce inefficient results also they have used very shallow encoders.

In [30], a new model is proposed to efficiently enable transfer knowledge from unstructured text by combining hybrid methods for recommendation. TMH can work with unstructured text via end-to-end mode by extracting useful content using a memory module. Major focus was on transfer learning from source to target domain, the effectiveness of sparsity was not considered. In [31], top sequential pattern mining algorithm has been used for identifying the most frequent recommendations. They tried to extend CDR by collecting the item similarity using ontology and collaborative filtering in items and users using prefix-span pattern mining. Sequential pattern mining is efficient way to find only frequent recommendation but not for reducing sparsity as they will face candidate generation problem.

Findings from this review summarize that where there is deficiency of data in one domain, there is comparatively more data in some other domain. However, transfer learning can be utilized in case both domains are related. Furthermore, both collaborative filtering and transfer learning can be joined to take out information from a source domain with sufficient data to boost accuracy. This will help in giving better recommendations in target domains (Table 2).

**Table 2** Evaluation of various works in cross-domain recommendation systems which dealt with CDRS most significant challenges

Algorithm	Year	Type	Context	Dataset	Sparsity	User drift	Metrics	Application
Mob hinter [8]	2008	User-based	Ratings	Movie lens	No	No	RMSE, MAE	Social network
CBT [9]	2009	User-based	Ratings	Movie lens, each movie, book-crossing	Yes	No	MAE	General
TSVD++ [10]	2009	User-based	Time	Netflix	No	Yes	RMSE	General
Folksonomy [11]	2010	Tags-based	Tags	Movie lens	No	No	Recall	Social network
CST [12]	2010	Transfer learning	Ratings	Movie lens, Netflix	Yes	No	RMSE, MAE	General
Item popularity-based ranking [13]	2014	Item-based	Items	Yahoo! movies, Netflix	Yes	No	Precision loss	General
FPMC-LR [14]	2013	User-based	Time	Gowalla	Yes	Yes	Convergence	Social network
FM-MCMC [15]	2014	User-based	Ratings	Amazon dataset	Yes	No	RMSE, MAE	General
MV-DNN [16]	2016	User-based	Ratings	Web, movie/TV view logs	Yes	No	Convergence, MAE	General
CRUS [17]	2016	User-based	Users	Yelp	Yes	No	Accuracy	General
Multi-criteria [18]	2016	Item-based	Users	Yahoo! movies	Yes	No	Convergence	General
Tree based fuzzy [19]	2016	Item-based	Users	Movie lens	Yes	No	Precision, recall	General
ontology-based KNN [20]	2017	Item-based	Users	Movie lens	Yes	No	MAE, accuracy	General
Friend Group [21]	2017	Item-based	Users	-	Yes	No	MSE	General
CCCFNet [23]	2017	Item-based	Ratings	Twitter	No	No	Precision, recall	Social network
Rating patterns approximation [22]	2018	Transfer learning	Ratings	Netflix, each movie	Yes	No	MAE, RMSE	General
Neural transfer learning [24]	2018	Transfer learning	Ratings	Mobile news, Amazon products	Yes	No	MAE	General

(continued)



Table 2 (continued)

Algorithm	Year	Type	Context	Dataset	Sparsity	User drift	Metrics	Application
Superhighway [25]	2018	Transfer learning	Ratings	KKBOX_RI	Yes	No	MAP	General
CIT [26]	2018	Transfer learning	Ratings	Movie lens, Netflix	Yes	No	MAE, RMSE	General
SLIM [27]	2018	Item-based	Users	Movie lens	Yes	No	MAE, RMSE	General
High-dimensional regression [28]	2019	Item-based	Users	Movie lens	Yes	No	Recall	General
Autoencoders and neighborhood [29]	2019	Item-based	Users	Movie lens	Yes	No	MAE	General
TMH [30]	2019	Hybrid	Item	Mobile news, Amazon product	Yes	No	RMSE	General
CD-SPM [31]	2019	Item-based	Ratings	Movie lens	No	No	RMSE, F1-score	General

## 5 Conclusion and Future Perspectives

Irrespective of the work and advantages one gets, there is an innovation in this research topic and most of the part of this topic is still unexplored. Exploiting information from lot of other domains to offer suggestion from the distinct perspectives has been addressed in many areas which include the retrieval of information, management of knowledge, user modeling and learning via machine. The problems of sparseness are faced by the cross-domain collaborative filtering and this is made by shifting of rating information in various existing domains. As future work, we think to investigate CDR for challenges including knowledge transfer and sparsity. Also, we can focus a much better fusion of information from auxiliary sources such as user emotions to target domains; which could result in better cross-domain recommendations. One of the machine learning models is transfer learning (TL) that utilizes information learned in one chore in a different but linked source domain, to resolve the chores in target domain. It is specifically used when there is less high-quality training data. So, to resolve the sparsity problem, we will focus on a novel transfer learning approach based on human emotions.

## References

1. Viktoratos I, Tsadiras A, Bassiliades N (2018) Combining community-based knowledge with association rule mining to alleviate the cold start problem in context-aware recommender systems. *Expert Syst Appl* 101:78–90
2. Melville P, Sindhvani V (2011) Recommender systems. In: *Encyclopedia of machine learning*. Springer, Boston, MA, pp. 829–838
3. Al Mamunur Rashid SKL, Karypis G, Riedl J (2006) ClustKNN: a highly scalable hybrid model-& memory-based CF algorithm. In: *Proceeding of WebKDD*
4. Shani G, Gunawardana A (2011) Evaluating recommendation systems. In: *Recommender systems handbook*. Springer, Boston, MA, pp 257–297
5. Sarwar B, Karypis G, Konstan J, Riedl J (2002) Incremental singular value decomposition algorithms for highly scalable recommender systems. In: *Fifth international conference on computer and information science*. Citeseer, Dec 2002, pp 27–28
6. Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. *Computer* 8:30–37
7. All in One Tutorial via Matrix Factorization (2019). [https://antk.readthedocs.io/en/latest/mf\\_tutorial.html](https://antk.readthedocs.io/en/latest/mf_tutorial.html)
8. Schifanella R, Panisson A, Gena C, Ruffo G (2008) Mobhinter.: epidemic collaborative filtering and self-organization in mobile ad-hoc networks. In: *Proceedings of the 2008 ACM conference on recommender systems*, Oct 2008. ACM, pp. 27–34
9. Li B, Yang Q, Xue X (2009) Can movies and books collaborate? Cross-domain collaborative filtering for sparsity reduction. *IJCAI* 9(July):2052–2057
10. Koren Y (2009) Collaborative filtering with temporal dynamics. In: *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, June 2009, pp 447–456
11. Lo Giusto G, Mashhadi AJ, Capra L (2010) Folksonomy-based reasoning for content dissemination in mobile settings. In: *Proceedings of the 5th ACM workshop on challenged networks*. ACM, Sept 2010, pp 39–46

12. Del Prete L, Capra L (2010) differs.: a mobile recommender service. In: 2010 eleventh international conference on mobile data management (MDM), pp 21–26. IEEE, May 2010
13. Kumar R, Verma BK, Rastogi SS (2014) Social popularity based SVD++ recommender system. *Int J Comput Appl* 87(14)
14. Cheng C, Yang H, Lyu MR, King I (2013) Where you like to go next: successive point-of-interest recommendation. In: Twenty third international joint conference on artificial intelligence, June 2013
15. Loni B, Shi Y, Larson M, Hanjalic A (2014) Cross-domain collaborative filtering with factorization machines. In: European conference on information retrieval. Springer, Cham, April 2014, pp 656–661
16. Song Y, Elkahky AM, He X (2016) Multi-rate deep learning for temporal recommendation. In: Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval. ACM, July 2016, pp 909–912
17. Xu Z, Jiang H, Kong X, Kang J, Wang W, Xia F (2016) Cross-domain item recommendation based on user similarity. *Comput Sci Inf Syst* 13(2):359–373
18. Shambour Q, Hourani M, Fraihat S (2016) An item-based multi-criteria collaborative filtering algorithm for personalized recommender systems. *Int J Adv Comput Sci Appl* 7(8):274–279
19. Sen M, Udgirkar S (2016) Recommendation system for tree structure data based on fuzzy preferences and clustering. *Int J Eng Res* 5(6):534–538
20. Subramaniaswamy V, Logesh R (2017) Adaptive KNN based recommender system through mining of user preferences. *Wirel Pers Commun* 97(2):2229–2247
21. Borgs C, Chayes J, Lee CE, Shah D (2017) Thy friend is my friend: Iterative collaborative filtering for sparse matrix estimation. In: Advances in neural information processing systems, pp 4715–4726
22. He M, Zhang J, Yang P, Yao K (2018) Robust transfer learning for cross-domain collaborative filtering using multiple rating patterns approximation. In: Proceedings of the eleventh ACM international conference on web search and data mining, ACM, Feb 2018, pp 225–233
23. Lian J, Zhang F, Xie X, Sun G (2017) CCCFNet: a content-boosted collaborative filtering neural network for cross domain recommender systems. In: Proceedings of the 26th international conference on world wide web companion. International world wide web conferences steering committee, pp 817–818
24. Hu G, Zhang Y, Yang Q (2018) MTNet: a neural approach for cross-domain recommendation with unstructured text
25. Lai KH, Wang TH, Chi HY, Chen Y, Tsai MF, Wang CJ (2018) Superhighway: bypass data sparsity in cross-domain CF. arXiv preprint [arXiv:1808.09784](https://arxiv.org/abs/1808.09784)
26. Taneja A, Arora A (2018) Cross domain recommendation using multidimensional tensor factorization. *Expert Syst Appl* 92:304–316
27. Wu G, Volkovs M, Soon CL, Sanner S, Rai H (2018) Noise contrastive estimation for scalable linear models for one-class collaborative filtering. arXiv preprint [arXiv:1811.00697](https://arxiv.org/abs/1811.00697)
28. Steck H (2019) Collaborative filtering via high-dimensional regression. arXiv preprint [arXiv:1904.13033](https://arxiv.org/abs/1904.13033)
29. Steck H (2019) Embarrassingly shallow autoencoders for sparse data. In: The world wide web conference. ACM, May 2019, pp 3251–3257
30. Hu G, Zhang Y, Yang Q (2019) Transfer meets hybrid: a synthetic approach for cross-domain collaborative filtering with text. In: The world wide web conference. ACM, May 2019, pp 2822–2829
31. Anwar T, Uma V (2019) CD-SPM: cross-domain book recommendation using sequential pattern mining and rule mining. *J King Saud Univ-Comput Inf Sci*

# Enabling Edge Computing in an IoT-Based Weather Monitoring Application



Kavita Srivastava and Sudhir Kumar Sharma

**Abstract** Internet of Things (IoT) applications employ several sensors for gathering data. These sensors are often placed at hard to reach locations since they need to be left unmonitored. Sensors collect data in real time and send it to a cloud server for further processing. Often the sensors generate a large volume of data which is redundant in nature. Transferring sensor data to a cloud server for processing leads to high bandwidth consumption, delay, and increase in operational cost and data security issues. Edge computing allows the sensor data to be stored and analyzed on an edge device, and only the data summary is sent to the cloud server. In this paper, an edge computing approach for managing and analyzing data in a weather monitoring application is proposed. The application has been built using a Raspberry Pi system. The computation has been performed by creating Microsoft Azure IoT Hub resource and Microsoft Azure IoT Edge solution. Results have been compared with AzureML cloud platform.

**Keywords** Edge computing · IoT Hub · IoT edge · Internet of Things · Device Provisioning System · Edge runtime · Edge modules

## 1 Introduction

Edge computing is an enabling technology that leverages Internet of Things (IoT) applications by providing a wide range of services. An IoT application solution is built by using one or more sensor nodes comprising microcontroller units (MCU) and some provisioning of data storage, analysis, and prediction. Often these facilities are provided by a cloud computing service. This makes the centralization of resources in IoT application. This scheme has certain drawbacks. It is not suitable for IoT applications which require quick response. For instance, suppose a weather

---

K. Srivastava (✉) · S. K. Sharma  
Institute of Information Technology and Management, Janakpuri, Delhi, India  
e-mail: [kavita332211@gmail.com](mailto:kavita332211@gmail.com)

S. K. Sharma  
e-mail: [Sharmasudhir08@gmail.com](mailto:Sharmasudhir08@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_18](https://doi.org/10.1007/978-981-15-2329-8_18)

monitoring application is designed to collect temperature and humidity, data detects a sudden rise in temperature and causes an alarm to ring. If the data has been sent to cloud service for analysis, it would cause delays making the system ineffective.

Edge computing performs the analysis of data near to device location. It means the analysis has been done at the place where data is collected. Edge computing benefits the IoT applications in several ways.

- It makes application distributed in nature. We can aggregate data of several nodes which are placed in the neighborhood of each other on a single edge device. Hence, for one IoT application consisting of thousands of nodes, we can use several edge devices.
- Employing edge devices results in less data to be transferred to the cloud, thereby reducing the requirement of data storage on cloud and less utilization of network bandwidth.
- The edge computing makes application more secure since data is stored only on the edge device and can be processed offline.
- The application becomes more scalable since we can easily employ more sensor nodes targeted to an existing edge device or to a new edge device.

In this study, we have proposed an edge computing-enabled weather monitoring application built from Raspberry Pi board and DHT11 temperature and humidity sensor. We have used Microsoft Azure IoT Edge for stream analysis and Microsoft AzureML for comparative analysis.

Rest of the paper is organized as follows. Section 2 provides the related research in edge computing. Section 3 discusses the architecture of edge computing-enabled IoT application. Section 4 provides workflow of implementation, Sect. 5 provides experiments performed, and Sect. 6 provides results and discussion.

## 2 Related Work

In the context of IoT, edge computing has not been explored much by the research community. Dupont et al. [1] presented Cloud4IoT platform for IoT functions migration horizontally and vertically. They explained two use cases for health care and remote engine diagnostic for roaming and offloading, respectively. Higashino et al. [2] have pointed out a number of challenges in edge computing like creation technology, networking technology, processing technology, and content curation technology. They discussed their ongoing research on disaster mitigation using edge computing concepts. Pan et al. [3] discussed IoT applications benefited from edge cloud. They discussed several benefits and challenges of using edge cloud infrastructure. Babou et al. [4] proposed a home edge computing (HEC) architecture for providing data storage and processing capability near user locations. They demonstrated their work on EdgeCloudSim simulator. Ramljak et al. [5] presented Bayesian reasoning-based framework for minimizing data transfer between edge and cloud. They provided BelifeCache modular framework to support their work. Zeng et al. [6] proposed a

system for edge-based computing on visual sensors for face recognition application. Hsieh et al. [7] presented a Docker container-based managed platform for smart city applications. They described the deployment of three different applications of air quality monitoring, image recognition, and sound classification. Hu et al. [8] proposed a retail point-of-sale system based on blockchain technology. Here, a node in blockchain network works as edge computing server. Sanchez et al. [9] proposed a system of performing computation through convolution neural networks (CNN) on edge devices which are placed near the camera sensors. Yu et al. [10] described edge computing based on containers which provide lightweight virtualization. They have used Docker container and deployed the container through Kubernetes orchestration tool. Merlino et al. [11] suggested an OpenStack-based middleware platform. They described how containers at edge, fog, and cloud level can be discovered and combined.

### 3 Architecture

An IoT solution is basically comprised of three parts—things, insight, and action. Things include devices, sensors, industrial PCs, gateways, and everything collecting the data. Insights include real-time analysis of data, historical insight, and making predictions. IoT projects are still complex to start with. They are incompatible with existing infrastructure and difficult to scale. Azure IoT provides secure, fast, open, and scalable solutions. With azure IoT, we can enhance and scale our IoT applications effortlessly. Azure IoT provides a way to establish bidirectional communication between cloud to device and device to cloud messages. Figure 1 shows the layered architecture of edge computing-enabled IoT application.

The architecture of application has three layers—device layer, edge layer, and cloud layer. These layers are explained below in detail.

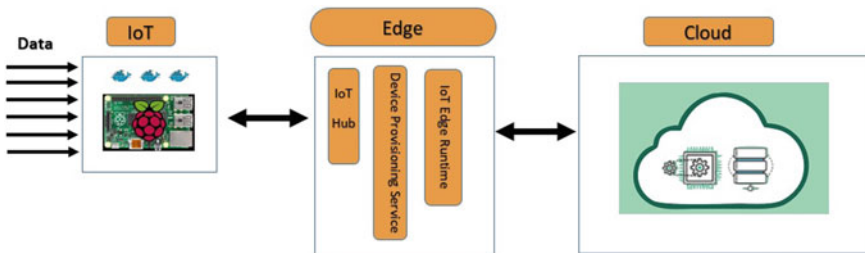


Fig. 1 Architecture of edge computing-enabled IoT application

### ***3.1 Device Layer***

Device layer is composed of sensors and microcontrollers. Sensors directly collect data from the environment. The sensors are employed for a wide range of use cases. We need to install Docker container on device node.

### ***3.2 Edge Layer***

Edge layer has three components—Iot Hub, Device Provisioning Service, and IoT Edge Runtime. IoT Hub provides services to devices and applications. It is responsible for establishing communication among sensor devices and backend application. We can monitor and control devices using IoT Hub, determine device status, and detect device failure from cloud. Device Provisioning Service is a way to register devices on cloud with zero touch. Device Provisioning Service can provision up to millions of devices. Azure Device Provisioning Service eliminates the need of human intervention completely. It also provides security and scalability. Device Provisioning Service provides automatic configuration and load balancing. Edge runtime is responsible for providing basic services such as management, connectivity, and security for devices which may be operating offline. Capabilities of runtime are provided in the form of modules which perform specific actions such as stream analytics or storage solution. Edge runtime modules can be chained together for pipeline processing of data. Modules are implemented as Docker containers. Custom modules can be written in the language of your choice such as C#, Java, or Python. All devices connected through IoT Hub run IoT Edge Runtime. Modules in edge runtime are deployed according to application use case.

### ***3.3 Cloud Layer***

Edge layer can provision millions of devices and filter the streaming data before it reaches the cloud layer. However, cloud layer is required for further processing of data. Cloud layer provides big data storage, business analytics, and machine learning.

In the next section, implementation of the application has been explained.

## **4 Implementation**

An IoT solution requires IoT sensor nodes and a backend application running on cloud. Microsoft Azure IoT solution provides the facility for controlling and monitoring IoT devices. The communication between device and backend needs to be

bidirectional. For instance, in case of weather monitoring application, sensor nodes collect data and send to the cloud. If any abnormal behavior is detected like sudden rise in temperature along with sudden fall in humidity, it may be an indication of fire outbreak. The backend cloud application must send a notification to the device causing an alarm to ring as well as a notification on user's mobile phone should also be sent.

Microsoft Azure provides a number of IoT-related services including IoT Central (a SaaS solution), IoT solution accelerator (PaaS solution), IoT Hub, Device Provisioning Service, IoT Edge, Azure Digital Twins, Time Series Insights, and Azure Maps for providing geographic information to applications. IoT Edge is suitable when we need to analyze data on the IoT devices rather than using cloud service to analyze data. By doing this, we can perform the computation on edge, and only few messages need to be sent to the cloud for storage. In our weather monitoring application, we have utilized the service of Azure IoT Edge. The hardware part of our application consists of Raspberry Pi board and DHT 11 temperature and humidity sensor. We have connected the sensor with Raspberry Pi board and installed the Raspbian operating system. Then, we have installed Docker container on Raspberry Pi followed by IoT Edge Runtime. After that we can login into Azure account and create IoT Hub resource, create a resource group, select a region nearest to the device location, and specify size and scale. We can obtain connection string to be used in the application code. Then, we registered the device by specifying device ID. Then, we have linked Device Provisioning Service to IoT Hub. For that, we have created a C# console application and installed `Microsoft.Azure.Devices.Provisioning.Service` library. Since we want to analyze data on the device edge rather than on cloud, we have configured IoT Edge. The workflow of application configuration and deployment is shown in Fig. 2. The application starts by fetching data from device. It computes the average of both temperature and humidity values. If the difference between current value and average is more than 10, a message is generated and sent to the cloud. Next, the application plots the data fetched. Figure 3 shows the computation flowchart.

## 5 Experiments

The device data has been collected for a period of one month in winter season of January 12, 2019, to February 10, 2019. It has been observed that variation in temperature and humidity values is rare for a particular day. This results in very few messages being sent to cloud for further processing. Plot of Day 1 shows the effect of turning on a heating device which resulted in some increase in temperature and decrease in humidity value. The same pattern followed for each day.

The results have shown that in spite of collecting large amount of data by sensor node, only few messages have been sent to the cloud backend. The data has been analyzed at IoT Edge before sending to the cloud. In order to perform a comparative analysis of computing on edge versus computing on cloud, we have performed



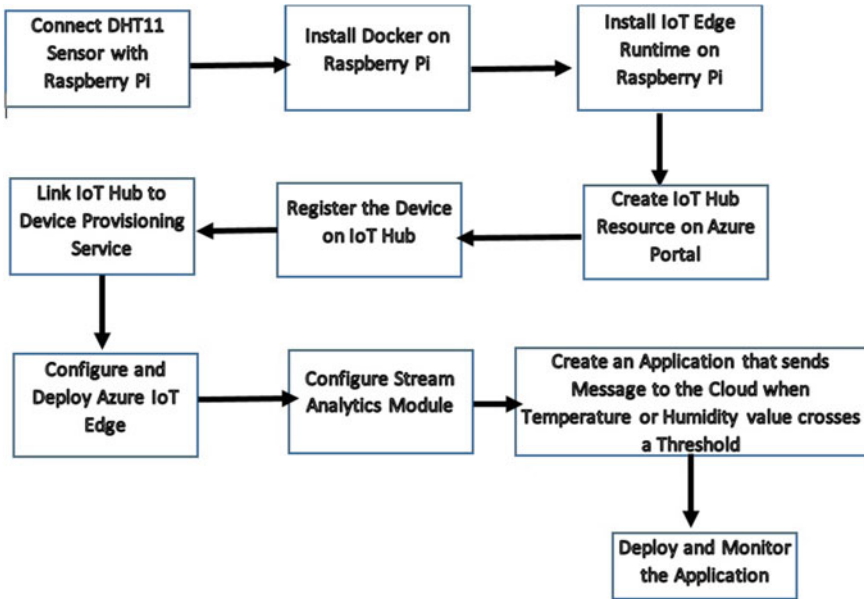


Fig. 2 Application workflow

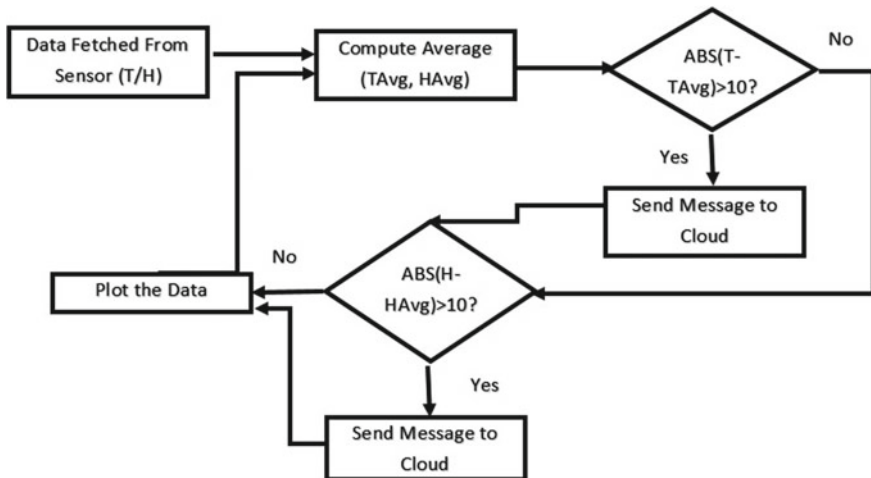
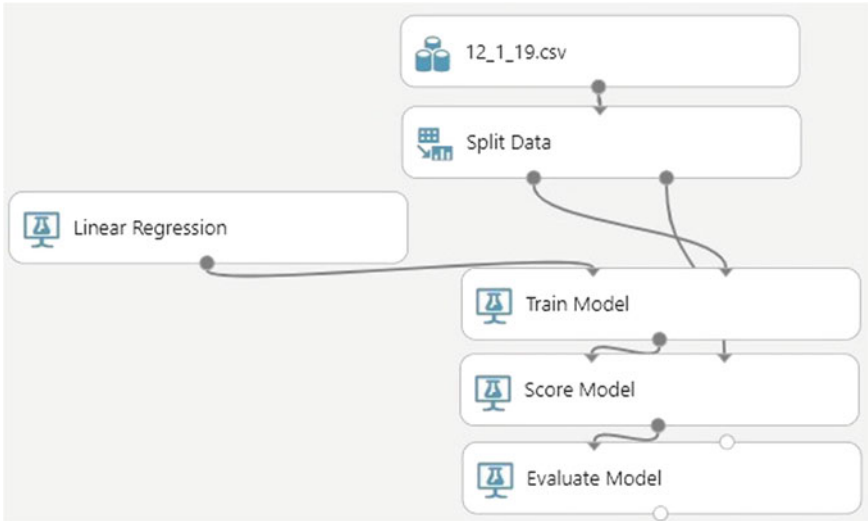


Fig. 3 Computation flowchart

regression analysis on cloud by utilizing AzureML platform or machine learning. The cloud experiment model is shown in Fig. 4.

The data has been uploaded using datasets module of AzureML Studio. Next, in experiment module, this dataset is divided into training set containing 80% of total



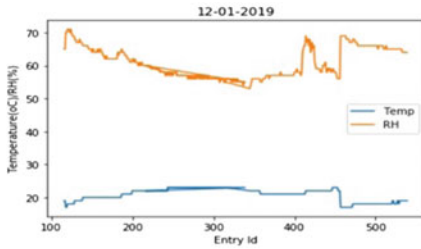
**Fig. 4** Regression analysis with AzureML

data and test set containing remaining 20% of data using split data function. A linear regression model is added to the experiment with online gradient descent method with learning rate = 0.1 and number of training epochs = 10. A train model function is added, and temperature column is selected from dataset. A score model function is added to the experiment to find predicted values in test dataset. Finally, an evaluation model function is added to measure the accuracy of trained dataset. The experiment is executed, and the results are shown in Fig. 5.

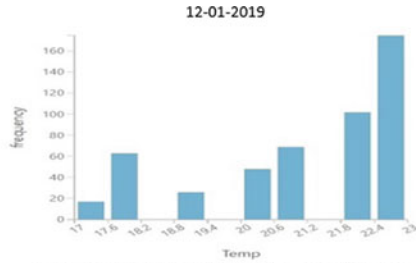
## 6 Results and Discussion

Figure 5a shows the plot of telemetry data obtained on edge device. Figure 5b shows the data of the same day visualized on the cloud server. Figure 5c indicates that scored labels have shown very small variation in data with standard deviation of 0.1675. Figure 5d shows the evaluation metrics.

The results indicated that analysis on edge and cloud is mostly similar; however, edge computing provides several benefits over cloud computing. Edge computing results in substantial reduction in bandwidth utilization and latency. It makes our application fast and cost effective.



(a) Plot of Temperature and Humidity Data Fetched by Edge Device.



(b) Temperature data visualization on cloud server.

Mean	15.0058
Median	15.0147
Min	14.6496
Max	15.2909
Standard Deviation	0.1675
Unique Values	96
Missing Values	0
Feature Type	Numeric Score

(c) Statistics obtained from Scored Labels

Mean Absolute Error	6.114204
Root Mean Squared Error	6.406236
Relative Absolute Error	3.755654
Relative Squared Error	11.509944
Coefficient of Determination	-10.509944

(d) Metrics obtained from Evaluation Model

Fig. 5 Comparative analysis on edge and cloud

## 7 Conclusion and Future Work

The edge computing has emerged as a technology that can be utilized in between the sensors node collecting data from environment and the cloud computing services. Edge computing has tremendous potential of making our IoT applications work efficiently by causing the resource utilization efficiently. Edge computing results in making IoT application distributed in nature by managing the data near the location where it is generated.

In this study, we have used Microsoft Azure IoT Edge for edge computing. We have employed temperature and humidity sensor with Raspberry Pi board and collected the data. The data has been processed by the edge device which can be a laptop or PC. The results have shown that very few messages have been transferred to the cloud service indicating abnormal conditions. The edge device has successfully analyzed the data. The resulting application is fast and cost effective.

Our future work will involve integrating more sensor nodes for collecting data and employing other edge modules.

## References

1. Dupont C, Giaffreda R, Capra L (2017). Edge computing in IoT context: horizontal and vertical Linux container migration. In: 2017 Global internet of things summit (GloTS), June 2017. IEEE, pp 1–4
2. Higashino T, Yamaguchi H, Hiromori A, Uchiyama A, Yasumoto K (2017) Edge computing and IoT based research for building safe smart cities resistant to disasters. In: 2017 IEEE 37th international conference on distributed computing systems (ICDCS), June 2017. IEEE, pp 1729–1737
3. Pan J, McElhannon J (2017) Future edge cloud and edge computing for internet of things applications. *IEEE Internet Things J* 5(1):439–449
4. Babou CSM, Fall D, Kashiara S, Niang I, Kadobayashi Y (2018) Home edge computing (HEC): design of a new edge computing technology for achieving ultra-low latency. In: International conference on edge computing, June 2018. Springer, Cham, pp 3–17
5. Ramljak D, Tom DA, Voigt D, Kant K (2018) Modular framework for data prefetching and replacement at the edge. In: International conference on edge computing, June 2018. Springer, Cham, pp 18–33
6. Zeng J, Li C, Zhang LJ (2018) A face recognition system based on cloud computing and AI edge for IOT. In: International conference on edge computing. Springer, Cham, June 2018, pp 91–98
7. Hsieh YC, Hong HJ, Tsai PH, Wang YR, Zhu Q, Uddin MYS, Venkatasubramanian N, Hsu CH (2018) Managed edge computing on Internet-of-Things devices for smart city applications. In: NOMS 2018–2018 IEEE/IFIP network operations and management symposium, April 2018. IEEE, pp 1–2
8. Hu B, Xie H, Ma Y, Wang J, Zhang LJ (2018) A robust retail POS system based on blockchain and edge computing. In: International conference on edge computing, June 2018. Springer, Cham, pp 99–110
9. Sanchez J, Soltani N, Kulkarni P, Chamarithi RV, Tabkhi H (2018) A reconfigurable streaming processor for real-time low-power execution of convolutional neural networks at the edge. In: International conference on edge computing, June 2018. Springer, Cham, pp 49–64
10. Yu Z, Wang J, Qi Q, Liao J, Xu J (2018) Boundless application and resource based on container technology. In: International conference on edge computing, June. Springer, Cham, pp 34–48
11. Merlino G, Dautov R, Distefano S, Bruneo D (2019) Enabling workload engineering in edge, fog, and cloud computing through OpenStack-based middleware. *ACM Trans Internet Technol (TOIT)* 19(2):28

# Green Cloud Job Scheduling and Load Balancing Using Hybrid Biogeography Based Optimization and Genetic Algorithm: A Proposed Approach



Yashika Sharma and Sachin Lakra

**Abstract** In the present era, technology plays a very significant role in human life. It makes human life easier and has become a crucial part of it. Like air, water and food, technology has become a necessity for human survival. Specifically, technologies related to the areas of computer science and the internet play a very wide role in day-to-day life. Everyday new internet technologies and computing devices are being launched into the market and they define the working style of human beings. With this ever-increasing demand and usage of technology; the amount of carbon footprint is also increasing as a byproduct of technology. The increased amount of carbon footprint accounts for considerable global warming. Energy requirement for the huge number of computing resources is also huge and so is the magnitude of heat produced resulting in global warming. Thus, there is an urgent need of a transition towards green computing. IT Companies/industries should integrate green agenda in their products. Green IT is not a technology itself but it is a transition from conventional computing to green computing. This transition may help to curb negative effects on the environment.

**Keywords** Cloud computing · Energy efficiency · Load balancing

## 1 Introduction

Green Computing is about optimum utilization of the available resources minimizing power usage. There are various approaches to achieve the optimal utilization: Hardware approach and Software approach. The hardware approach comprises of change in equipment design, equipment recycling, application architecture and power management, etc. On the other hand, software approaches include virtualization, cloud computing, grid computing, algorithm optimization, etc.

---

Y. Sharma (✉) · S. Lakra  
Manav Rachna University, Faridabad, India  
e-mail: [Yashika.sharma85@gmail.com](mailto:Yashika.sharma85@gmail.com)

S. Lakra  
e-mail: [sachin@mru.edu.in](mailto:sachin@mru.edu.in)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_19](https://doi.org/10.1007/978-981-15-2329-8_19)

Nowadays a new era is taking hold in the IT marketplace, which is introducing computing services to the user on demand in a pay-per-use manner anytime everywhere. However, consumers need to be linked to a fast-speed internet connection to be able to harness these services. This era is known as “Cloud Computing”. Cloud computing is making the provision of computing facilities as a provider as opposed to a product, in which shared assets, software programs and statistics are provided to users over the network. Cloud computing vendors supply software through the internet, which may be given the right of usage through web browsers. At the same time, business software and information are saved on servers at a distant location. Cloud computing can offer three types of carrier modes, inclusive of “IaaS”, “PaaS” and “SaaS”. The above mentioned facilities/services of Cloud Computing eliminate the need of individual drives and software thus minimizing the usage of resources such as water used for cooling data centers. Cloud computing is thus making a pavement for the transition towards Green IT from conventional IT; an eco-friendlier and energy efficient way of computing. Thus, reducing “SaaS” is a provider which furnishes programs available on the cloud-computing infrastructure to users, hosted through carrier companies. “PaaS” refers to services, which offer high-degree integrated surroundings to layout, construct, run, take a look at, install and replace the packages created by the use of development languages and tools, including, Java, python, internet, etc., supplied by way of provider companies to the cloud. “IaaS” refers to the offerings supplied to customers to hire electricity, garage, community and different primary computing resources, using which customers can install and run any software program including working structures and programs. The beauty of cloud computing is that anytime anywhere anyone can access the available services by paying subscription charges to the service provider. Clients need not bother about the service maintenance, updated versions, etc. (Fig. 1).

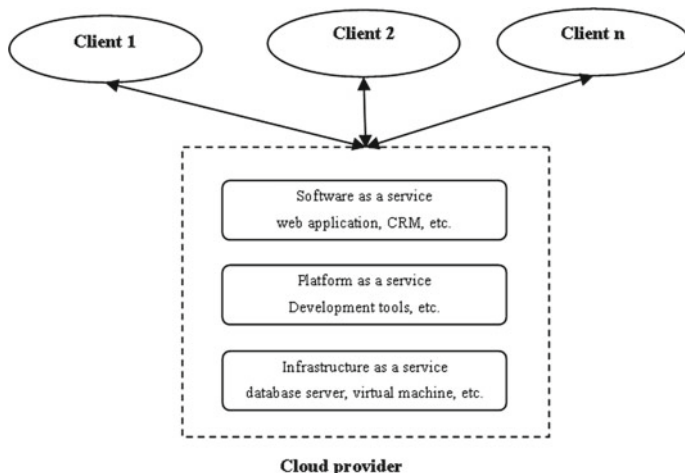


Fig. 1 Architecture of cloud computing

## 2 Related Work

Dastjerdi et al. [1], describe that since a cloud service individually cannot fulfill all of a user's needs, a framework of cloud offerings is required. A cloud carrier framework, which would include multiple duties such as innovation, realizing compatibility, choice, and deployment, is a complicated technique and users would find it difficult to select the optimal technique out of hundreds and if not hundreds, of all feasible compositions possible. The service framework in the cloud would raise new challenges because of diversity of customers with distinctive knowledge needing their programs to be rolled-out throughout different geographical places with a number of constraints.

Hung et al. [2], describe that to obtain good overall performance, numerous servers in cloud data centers (DCs) work in synch to ensure high availability and reliability of cloud computing offerings, specifically, in phases of multitasking. Creative and efficient methods are needed for handling the collapse of such computing nodes. A good amount of research work has been done to deal with such hassles, however, it cannot always assure necessary minimal performance. The authors, in this paper, have developed a scheduling set of regulations, primarily based on fee and bandwidth, which ensures the possibility of green recuperation on heterogeneous computing environments.

Menzel et al. [3], describe that with increasing vendors supplying a huge range of computing facilities, the movement of existing system towards the new era would require finding a right combination of computing facilities and virtual machines (digital system) descriptions. A better way, therefore, for internet applications is to automate the assessment ensuring the service excellence without compromising the contradictory choice standards such as throughput and price.

Prasad et al. [4], give a cloud aid obtaining technique that presently may not be most effective, but automates the process of selecting the most appropriate cloud company. Further, it implements dynamic pricing. Three viable mechanisms are advised for procurement of cloud beneficial useful resource: Cloud-Dominant Strategy Incentive Compatible (C-DSIC), Cloud-Bayesian Incentive Compatible (C-BIC), and Cloud Optimal (C-OPT). C-DSIC is a technique primarily founded on the VCG mechanism. The C-BIC mechanism achieves price range balance but does not fulfill individual rationality. In C-DSIC and C-BIC, the cloud vendor who gives the lowest price per unit QoS is asserted the winner. In C-OPT, the cloud provider with the least virtual charges is stated the winner.

Rahman et al. [5], describe the explosive boom of cloud computing in modern years which has led to a big growth in both the amount of website visitors and the sort of issuer requests to cloud servers. This boom style of load poses crucial stressful situations to the cloud load balancer in inexperienced balancing of the burden, which is already a daunting process. The cloud load balancing is an area which is well-researched in which several solutions to balance load had been proposed.

Sarbazi-Azad et al. [6], describe this bankruptcy introducing the fundamental principles of marketplace-oriented cloud computing structures and affords a reference

version. The model, together with the present-day technologies offered within the chapter, make a contribution toward the mainstream adoption of Cloud computing era.

### 3 Problem Definition/Need of Study

Scheduling in the context of cloud computing structures is a great challenge. A disbursed answer is desired usually for this, since it is not always commonly realistic, viable and cost-effective to preserve one or extra idle services simply to satisfy the desired requirements. It is not feasible to assign jobs to appropriate servers and customers in my opinion for efficient scheduling as cloud has a complex structure and components are available at some stage in a massive distributed environment. Scheduling algorithms are categorized as static and dynamic. Static algorithms are generally suitable for homogeneous environments and might produce great consequences in given environments. However, they may not be commonly suitable and cannot meet the dynamic modifications to the attributes at some level within the execution time. Dynamic algorithms are more suitable and think about one-of-a-type of attributes previous to and during run-time. Scheduling is the system of enhancing the overall performance of a system through a redistribution of load amongst processors.

#### 3.1 *Hybrid Cloud*

A hybrid cloud is a framework in which a commercial enterprise employer provides and manages various types of resources in-house and has other types provided externally. A corporation would possibly use a public cloud's offerings and provide for archived statistics, however, continue to preserve in-house storage for operational client records. This deployment version helps businesses to make use of secured packages through website hosting on the local cloud, at the same time taking part in value addition through maintaining shared statistics and programs on a general cloud available publicly. This is the model that is likewise used to handle the situation of cloud bursting, in which the cloud infrastructure defined private seems unable to cope with overload and calls the public cloud to handle the load (Fig. 2).

#### 3.2 *Features of Cloud Computing*

The key features of cloud computing can be listed as below:

**On-Demand and Self Provisioning.** This is wherein the cloud provides computing abilities based totally on the users' needs. Users' needs may change often,



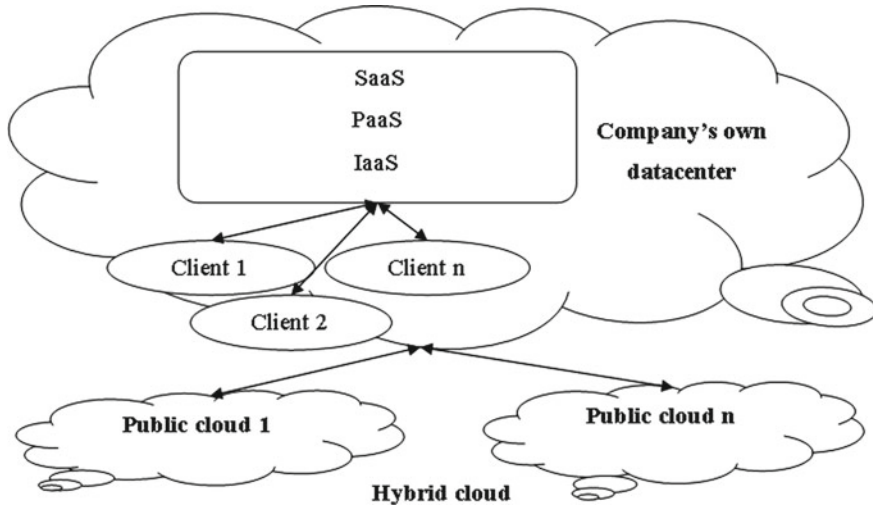


Fig. 2 Structure of the hybrid cloud

due to which it is referred to as “on-demand”. The pro of this is that, most of these provisioning strategies do not need human intervention.

**Shared Infrastructure.** The use of a virtualized software model, allows the sharing of networking abilities, services and storage. The infrastructure of cloud, no matter what the deployment model is, ensures the provision of maximum of the available infrastructure to its customers.

**Dynamic Provisioning.** Run time provisioning permits for the availability of offerings based totally on contemporary demand necessities. This is completed routinely by the use of software automation, permitting the growth and compaction of carrier functionality, as required. The resulting dynamic scaling needs to be carried out at the same time while keeping excessive ranges of reliability and safety in mind.

**Network Access.** The capacities of systems are to be provided to clients through a network and can be accessed from one-of-a-kind devices which include desktop computers, cell phones, smart phones and pill devices. Deployments of offerings within the cloud consist of the use of commercial enterprise programs on the one hand, and the use of modern-day software on the most recent smart phones, on the other hand.

**Resource Pooling.** The issuer pools its computing resources to serve a few consumers. The usage of a multi-tenant version, with one of a kind physical and digital set of resources, is dynamically assigned and reassigned in line with user requirements.

**Elasticity and Flexibility.** Cloud infrastructure may be made available to the system user in a short time period. Suppose the traffic on a website is increasing on regular basis and net website online traffic increases up to one hundred percent. If the website is hosted on a personal server, there may be a strong possibility for it to in reality “go down” and prevent completion of work due to software and hardware

barriers. In such instances, the Cloud dynamically allocates necessary resources which will ensure an easy operation, and when the load decreases once more, sources are automatically restored to their original state.

**Managed Metering.** Managed metering is aimed to focus on managing as well as optimizing the provider. It also provides statistics of reporting and billing. In this way, clients are charged for the services they have used and are billed accordingly. More precisely, cloud computing permits for the scalable deployment and sharing of offerings, as desired, from almost every possible location and for this the purchaser may be charged entirely for real utilization.

### *3.3 Challenges of Cloud Computing*

The following are a number of disturbing situations associated with cloud computing, and although a number of those may additionally cause a slowdown whilst turning in extra services within the cloud. A majority of the situations can also provide opportunities, if resolved with due care and interest inside the Cloud.

**Data Location.** It is because of cloud computing technology that cloud servers are allowed to live everywhere. For this reason, companies would not recognize the coverage area of the particular server used to keep and approach the available applications and data. It is far more essential to remember the fact that most of the cloud provider carriers (CSPs) can also especially outline where records are to be placed.

**Cloud Security Policy.** Some CSPs may additionally have a good deal less transparency as compared to others. The clarification for this mismatch is that the policies can be proprietary. Due to which, problems with the organization's data compliance needs may arise. The corporation wants to have distinctive records of the service level agreements (SLAs) that stipulated the degree of safety supplied by using the CSPs.

**Lack of Standards.** The cloud has documented interfaces; however, there are no standards related to those in existence, and as a result it may happen that maximum clouds may not communicate with each other.

**Continuously Evolving.** Customer necessities are constantly developing along with the requirements for interfaces, networking, and storage. Because of this a "cloud," mainly a public cloud, does no longer continue to be static and is likewise constantly evolving.

## **4 Cloud Scheduling**

Cloud Scheduling seeks "the best" technology time table for the datacenters to supply the required call for transmission losses at minimum production value. Numerous investigations on cloud load balancing have been undertaken till date, as higher

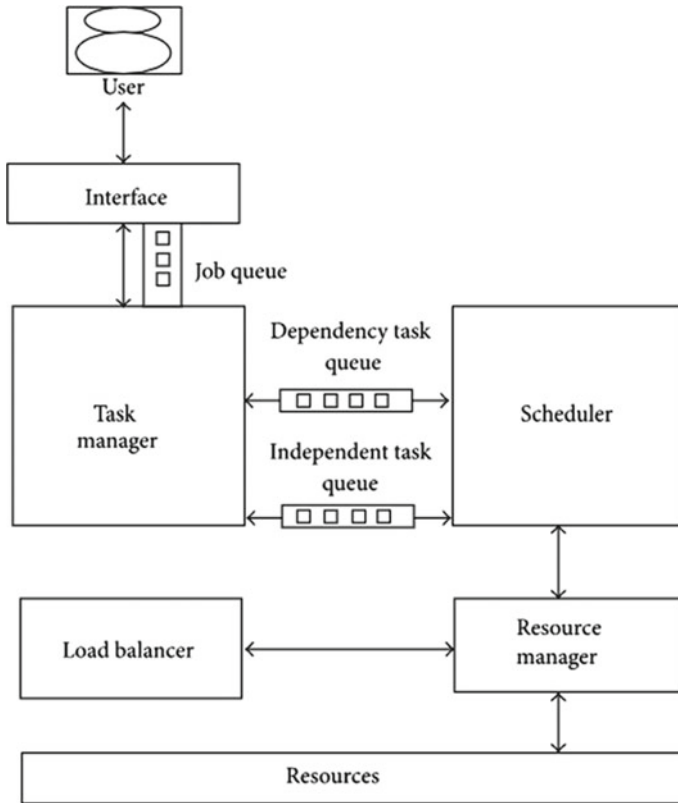
solutions could bring about extensive good value advantages. Previously a number of derivative-based procedures such as Lagrangian multiplier technique had been implemented to solve such problems. Those strategies require that incremental price curves are monotonically growing in nature. However in exercise, the input output traits of cutting-edge producing gadgets are extraordinarily nonlinear due to dynamic loadings, ramp-charge limits, multi-fuel alternatives, and many others. Their traits must be approximated to satisfy the necessities of classical dispatch algorithms. Because of such approximation the answer is handiest sub-most excellent and hence a big quantity of sales loss occurs over time. Fantastically nonlinear traits of these call for solution strategies that have no regulations directly to the form of the gasoline value curves.

#### ***4.1 Optimization as a Solution to Cloud Scheduling***

Biogeography based Optimization (BBO) is a recently proposed metaheuristic method for solving difficult combinatorial optimization problems. Biogeography depicts the way species drift from one island to any other, the way new species evolve, and the way species grow to be nonexistent. A habitat is any island (vicinity) that is geographically remote from other islands. Areas which might be well suited as houses for biological species are stated to have an excessive habitat suitability index (HSI). Factors that affect habitat suitability index consist of rainfall, range of flora, topographic capabilities, land vicinity, and temperature. The variables that represent habitability are known as suitability index variables (SIV). Suitability index variables may be considered as the independent variables of the habitat, and habitat suitability index can be calculated using these variables. This characteristic of BBO set of rules may be utilized to address the scheduling problem in clouds. Various characteristics of jobs such as cycles required for execution, priority, and arrival can be taken into account to get the optimum throughput possible for the job. The same concept can be repeated on various jobs in parallel and iteratively to get optimal solution. The BBO algorithm offers well dispensed and parallel processing solution to the cloud balancing problem. The Genetic algorithm operators, namely, mutation and crossover are to be incorporated into the BBO algorithm as part of the research work being proposed.

#### ***4.2 Architecture of Scheduling and Load Balancing***

Figure 3 shows the design of scheduling and load balancing in the proposed work, in which the scheduler/load balancer needs the reason to discover the most appropriate Virtual Machine (VM) and allocate the jobs from users towards selected VM(s) based



**Fig. 3** Architecture of resource allocation via load balancing and scheduling

on some fitness function or a given set of rules. The scheduler allocates jobs to the most appropriate VM(s), the VM being used minimum at that specific job arrival time.

It is the task of Load Balancer to identify the migration of various users' job from a fully utilized VM to an empty VM or at run-time, every time load Balancer identifies a free Virtual Machine using the status information available. A resource monitor is needed such that it communicates with every VM's resources and can collect the VM's capability (current load, maximum load (in MIPS)) on each VM, and no. of jobs in waiting or in current execution/queue at the VM. The task needs are given by the cloud users which include the length of the tasks (MIPS) designated to be executed and transfers the needs to the scheduler for decision making.

### 4.3 Working of Load Balancer

The working of a load balancer is depicted in Fig. 4. The various steps involved are:

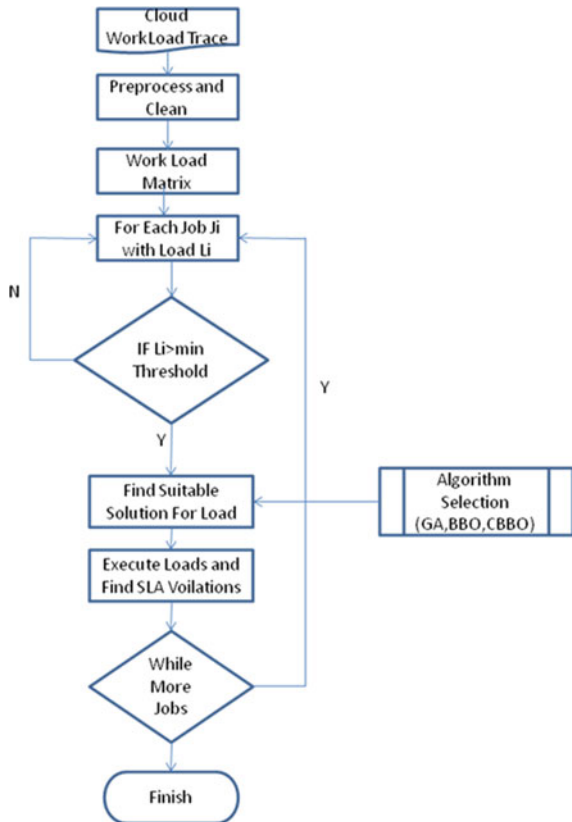
**Step 1 Importing Work Load Traces.** The first task of load balancer is to identify the incoming tasks to it and for this the cloud MAP reduce traces can be utilized which are available from OpenCloud[x] accessible at “<http://ftp.pdl.cmu.edu/pub/datasets/hla/dataset.html>”, OpenCloud is a research cluster at Carnegie Mellon University (CMU) managed by the CMU Parallel Data Lab.

**Step 2 Cleaning of Work Load Traces.** As the data set contains various unwanted and noisy data, preprocessing will be done to clean the data set.

**Step 3 Creation of Workload Matrix.** The workload matrix is created from Hadoop logs which can be understood by MATLAB and optimization algorithms can be applied.

**Step 4 Optimization.** In this phase for each log item a selected optimizer under test is applied, the optimizer set of optimizers will be GA, BBO and Improved BBO algorithms. This phase provides the thresholder Load Balancing cost of the algorithm

Fig. 4 Proposed flow chart of the green cloud load balancing algorithm



and optimal load assignment to VM (or identity of VM where we can transfer the tasks).

**Step 5 Assignment and Evaluation.** The fittest most VM is selected for execution and various SLA (Service Level Agreement) violations are evaluated for example if the job actually performed in specified time or not, how much cost it took for the given tasks MIPS, how much Energy it Consumed? Which Algorithm was more Green Power?

#### **4.4 Expected Outcome**

The proposed algorithm aims to develop an optimized scheduling solution for the cloud environment which will provide maximum utilization of the available bandwidth and will minimize the idle time. This will ultimately lead to an enhancement in energy efficiency and reduction in the carbon footprint of the scheduling solution.

### **5 Conclusion**

Cloud Computing is a way to move towards Green IT by eliminating the need of individual hardware and by providing a robust platform over the network. The cloud infrastructure can further be made more cost and energy efficient by scheduling its resources effectively. The dynamic nature of the cloud infrastructure makes it difficult to improve the system using an approximation or calculus based solution rather the utilization can be optimized using heuristic optimization techniques. Different optimization techniques/algorithms can be implemented and can be analyzed for the performance. Two or more optimization techniques/algorithms can be combined together and a hybrid technique or algorithm can be implemented to optimize the resource utilization and load balancing of the cloud infrastructure.

### **References**

1. Dastjerdi A, Buyya R (2012) An autonomous reliability-aware negotiation strategy for cloud computing environments. In: Proceedings of 12th IEEE/ACM international symposium on cluster, cloud and grid computing. IEEE, pp 284–291
2. Hung PP, Van Nguyen M, Aazam M, Huh E-N (2014) Task scheduling for optimizing recovery time in cloud computing. In: 2014 international conference on IEEE, computing, management and telecommunications (ComManTel), 2014, pp 188–193
3. Menzel M, Ranjan R, Wang L, Khan S, Chen J (2014) CloudGenius: a hybrid decision support method for automating the migration of web application clusters to public clouds. In: IEEE transactions on computers 2014, pp 1336–1348
4. Prasad AS, Rao S (2014) A mechanism design approach to resource procurement in cloud computing. IEEE Trans Comput 63(1):17–30

5. Rahman M, Iqbal S, Gao J (2014) Load balancer as a service in cloud computing. In: 2014 IEEE 8th international symposium on service oriented system engineering (SOSE), 2014. IEEE, pp 204–211
6. Sarbazi-Azad H, Zomaya A (2014) Market-oriented cloud computing and the cloudbus toolkit. Wiley-IEEE Press, Large Scale Network-Centric Distributed Systems
7. Takako Endo P, Batista MS, Goncalves GE, Rodrigues M, Sadok D, Kelner J, Sefidcon A, Wuhib F, Self-organizing strategies for resource management in cloud computing: state-of-the-art and challenges. In: 2nd IEEE Latin American conference on cloud computing and communications (LatinCloud), 2013. IEEE, pp 13–18
8. Tawfeek MA, El-Sisi A, Keshk AE, Torkey FA Cloud task scheduling based on ant colony optimization. In: 2013 8th international conference on computer engineering & systems (ICCES), 2013. IEEE, pp 129–137
9. Verma A, Kaushal S Budget constrained priority based genetic algorithm for workflow scheduling in cloud. In: Fifth international conference on advances in recent technologies in communication and computing (ARTCom 2013), 2013. IET, pp 216–222
10. Sun W, Zhang N, Wang H, Yin W, Qiu T (2013) PACO: a period ACO based scheduling algorithm in cloud computing. In: 2013 international conference on cloud computing and big data (CloudCom-Asia), 2013. IEEE, pp 482–486

# Detection of Hazardous Analyte Using Transparent Gate Thin-Film Transistor



Ajay Kumar, Amit Kumar Goyal, Manan Roy, Neha Gupta, MM Tripathi and Rishu Chaujar

**Abstract** The work mainly focuses on the detection of a hazardous analyte like silicon carbide using TGTFT which has an ITO gate, a bio-receptor of silicon nitride and further compared with the same device with an additional analyte, i.e., silicon carbide added into it. Transfer characteristics and some more electrical properties have been simulated and compared. The drain current ( $I_d$ ) of the analyte TFT increased by 21.59% in contrast to the device which has air. A substantial increment of 17.34% in the electric field of the analyte-added TFT was observed in contrast to the air-filled TGTFT. Some changes were observed in valence band energy (VBE) and conduction band energy (CBE) of analyte device and without analyte device. The addition of analyte changes chemical composition of interface, i.e., changes the electron concentration at interface and therefore altering the effect of specific gate voltage as potential of the interface changes resulting in different results from the one with no analyte. Therefore, changes in the electrical properties of the device pave the way of hazardous analyte detection.

**Keywords** TGTFT · ITO · Bio-receptor · Analyte

---

A. Kumar (✉) · A. K. Goyal · M. Roy  
ECE Department, JIIT, Sector 62, Noida, UP, India  
e-mail: [ajaykumar@dtu.ac.in](mailto:ajaykumar@dtu.ac.in)

A. K. Goyal  
e-mail: [amit.goyal@jiit.ac.in](mailto:amit.goyal@jiit.ac.in)

M. Roy  
e-mail: [manankkt@gmail.com](mailto:manankkt@gmail.com)

N. Gupta  
AHS Department, AGITM (NIEC), New Delhi, India  
e-mail: [gupta.neha@niecdelhi.ac.in](mailto:gupta.neha@niecdelhi.ac.in)

M. Tripathi · R. Chaujar  
EE Department and Applied Physics Department, DTU, New Delhi, India  
e-mail: [mmtripathi@dce.ac.in](mailto:mmtripathi@dce.ac.in)

R. Chaujar  
e-mail: [chaujar.rishu@dtu.ac.in](mailto:chaujar.rishu@dtu.ac.in)



## 1 Introduction

Biosensor is frequently used as an analytical device for chemical substances detection that incorporates a biological component with a physicochemical detector [1, 2]. The sensitive biological elements like tissues, organelles, microorganisms, antibodies, enzymes, cell receptors, and nucleic acid, are biologically developed chemicals which interact, tie, or detect with the analyte under the simulation [3]. Bio-TFTs [4–6] are generally designed by a transistor device coupled with the bio-sensitive layer that can successfully detect hazardous biomolecules such as proteins and nucleic acids. By using field-effect transistor (FET), a bio-TFT has configured. Bio-TFT acts as a transducer which separated  $\text{SiO}_2$  (e.g., an insulating layer) from receptors or probe molecules (e.g., the biological recognition element) [7] which are biased to the target molecule called analyte. Silicon carbide can affect badly when breathed in. Silicon carbide can cause irritation in the eyes and nose on contact.

TFTs are widely used as imaging sensor arrays (ISAs) and active matrix displays applications with low power consumption [8, 9]. For the latest upcoming dynamic displays, staggered top-gate thin-film transistors (STG-TFTs) have been recognised to be the most persuasive option. To improve the gate controllability, a transparent material has been used in the gate. Many transparent materials with their oxide components are available; indium tin oxide (ITO) [10–18] is the frequently used transparent material due to its electrical and conductive properties like higher mobility [19], moderate temperature fabrication, uniform structure, and fluent transparency. It also requires lower deposition temperature [20, 21], low resistivity ( $10^{-5} \Omega \text{ cm}$ ), with high hall mobility ( $53.5 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ ), and high concentration ( $10^{21} \text{ cm}^{-3}$ ) [22]. In display matrices of TGTFT, Si is widely operated as the substrate. Although it shows a lot of hardships like heavy base and imprudent fragile built, Si is replaced with a unique substrate, which is universally accepted to fabricate screens which are light, thin, and non-breakable.

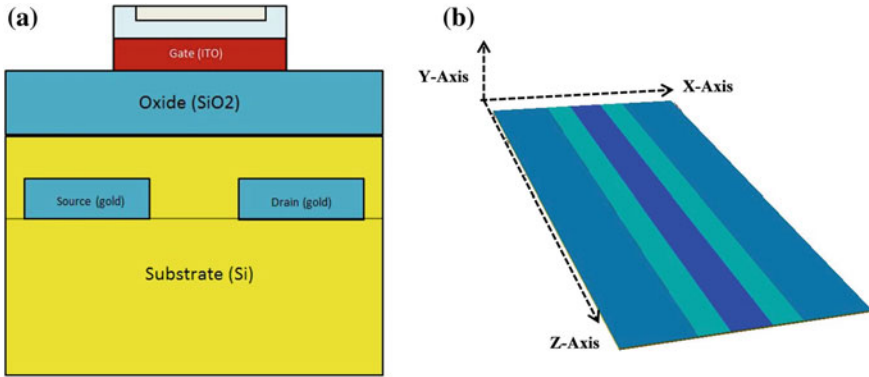
## 2 Device Design and Simulation Methodology

The design for TGTFT incorporated in the work is presented in Fig. 1. Figure 1a depicts the top geometry of the device; however, Fig. 1b reflects the three-dimensional geometry of the TGTFT. The device specifications are shown in Table 1.

ATLAS device simulator is used for this work [23]. For the simulation purpose, FLDMOB (parallel electric field dependence) and KLASRH model are considered [23]. The following models and equations [23] are used for the TGTFT simulation process.

### 1. Poisson's equation

$$\text{div}(\varepsilon \nabla \varphi) = -\rho \quad (1)$$



**Fig. 1** a Two-dimensional structure of TGTFT; b three-dimensional top view of TGTFT

**Table 1** TGTFT design parameters

Parameters	Width (nm)	Thickness (nm)
Gate	10	0.02
Oxide	24	0.02
Channel	24	0.09
Source	4	0.02
Drain	4	0.02
Substrate	24	0.07

where  $\epsilon$  is permittivity,  $\phi$  is potential, and  $\rho$  is space charge density.

2. *Current continuity equation*

For holes,

$$\frac{\partial p}{\partial t} = \frac{1}{q} \text{div } \vec{J}_p + G_p - R_p \tag{2}$$

For electrons,

$$\frac{\partial n}{\partial t} = \frac{1}{q} \text{div } \vec{J}_n + G_n - R_n \tag{3}$$

where  $\vec{J}_p$  is holes' current density, and  $\vec{J}_n$  is electrons' current density;  $G_p$  is the generation rate for holes ( $p$ ), and  $G_n$  is the generation rate for electrons ( $n$ );  $R_p$  and  $R_n$  are holes and electrons recombination rates, and ' $q$ ' is the charge.

3. *Recombination*

3.1 *Shockley-Read-Hall (SRH)*

$$R_{SRH} = \frac{pn - n_{ie}^2}{\tau_p[n + n_{ie}e^{-(E_{TRAP}/kT_L)}] + \tau_n[p + n_{ie}e^{-(E_{TRAP}/kT_L)}]} \tag{4}$$

where  $E_{TRAP}$  changes in the fermi levels;  $\tau_n$  is electron lifetime; and  $\tau_p$  is the hole lifetime.

3.2 Auger recombination

$$R_{Auger} = AUGN(pn^2 - nn_{ie}^2) + AUGP(np^2 - pn_{ie}^2) \tag{5}$$

where  $AUGP = 1.8 \times 10^{-31} \text{ cm}^6/\text{s}$  and  $AUGN = 8.3 \times 10^{-32} \text{ cm}^6/\text{s}$ .

4. Parallel electric field-dependent mobility

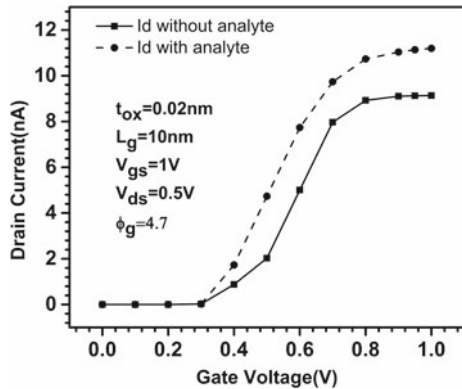
$$\mu(E) = \mu_0 \left[ \frac{1}{1 + (\mu_0 E / v_{sat})^\beta} \right]^{1/\beta} \tag{6}$$

### 3 Simulation Outcomes With Explanations

Figure 2a illustrates the transfer characteristics of the TGTFTs with and without analyte. When the hazardous analyte is placed on the gate, then the current changes; this change detect the presence of hazardous analyte in that particular environment. Moreover, the switching ratio ( $I_{on}/I_{off}$ ) of TGTFT has been evaluated with analyte and without analyte, and it increases in the presence of analyte as compared to one filled with air and is shown in Fig. 3a owing to increase in  $I_{on}$  (Fig. 2).

The detection of the presence of analyte in the environment has been detected by the change in threshold voltage ( $V_{th}$ ), and it is found to be 0.2 V for the TGTFT

**Fig. 2** TGTFT’s transfer characteristics with and without analyte



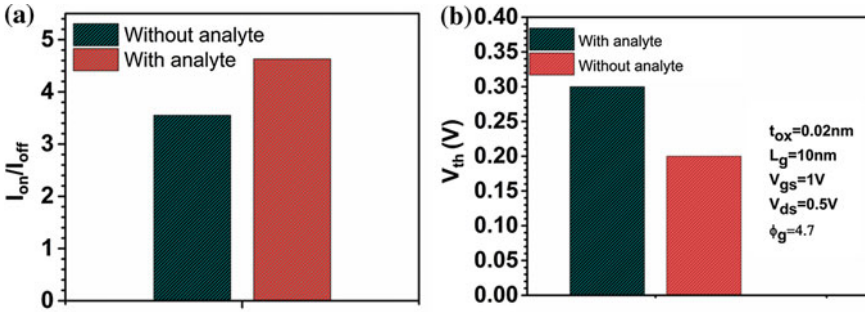


Fig. 3 a TGTFT's  $I_{on}/I_{off}$  with and without analyte; b TGTFT's  $V_{th}$  with and without analyte

without analyte and 0.3 V for TGTFT with analyte and is depicted in Fig. 3b, and the parallel value of  $I_d$  is also presented in the graph. The change in threshold voltage reflects the presence of analyte.

Figure 4a reflects the electric field of TGTFT with analyte and then without analyte, and Fig. 4b reflects the contour diagram of electric field for the TGTFT. The electric field is showing peaks at the ends of the source and drain near which the junction evolution happens. The electric field of source is more than that of drain due to pinch-off phenomenon at the drain surface. The addition of analyte changes chemical composition of interface, i.e., changes the electron concentration at interface and therefore altering the effect of specific gate voltage as potential of the interface changes resulting in different results from the one with no analyte. In physics language, the electron mobility concludes how fast an electron can move through a metal or semiconductor when hauled by an electric field. It is higher at source and drain as they indulge moving charge carriers  $n+$  and is shown in Fig. 5a and its contour in 5b.

The addition of analyte changes chemical composition of interface, i.e., changes the electron concentration at interface and therefore altering the effect of specific

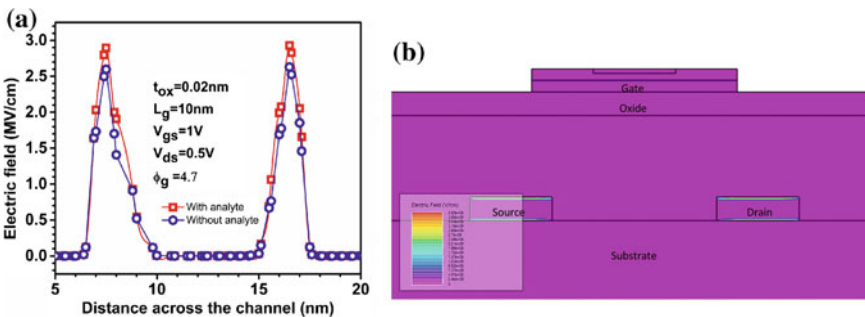


Fig. 4 a Electric field along the channel for TGTFT and CTFT; b contour diagram of electric field of TGTFT with analyte

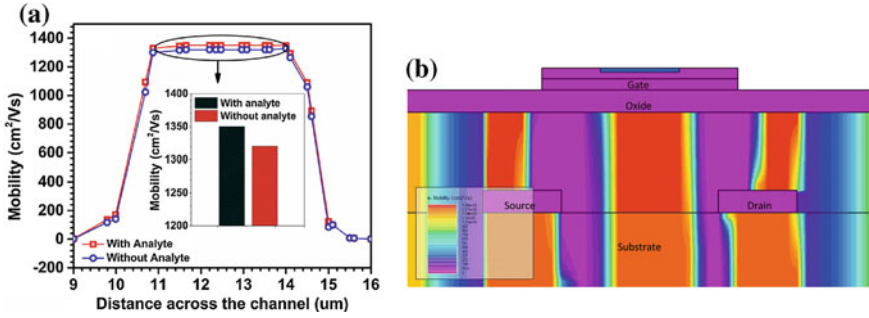


Fig. 5 a Electron mobility along the channel for TGFTT and CTFT; b contour diagram of electron mobility of TGFTT

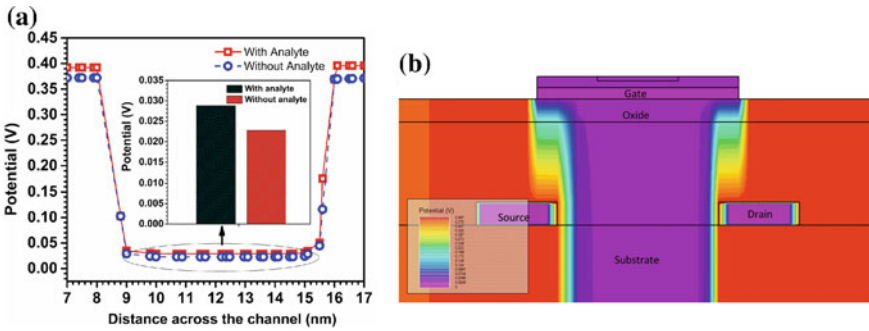
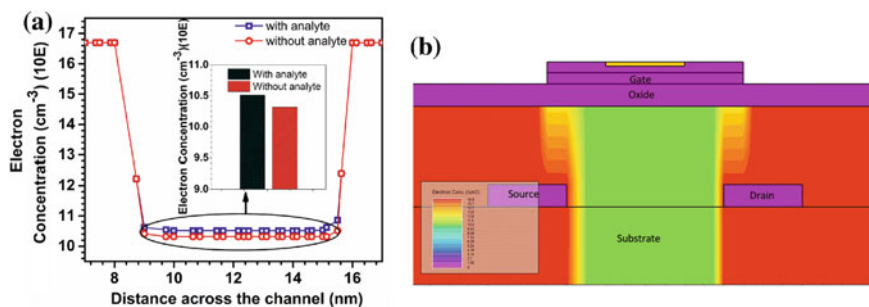


Fig. 6 a Surface potential along the channel for TGFTT with and without analyte; b contour diagram of surface potential for TGFTT with analyte

gate voltage as potential of the interface changes with respect to the potential when analyte is absent. The potential along the channel of transparent gate TFT was found more than that of without analyte, i.e., air-filled device, and is depicted in Fig. 6a and mobility contour of TGFTT in Fig. 6b. The change in potential of surface identifies the existence of hazardous analyte in the environment. Furthermore, electron concentration is observed in the presence of hazardous analyte and with air filled.

From Fig. 7a, it is observed that in the presence of hazardous analyte, the concentration of electron increases which causes change in drain current as shown in Fig. 2, switching ratio (SR) as depicted in Fig. 3a and threshold voltage (shown in Fig. 3b). The same result is clearly reflected in the contour plot of electron concentration as shown in Fig. 7b.



**Fig. 7** **a** Electron concentration along the channel for TGTFT with and without analyte; **b** contour diagram of electron concentration with analyte

## 4 Conclusion

Here, the three-dimensional schematic of TGTFT is made with the help of ITO gate and Au (Gold) contacts along with a bio-receptor of silicon nitride and is compared with the same device with additional analyte, i.e., silicon carbide in it, to detect its presence. Drain current to input voltage and other electric effects is studied. The  $I_d$  of the analyte TFT increased by 21.59% in contrast to the basic structure. Substantial increase of 17.34% in the electric field of the analyte-added TFT in contrast to conventional TFT is seen. The VBE and CBE of TGTFT are also varied in analyte device than without analyte structure. The addition of analyte changes chemical composition of interface and therefore altering the effect of specific gate voltage resulting in different results from the one with no analyte. Finally, because of changes in the property of device that contains the silicon carbide analyte, its presence can be detected.

**Acknowledgements** The authors fully acknowledge help and support from IIIT, Noida.

## References

1. Karube I, Muramatsu H (1988) Analytical device and method utilizing a piezoelectric crystal biosensor. Google Patents
2. Oshiman E, Shiraki Y (2010) Analytical device having temperature detection unit. Google Patents
3. Lazcka O, Del Campo FJ, Munoz FX (2007) Pathogen detection: a perspective of traditional methods and biosensors. *Biosens Bioelectron* 22(7):1205–1217
4. Stewart KA, Gouliouk V, Kesler DA, Wager JF (2017) Sputtered boron indium oxide thin-film transistors. *Solid-State Electron* 137:80–84
5. Shen Y-C, Yang C-H, Chen S-W, Wu S-H, Yang T-L, Huang J-J (2014) IGZO thin film transistor biosensors functionalized with ZnO nanorods and antibodies. *Biosens Bioelectron* 54:306–310
6. Kim SM, Park S, Lee WJ, Yoon MH (2017) Electronics and bioelectronic interfaces. *Handbook of Solid State Chemistry*, pp 75–92

7. Sibley DR, Monsma FJ Jr (1992) Molecular biology of dopamine receptors. *Trends Pharmacol Sci* 13:61–69
8. Gong Y, Jackson TN (2017) Offset drain ZnO thin-film transistors for high-voltage operation. *IEEE Electron Device Lett* 38(8):1047–1050
9. Zhao J, Tang W, Yu P, Guo X (2016) Low voltage organic thin-film transistor with reduced sub-gap DOS for power efficient logic circuits. In: 2016 7th international conference on computer aided design for thin-film transistor technologies (CAD-TFT), IEEE, pp 1–1
10. Kumar A (2017) Effect of trench depth and gate length shrinking assessment on the analog and linearity performance of TGRC-MOSFET. *Superlattices Microstruct* 109:626–640
11. Kumar A, Gupta N, Chaujar R (2016) Power gain assessment of ITO based transparent gate recessed channel (TGRC) MOSFET for RF/wireless applications. *Superlattices Microstruct* 91:290–301
12. Kumar A, Gupta N, Chaujar R (2016) TCAD RF performance investigation of transparent gate recessed channel MOSFET. *Microelectron J* 49:36–42
13. Kumar A, Gupta N, Chaujar R (2016) Analysis of novel transparent gate recessed channel (TGRC) MOSFET for improved analog behaviour. *Microsyst Technol* 22(11):2665–2671
14. Kumar A, Gupta N, Chaujar R (2017) Effect of structured parameters on the hot-carrier immunity of transparent gate recessed channel (TGRC) MOSFET. *Microsyst Technol* 23(9):4057–4064
15. Kumar A, Tiwari B, Singh S, Tripathi MM, Chaujar R (2018) Radiation analysis of N-channel TGRC-MOSFET: an X-Ray dosimeter. *IEEE Trans Electron Devices* 65(11):5014–5020
16. Kumar A, Tripathi M, Chaujar R (2017) Investigation of parasitic capacitances of In<sub>2</sub>O<sub>5</sub>Sn gate electrode recessed channel MOSFET for ULSI switching applications. *Microsyst Technol* 23(12):5867–5874
17. Kumar A, Tripathi M, Chaujar R (2018) Reliability issues of In<sub>2</sub>O<sub>5</sub>Sn gate electrode recessed channel MOSFET: impact of interface trap charges and temperature. *IEEE Trans Electron Devices* 65:860–866
18. Kumar A, Tripathi MM, Chaujar R (2018) In<sub>2</sub>O<sub>5</sub>Sn based transparent gate recessed channel MOSFET: RF small-signal model for microwave applications. *AEU—Int J Electron Commun* 93:233–241
19. Xu J, Xie H, Liu G, Zhang L, Tong X, Dong C (2016) Ambient effects on the light illumination stability of amorphous InGaZnO thin film transistors. In: 2016 7th international conference on computer aided design for thin-film transistor technologies (CAD-TFT), IEEE, pp 1–1
20. Singh V, Suman C, Kumar S (2006) Indium tin oxide (ITO) films on flexible substrates for organic light emitting diodes. In: *Proceedings of ASID* p 388
21. Murali R, Meindl JD (2007) Modeling the effect of source/drain junction depth on bulk-MOSFET scaling. *Solid-state electronics* 51(6):823–827
22. Minami T (2005) Transparent conducting oxide semiconductors for transparent electrodes. *Semicond Sci Technol* 20(4):S35–S44
23. Silvaco I (2011) ATLAS user's manual. Santa Clara, CA, Version 5

# A Robustness Analysis of Different Nonlinear Autoregressive Networks Using Monte Carlo Simulations for Predicting High Fluctuation Rainfall



Tien-Thinh Le, Binh Thai Pham, Vuong Minh Le, Hai-Bang Ly  
and Lu Minh Le

**Abstract** In this study, the main objective is to carry out the robustness analysis of an artificial intelligence (AI) approach, namely nonlinear autoregressive neural networks (NAR) using Monte Carlo simulations for predicting the high fluctuation rainfall. Various algorithms of the NAR including Levenberg–Marquardt (LM), Bayesian regularization (BR) and scaled conjugate gradient (SCG) were developed. Statistical criteria, namely coefficient of determination ( $R^2$ ), root mean squared error (RMSE) and mean absolute error (MAE), were used to quantify the impact of fluctuations on the prediction output. Results showed that SCG algorithm was not sufficiently robust, while LM and BR methods exposed a strong capability in forecasting daily rainfall. In addition, prediction using BR was slightly better than LM, especially in terms of standard deviation of  $R^2$ , RMSE and MAE distributions over 500 Monte Carlo realizations.

**Keywords** Statistical analysis · Nonlinear autoregressive neural networks · Rainfall · Training algorithm · Fluctuation

## 1 Introduction

Rainfall has a direct impact on various domains including agriculture [1], water resource management [2] and hydroelectricity [3]. Heavy rain was also the principal

---

T.-T. Le · V. M. Le  
NTT Hi-Tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City 700000, Vietnam

B. T. Pham (✉) · H.-B. Ly  
University of Transport Technology, Hanoi 100000, Vietnam  
e-mail: [binhpt@utt.edu.vn](mailto:binhpt@utt.edu.vn)

H.-B. Ly  
e-mail: [banglh@utt.edu.vn](mailto:banglh@utt.edu.vn)

L. M. Le  
Faculty of Engineering, Vietnam National University of Agriculture, Gia Lam, Hanoi 100000, Vietnam  
e-mail: [lmlu@vnu.edu.vn](mailto:lmlu@vnu.edu.vn)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_21](https://doi.org/10.1007/978-981-15-2329-8_21)



cause of many natural disasters, for instance, flood [4] or landslide [5]. Such atmospheric process production exposed high order of fluctuations in time as measured in meteorological stations [6, 7]. Consequently, this variability makes analysis of rain phenomenon more complicated, especially in forecasting [8].

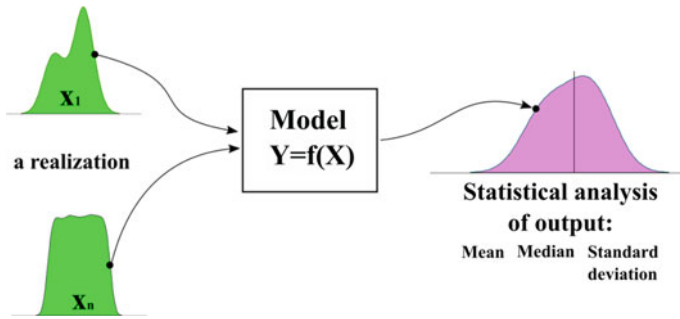
Recently, artificial intelligent (AI)-based models have been widely proposed for forecasting rainfall, especially using artificial neural networks (ANN) [9, 10]. Chattopadhyay et al. [11] have developed an ANN model trained using conjugate gradient descent technique for predicting the average summer monsoon rainfall in India. In another study, Chand [12] introduced a statistical comparison between a neural network and a traditional autoregressive integrated moving average method for the forecast of rainfall pattern in Hyderabad region, India, for a 104-year period. Dabhi and Chaudhary [13] have improved the neural network in rainfall prediction by reducing the stochastic noise in meteorological data using wavelet techniques. However, rain is one of the most fluctuated nonlinear problems of the atmospheric processes [14], and therefore, the study on the prediction of rainfall taking into account high order variations of data remains challenging for researchers, particularly from statistical point of view. For that reason, the main objective of this work is to develop a robust and efficient neural network model in order to track the nonlinear behavior of rainfall time series as well as its fluctuation level using Monte Carlo method. With this aim, a nonlinear autoregressive neural network (NAR) was proposed and validated using the data collected from the meteorological stations. Three training algorithms such as Levenberg–Marquardt (LM), Bayesian regularization (BR) and scaled conjugate gradient (SCG) were, respectively, used to train the developed AI models. Monte Carlo simulations were finally performed in order to evaluate the robustness of each training algorithm under a high order of data fluctuations.

## 2 Material and Method

### 2.1 Methods Used

In this study, a prediction model based on nonlinear autoregressive neural networks (NAR) was adopted for investigating rainfall time series. In such model, a stochastic difference equation was introduced to correlate previous and future values of the atmospheric random process [15]. In this case of meteorological phenomena, the stochastic equation is nonlinear because of the behavior of the considered time-dependent variable, i.e., rainfall [16]. In the literature, NAR has also been constructed for predicting various time series data such as industrial index of production [17] or air temperature [18].

Various criteria, namely coefficient of determination ( $R^2$ ), root mean squared error (RMSE) and mean absolute error (MAE), were used in this research in order to quantify the performance of NAR model. Details of formulation of such criteria could be found in [19].



**Fig. 1** Monte Carlo method for accounting variability of input under statistical analysis of output result

In this paper, three training algorithms such as Levenberg–Marquardt (LM), Bayesian regularization (BR) and scaled conjugate gradient (SCG) [20–22] have been used in order to evaluate the performance of NAR model under high fluctuation of data. Monte Carlo method [23] was applied to statistically characterize the quality of prediction model using each previously mentioned training algorithm. A schematization of Monte Carlo method accounting variability of input on the output result is presented in Fig. 1. A convergence function was also introduced in order to estimate the stationary domain of Monte Carlo runs, as in the following [24]:

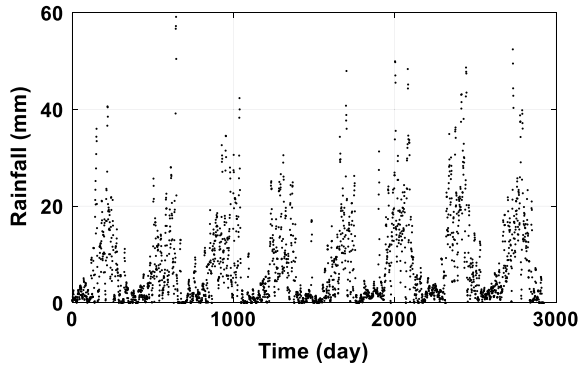
$$f = \frac{1}{NMC} \sum_{i=1}^{NMC} G_i, \tag{1}$$

where NMC is the number of Monte Carlo runs and  $G$  is the random variable of interest (such as  $R^2$ , RMSE, MAE in this paper). The convergence function was helpful in order to detect optimal number of Monte Carlo realizations and ensure a reliable statistical analysis later.

## 2.2 Data Used

In this work, values of rainfall time series (in mm) were collected from a rain gauge at Cao Phong district, Hoa Binh, Vietnam, for a period between 2006 and 2013. High order of data fluctuation is shown in Fig. 1. First 70% of data were used for training the NAR model, and last 30% values were served for the testing part (Fig. 2).

**Fig. 2** Time series form of rainfall data

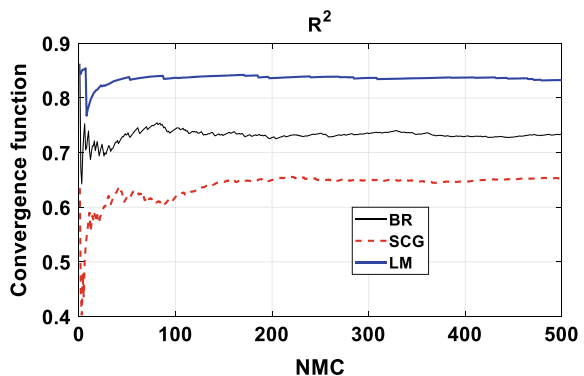


### 3 Results and Discussion

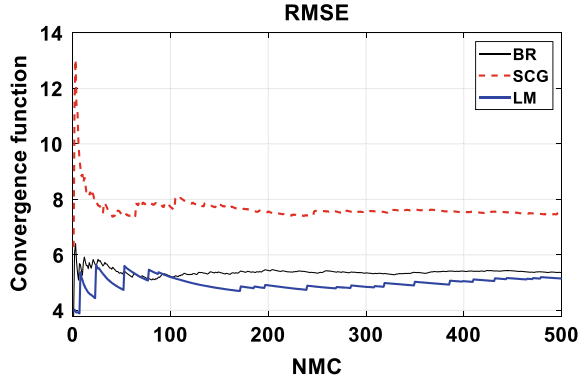
Rainfall data were used to train the NAR model using three training algorithms presented in the previous section. 500 Monte Carlo runs were also conducted in order to take into account data fluctuation’s influences on the efficiency of training algorithms. Distribution of quality parameters for the testing part such as  $R^2$ , RMSE and MAE was obtained over 500 Monte Carlo realizations. Convergence function of  $R^2$ , RMSE and MAE is presented in Figs. 3, 4 and 5, respectively. Statistical analysis was finally applied in order to evaluate the performance of each training algorithms. Summary of statistical information is highlighted in Table 1, including quantiles Q25, Q50, Q75, mean and standard deviation of  $R^2$ , RMSE and MAE distributions.

As observed in Figs. 3, 4 and 5, all three criteria such as  $R^2$ , RMSE and MAE reached the stationary solution over 500 Monte Carlo realizations. It means such number of runs was enough to explore statistical behavior of the problem. It is also seen that NAR model using SCG training algorithm exhibits highest order of fluctuation around the stationary solution with respect to all three criteria  $R^2$ , RMSE and MAE. Such observation was confirmed regarding statistical analysis in Table 1

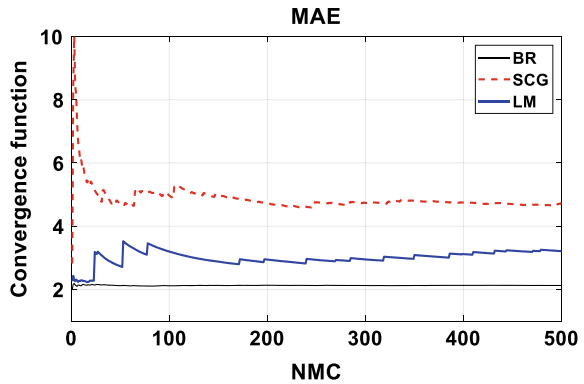
**Fig. 3** Convergence function of  $R^2$  over 500 Monte Carlo realizations using three training algorithms



**Fig. 4** Convergence function of RMSE over 500 Monte Carlo realizations using three training algorithms



**Fig. 5** Convergence function of MAE over 500 Monte Carlo realizations using three training algorithms



**Table 1** Summary of statistical information

Criteria	Algorithm	Q25	Median	Q75	Mean	StD
$R^2$	BR	0.584	0.863	0.863	0.737	0.189
	SCG	0.573	0.710	0.778	0.645	0.192
	LM	0.842	0.854	0.861	0.832	0.105
RMSE	BR	3.776	3.779	7.052	5.325	2.362
	SCG	5.083	6.226	8.468	8.115	6.744
	LM	3.815	3.936	4.081	5.342	7.499
MAE	BR	2.000	2.002	2.319	2.121	0.174
	SCG	2.839	3.590	5.053	5.157	5.943
	LM	2.103	2.220	2.324	3.390	6.763

(mean of  $R^2$  is the lowest value, while mean of RMSE and MAE are the highest values). It could be deduced that SCG algorithm was not robust and efficient to take into account fluctuation of rainfall data using NAR prediction model.

On the contrary, BR and LM algorithms exhibited strong capability for predicting rainfall. As indicated in Table 1 for mean value, mean of  $R^2$  distribution is 0.737 and 0.832, mean of RMSE distribution is 5.325 and 5.342, mean of MAE distribution is 2.121 and 3.390 using BR and LM algorithms, respectively. In the case of standard deviation, StD of  $R^2$  distribution is 0.189 and 0.105, StD of RMSE distribution is 2.362 and 7.499, and StD of MAE distribution is 0.174 and 6.763 using BR and LM algorithms, respectively. Note that high value of standard deviation obtained using LM algorithm is the consequence of large outlier values in the distribution. It means that the LM algorithm was sensitive to outlier in this problem and could generate extreme values of prediction. However, NAR model training with LM algorithm gave the best results with respect to the mean value of  $R^2$  and RMSE distributions.

To conclude, from overall statistical analysis with respect to distribution of  $R^2$ , RMSE and MAE over 500 Monte Carlo runs, nonlinear autoregressive neural networks using Bayesian regularization algorithm for training were the most robust and efficient (i.e., low coefficient of variation of all three  $R^2$ , RMSE and MAE distributions). Similarly, BR was proved as the most effective training technique in terms of predictive ability in the work of Kayri [25]. In a series of paper, Okut et al. [26, 27] have also demonstrated the highest capability of BR algorithm in capturing nonlinear relationships in data. In other investigations, BR technique exhibited the best performance in comparison with other training methods [28, 29], especially for time series data [30, 31].

## 4 Conclusion

In this paper, a robustness analysis of the NAR using different algorithms (LM, BR and SCG) was done for the prediction of high fluctuation rainfall. Monte Carlo simulations were used to evaluate the convergence of these algorithms. Various quality assessment criteria, namely  $R^2$ , RMSE and MAE, were used in order to evaluate the performance of the developed model. Results showed that BR algorithm was the most robust and efficient for the prediction of rainfall compared with other algorithms (LM and BR). The developed model was a potential candidate for the prediction of rainfall and could contribute to improve water resource management in many fields such as urban drainage systems, agriculture and hydroelectric production. However, in the future study, the influence of other meteorological variables such as temperature or relative humidity should be taken into account as they play also an important role in the rain process. Beside, noise reduction techniques might also be interested in order to reduce the impact of outlier data on training AI model.

**Conflict of Interest** The author declares no conflict of interest.

## References

1. Dash Y, Mishra SK, Sahany S, Panigrahi BK (2018) Indian summer monsoon rainfall prediction: a comparison of iterative and non-iterative approaches. *Appl Soft Comput* 70:1122–1134. <https://doi.org/10.1016/j.asoc.2017.08.055>
2. Hartmann H, Snow JA, Stein S, Su B, Zhai J, Jiang T, Krysanova V, Kundzewicz ZW (2016) Predictors of precipitation for improved water resources management in the Tarim River basin: creating a seasonal forecast model. *J Arid Environ* 125:31–42. <https://doi.org/10.1016/j.jaridenv.2015.09.010>
3. Haddad MS (2011) Capacity choice and water management in hydroelectricity systems. *Energy Econ* 33:168–177. <https://doi.org/10.1016/j.eneco.2010.05.005>
4. Toda K (2007) Urban flooding and measures. *J Disaster Res* 2:143–152. <https://doi.org/10.20965/jdr.2007.p0143>
5. Pham BT, Bui DT, Prakash I, Dholakia MB (2017) Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *CATENA* 149:52–63. <https://doi.org/10.1016/j.catena.2016.09.007>
6. Villarini G, Seo B-C, Serinaldi F, Krajewski WF (2014) Spatial and temporal modeling of radar rainfall uncertainties. *Atmos Res* 135–136:91–101. <https://doi.org/10.1016/j.atmosres.2013.09.007>
7. Sang XL, Su YZ, Xiao HJ, Wang H, Xu JX (2013) Prediction of rainfall in Da-Dong-Yong hydrologic station based on wavelet neural network. <https://www.scientific.net/AMR.726-731.3279>
8. Darji MP, Dabhi VK, Prajapati HB (2015) Rainfall forecasting using neural network: a survey. In: 2015 international conference on advances in computer engineering and applications, pp 706–713
9. Abbot J, Marohasy J (2017) Skilful rainfall forecasts from artificial neural networks with long duration series and single-month optimization. *Atmos Res* 197:289–299. <https://doi.org/10.1016/j.atmosres.2017.07.015>
10. Haviluddin M, Hardwinarto S., Sumaryono Aipassa M (2015) Rainfall monthly prediction based on artificial neural network: a case study in Tenggarong Station, East Kalimantan—Indonesia. *Procedia Comput Sci* 59:142–151. <https://doi.org/10.1016/j.procs.2015.07.528>
11. Chattopadhyay S (2006) Anticipation of summer monsoon rainfall over India by artificial neural network with conjugate gradient descent learning. [arXiv:nlin/0611010](https://arxiv.org/abs/0611010)
12. Chand RV (2006) Modelling and prediction of rainfall using artificial neural network and ARIMA techniques. *J Ind Geophys Union* 10(2):141–151
13. Dabhi VK, Chaudhary S Hybrid wavelet-postfix-GP model for rainfall prediction of Anand Region of India. <https://www.hindawi.com/journals/aai/2014/717803/>
14. Guhathakurta P (2008) Long lead monsoon rainfall prediction for meteorological sub-divisions of India using deterministic artificial neural network model. *Meteorol Atmos Phys* 101:93–108
15. Billings SA (2013) Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains. Wiley
16. Wu CL, Chau KW, Fan C (2010) Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *J Hydrol* 389:146–167
17. Potdar K, Kinnerkar R (2017) A non-linear autoregressive neural network model for forecasting Indian index of industrial production. In: 2017 IEEE region 10 symposium (TENSYP), pp 1–5
18. Islam MP, Morimoto T (2017) Non-linear autoregressive neural network approach for inside air temperature prediction of a pillar cooler. *Int J Green Energy* 14:141–149. <https://doi.org/10.1080/15435075.2016.1251925>
19. Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res* 30:79–82. <https://doi.org/10.3354/cr030079>

20. Hagan MT, Menhaj MB (1994) Training feedforward networks with the Marquardt algorithm. *IEEE Trans Neural Netw* 5:989–993. <https://doi.org/10.1109/72.329697>
21. Sharma A, Goyal MK (2017) A comparison of three soft computing techniques, Bayesian regression, support vector regression, and wavelet regression, for monthly rainfall forecast. *J Intell Syst* 26:641–655. <https://doi.org/10.1515/jisys-2016-0065>
22. Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* 6:525–533. [https://doi.org/10.1016/S0893-6080\(05\)80056-5](https://doi.org/10.1016/S0893-6080(05)80056-5)
23. Mordechai S (2011) Applications of Monte Carlo method in science and engineering
24. Soize C (2005) Random matrix theory for modeling uncertainties in computational mechanics. *Comput Methods Appl Mech Eng* 194:1333–1366. <https://doi.org/10.1016/j.cma.2004.06.038>
25. Kayri M (2016) Predictive abilities of Bayesian regularization and Levenberg–Marquardt algorithms in artificial neural networks: a comparative empirical study on social data. *Math Comput Appl* 21:20. <https://doi.org/10.3390/mca21020020>
26. Okut H, Gianola D, Rosa GJM, Weigel KA (2011) Prediction of body mass index in mice using dense molecular markers and a regularized neural network. *Genet Res (Camb)* 93:189–201. <https://doi.org/10.1017/S0016672310000662>
27. Okut H, Wu X-L, Rosa GJ, Bauck S, Woodward BW, Schnabel RD, Taylor JF, Gianola D (2013) Predicting expected progeny difference for marbling score in Angus cattle using artificial neural networks and Bayesian regression models. *Genet Sel Evol* 45:34. <https://doi.org/10.1186/1297-9686-45-34>
28. Bruneau P, McElroy NR (2006) logD7.4 modeling using Bayesian regularized neural networks. Assessment and correction of the errors of prediction. *J Chem Inf Model* 46:1379–1387. <https://doi.org/10.1021/ci0504014>
29. Saini LM (2008) Peak load forecasting using Bayesian regularization, resilient and adaptive backpropagation learning based artificial neural networks. *Electr Power Syst Res* 78:1302–1310. <https://doi.org/10.1016/j.epsr.2007.11.003>
30. Lauret P, Fock E, Randrianarivony RN, Manicom-Ramsamy J-F (2008) Bayesian neural network approach to short time load forecasting. *Energy Convers Manag* 49:1156–1166. <https://doi.org/10.1016/j.enconman.2007.09.009>
31. Ticknor JL (2013) A Bayesian regularized artificial neural network for stock market forecasting. *Expert Syst Appl* 40:5501–5506. <https://doi.org/10.1016/j.eswa.2013.04.013>

# Daily Rainfall Prediction Using Nonlinear Autoregressive Neural Network



Vuong Minh Le, Binh Thai Pham, Tien-Thinh Le, Hai-Bang Ly  
and Lu Minh Le

**Abstract** In this paper, a prediction model using Nonlinear Autoregressive Neural Networks with external variables (NARX) was proposed in order to forecast daily rainfall at Hoa Binh city, Vietnam. For this aim, eight-year time series of meteorological data were first collected, involving temperature, wind speed, relative humidity, solar radiation as input variables and daily rainfall as output variable. NARX-based daily rainfall prediction model was then constructed and validated using various criteria such as coefficient of correlation ( $R$ ), root mean squared error (RMSE) and mean absolute error (MAE). Results show a good statistical correlation between measured and predicted rainfall values, i.e.,  $R = 0.8846$ ,  $RMSE = 5.3793$  mm, and  $MAE = 3.0218$  mm. Therefore, it is reasonably stated that the developed model is promising for the forecast of daily rainfall.

**Keywords** Nonlinear Autoregressive Neural Networks · Rainfall · Artificial intelligence · Forecasting

---

V. M. Le · B. T. Pham (✉) · T.-T. Le  
NTT Hi-Tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City 700000, Vietnam  
e-mail: [binhpt@utt.edu.vn](mailto:binhpt@utt.edu.vn)

V. M. Le  
e-mail: [vuongminhle09@gmail.com](mailto:vuongminhle09@gmail.com)

T.-T. Le  
e-mail: [tienthinhle.vn@gmail.com](mailto:tienthinhle.vn@gmail.com)

H.-B. Ly  
University of Transport Technology, Trieu Khuc, Thanh Xuan, Hanoi 100000, Vietnam  
e-mail: [banglh@utt.edu.vn](mailto:banglh@utt.edu.vn)

L. M. Le  
Faculty of Engineering, Vietnam National University of Agriculture, Gia Lam, Hanoi 100000,  
Vietnam  
e-mail: [lmlu@vnua.edu.vn](mailto:lmlu@vnua.edu.vn)



## 1 Introduction

Rainfall is now one of the most important factors which have major effects on transport infrastructures in any urban area [1, 2]. Transport of goods and people can be vulnerable to climate change, especially to extreme daily rainfall. Indeed, high precipitation intensity may yield a lot of perturbations in the transport network [3, 4]. For example, it can cause flooding which damages transport infrastructures, delays traffics or worst destroys drainage systems, rail rails [5, 6]. In recent years, many climate change-related events have happened which badly influenced the transportation infrastructures, for example, the extreme precipitation in Baden-Württemberg, Germany in 2013 which has huge impacts on road infrastructures [7]. Being a state with one of the most developed economies in Europe, the extreme precipitation in Baden-Württemberg has damaged and destroyed roads, buildings and civilian houses. In the report analyzed after the flood, the precipitation had a record of 22.75 trillion liters over Germany in only 4 days. Today, rainfall forecasting is still a challenge and a very important task influencing the economic development of any urban area.

Many artificial intelligent (AI) approaches have been introduced in the past decades in order to be able to forecast several meteorological parameters associated to climate change such as solar radiation using artificial neural network [8], wind power using statistical method [9] or especially the precipitation [10]. In [11], Chang et al. have proposed a neural-fuzzy network-based model for forecasting the watershed rainfall on the upstream of the Tahan River in northern Taiwan, allowing to predict the rainfall with lead time of 1–2 h. Kisi and Shiri [12] have proposed a precipitation forecasting technique using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models at the Aegean region in the west of Turkey. An AI model using rule-based fuzzy inference system has been also used for rainfall forecasting in the Cairo airport station, Egypt [13]. The method allows predicting the rainfall up to 2–6 h ahead of the event, and the prediction results are quite satisfactory. Despite the fact that several AI techniques have been proposed and have produced very promising results on rainfall forecasting, this type of meteorological event still remains as one of the most stochastic nonlinear phenomenon in nature.

In this paper, a Nonlinear Autoregressive Neural Networks (NARX) model has been proposed in order to rigorously take into account the nonlinear behavior of precipitation time series to predict daily rainfall at Hoa Binh city, Viet Nam where is one of the most prone regions in Vietnam for floods [14] because of its typical climate and location. Hoa Binh is currently faced with several problems mainly expanding urbanization, especially for drainage systems. More recently, various master plans have been proposed to develop infrastructures for the city. The weather forecasting systems play now an essential role in order to prevent natural disasters like flood.

## 2 Material and Method

### 2.1 Method Used

In this work, Nonlinear Autoregressive Neural Networks (NARX) model was constructed and validated for predicting daily rainfall. The principal idea of this model is to predict future values of precipitation time series based on its history involving other meteorological variables such as temperature, wind speed, relative humidity and solar radiation [15]. Such external variables exhibit significant influences on the time series of interest and they are needed to be taken into account [16, 17]. NARX model has been broadly used for forecasting time series in various domains of science, for instance for energy consumption [18], for daily solar radiation [19], for air pollution [20] and for electric loads [21]. In this paper, NARX model was trained using Levenberg–Marquardt algorithm [22].

Three quality assessment criteria such as coefficient of correlation ( $R$ ), root mean squared error (RMSE) and mean absolute error (MAE) were introduced in order to quantify the performance of the developed NARX model.  $R$  was widely applied to exposure linear correlation between predicted and measured vector data [23].  $R$  yields a value between 0 and 1, where 1 is total correlation and 0 means zero-correlation. RMSE calculates the average square root of squared deviation between the predicted and actual values [24] while MAE measures the absolute normalized deviation between the predicted and measured data [25]. As an interpretation, low values of RMSE and MAE expose good correlation between two vectors such as predicted and measured data.  $R$ , RMSE and MAE were calculated using the following equations [26]:

$$R = \frac{\sum_{i=1}^N (t_i - \bar{t})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^N (t_i - \bar{t})^2 \sum_{i=1}^N (o_i - \bar{o})^2}}, \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (t_i - o_i)^2}, \quad (2)$$

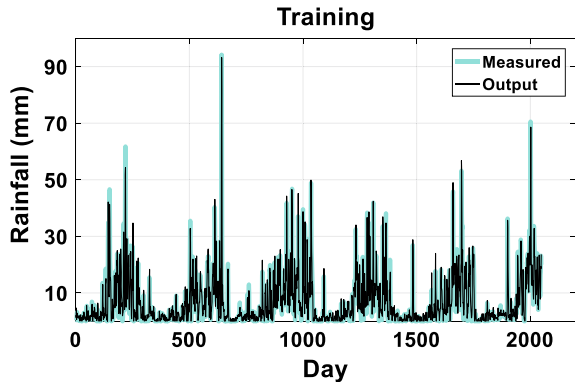
$$\text{MAE} = \frac{\sum_{i=1}^N |t_i - o_i|}{N}, \quad (3)$$

where  $N$  is the number of observations,  $t_i$  and  $\bar{t}$  are actual and mean actual values while  $o_i$  and  $\bar{o}$  are predicted and mean predicted values, respectively, ( $i = 1:N$ ).

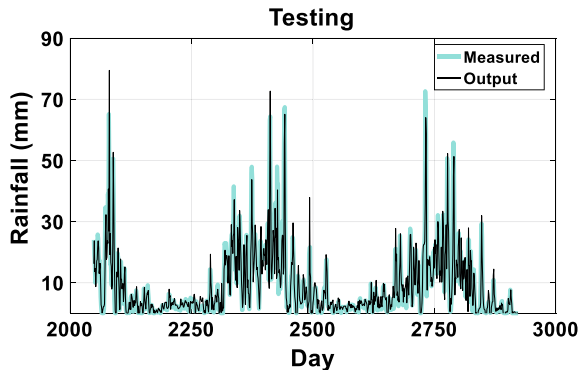
## 2.2 Data Used

In this study, daily meteorological data including precipitation intensity (mm), maximum temperature ( $^{\circ}\text{C}$ ), minimum temperature ( $^{\circ}\text{C}$ ), wind speed (km/h), relative humidity (–) and solar radiation ( $\text{mJ}/\text{m}^2$ ) were extracted from a meteorological station at Hoa Binh city, Vietnam, for a period of 8 years between 2006 and 2013. The rain gauge was located at Latitude 20.763 m and Longitude 105.312 m. Time series form of rainfall data is presented in Figs. 1 and 2, which exhibit large order of nonlinearity and fluctuation. In terms of statistical fluctuation, coefficient of variations of maximum temperature, minimum temperature, wind speed, relative humidity, solar radiation and rainfall intensity are 25.12, 29.04, 25.90, 13.76, 42.80, 134.51 (%), respectively. Data were split into two parts, first 70% values were used for training NARX model and last 30% values were used for testing the performance of NARX. Rainfall time series of the training part is shown in Fig. 1 while rainfall time series for the testing part is visualized in Fig. 2.

**Fig. 1** Time series for measured data and output of NARX for the training part



**Fig. 2** Time series for measured data and output of NARX for the testing part



### 3 Results and Discussion

Figure 1 shows the results of training part while Fig. 2 explores the testing rainfall time series using NARX model. Figure 3 presents error distribution and Fig. 4 exposes scatter regression plot between output of NARX model and measured values. Statistical parameters such as  $R$ , RMSE, MAE, mean, median and standard deviation of error including slope of linear fit were indicated in Table 1. As observed in Figs. 1 and 2, a strong correlation between measured and predicted data was obtained for both training and testing part. Regarding Table 1 for the training part,  $R = 0.9355$ , RMSE

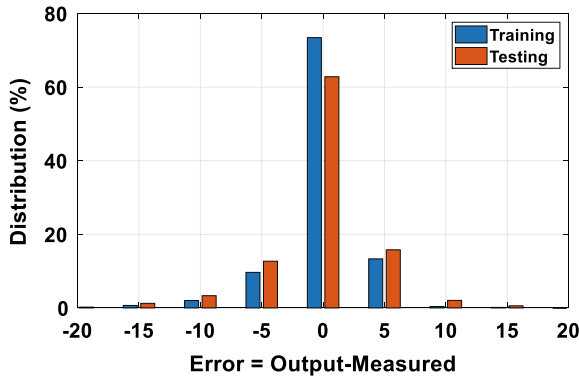


Fig. 3 Distribution of error between predicted and measured rainfall

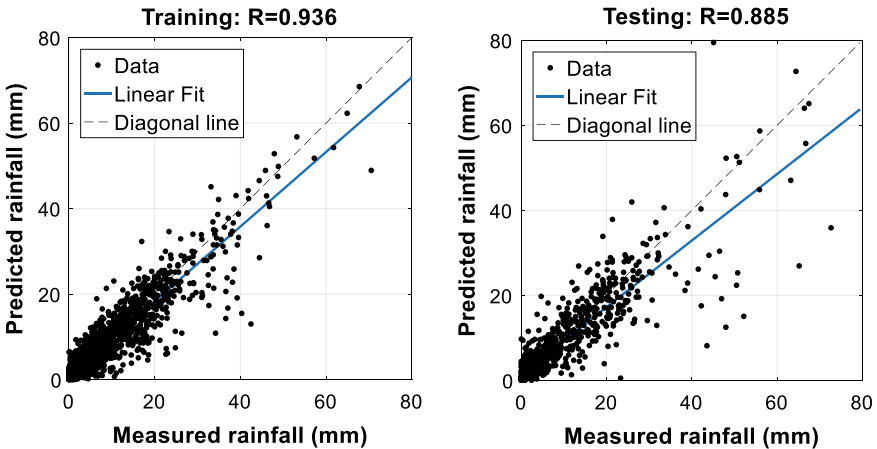


Fig. 4 Scatter plot of output versus measured rainfall values for: a the training part and b the testing part.  $R$  and linear fit are also indicated

**Table 1** Summary of statistical analysis

Statistical variable	Training data	Testing data
<i>R</i>	0.9355	0.8846
RMSE (mm)	3.3395	5.3793
MAE (mm)	2.0780	3.0218
Error mean (mm)	−0.0014	−0.3754
Error median (mm)	0.3016	0.2645
Error StD (mm)	3.3403	5.3692
Slope	0.87	0.78
Intercept	0.87	1.6
Slope angle (°)	41.02	37.95

= 3.3395, MAE = 2.0780, for the testing part,  $R = 0.8846$ , RMSE = 5.3793, MAE = 3.0218. These results confirmed the high performance of the proposed NARX model. The prediction model exhibited also a strong capability in order to predict nonlinear information inside data, including both rainy and no rainy days (Figs. 1 and 2).

Figure 3 presents the distribution of error—Error = (Output-Measured) (resolution of 5 mm). Both training and testing parts show a zero-high concentrated distribution (i.e., there were almost 70% of population at bin-zero). As highlighted in Table 1, error mean = −0.0014 and −0.3754, error median = 0.3016 and 0.2645, error standard deviation = 3.3403 and 5.3692, respectively, for the training and testing data.

Figure 4 presents measured and predicted rainfall in a scatter form including linear fit with two controlled parameters such as slope and intercept. As indicated in Table 1, slope = 0.87 and 0.78, intercept = 0.87 and 1.6, slope angle (°) = 41.02 and 37.95, respectively, for training and testing data.

In conclusion, NARX model was a potential candidate for forecasting rainfall in Hoa Binh city, Vietnam, as a contribution for the drainage infrastructure system. Such model was constructed and validated in accounting also other important meteorological variables such as temperature, wind speed, relative humidity and solar radiation. Such modeling methodology was relevant without dropping any significant statistical information within a context of atmospheric process. NARX was also proved as the most efficient model for forecasting daily rainfall in Nilgiris, India, when comparing with other algorithms such as forward back propagation neural network, cascade-forward back propagation neural network or distributed time delay neural network [27]. In other investigations, NARX technique exhibited also highly capability in predicting and exploring the relationship between meteorological variables [28, 29], especially for rainfall time series [30–33].

## 4 Conclusion

In this research, the NARX model was constructed for predicting daily rainfall at Hoa Binh city, Vietnam. Meteorological data including temperature, wind speed, relative humidity, solar radiation and rainfall were collected for an eight-year period. First 70% of data were prepared to train the NARX model while last 30% of data were used for testing the performance of the developed model.  $R$ , RMSE, MAE, mean and standard deviation of error including slope of linear fit were introduced in order to assess the quality of NARX model. The developed NARX model provided efficient information and could be used to improve the weather warning systems. In further researches, sensitivity analysis of external variables on the prediction of rainfall should be investigated in order to better understand meteorological processes.

**Conflict of Interest** The author declares no conflict of interest.

## References

1. Nemry F, Demirel H (2012) Impacts of climate change on transport: a focus on road and rail transport infrastructures. European Commission, Joint Research Centre (JRC), Institute for Prospective Technological Studies (IPTS)
2. Willems P, Olsson J, Arnbjerg-Nielsen K, Beecham S, Pathirana A, Gregersen IB, Madsen H, Nguyen V-T-V (2012) Impacts of climate change on rainfall extremes and urban drainage systems. IWA Publishing
3. Qing-Chang L, Zhong-Ren P, Junyi Z (2013) Critical transportation infrastructure identification and prioritization under flooding risks. In: Transportation research board 92nd annual meeting
4. Toda K (2007) Urban Flooding and Measures. *J Disaster Res* 2:143–152. <https://doi.org/10.20965/jdr.2007.p0143>
5. Suarez P, Anderson W, Mahal V, Lakshmanan TR (2005) Impacts of flooding and climate change on urban transportation: a system wide performance assessment of the Boston Metro Area. *Trans Res Part D: Transp Environ* 10:231–244
6. Pregolato M, Ford A, Wilkinson SM, Dawson RJ (2017) The impact of flooding on road transport: a depth-disruption function. *Transp Res Part D: Transp Environ* 55:67–81
7. Keller S, Atzl A (2014) Mapping natural hazard impacts on road infrastructure—the extreme precipitation in Baden-Württemberg, Germany, June 2013. *Int J Disaster Risk Sci* 5:227–241
8. Yadav AK, Chandel SS (2014) Solar radiation prediction using artificial neural network techniques: a review. *Renew Sustain Energy Rev* 33:772–781
9. Sideratos G, Hatzigiorgiou ND (2007) An advanced statistical method for wind power forecasting. *IEEE Trans Power Syst* 22:258–265
10. Wang W-C, Chau K-W, Cheng C-T, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J Hydrol* 374:294–306
11. Chang F-J, Chiang Y-M, Tsai M-J, Shieh M-C, Hsu K-L, Sorooshian S (2014) Watershed rainfall forecasting using neuro-fuzzy networks with the assimilation of multi-sensor information. *J Hydrol* 508:374–384
12. Kisi O, Shiri J (2011) Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. *Water Resour Manage* 25:3135–3152

13. Asklany SA, Elhelow K, Youssef IK, El-wahab MA (2011) Rainfall events prediction using rule-based fuzzy inference system. *Atmos Res* 101:228–236
14. Tien Bui D, Pradhan B, Lofman O, Revhau I, Dick ØB (2013) Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province. *Vietnam Nat Hazards* 66:707–730. <https://doi.org/10.1007/s11069-012-0510-0>
15. Billings SA (2013) *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. Wiley
16. Eccel E (2012) Estimating air humidity from temperature and precipitation measures for modelling applications. *Meteorological Applications*. 19:118–128. <https://doi.org/10.1002/met.258>
17. Lepore C, Allen JT, Tippet MK (2016) Relationships between hourly rainfall intensity and atmospheric variables over the contiguous United States. *J Climate* 29:3181–3197. <https://doi.org/10.1175/JCLI-D-15-0331.1>
18. Ruiz LGB, Cuéllar MP, Calvo-Flores MD, Jiménez MDCP (2016) An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings. *Energies*. 9:684. <https://doi.org/10.3390/en9090684>
19. Boussaada Z, Curea O, Remaci A, Camblong H, Mrabet Bellaaj N (2018) A Nonlinear Autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation. *Energies* 11:620. <https://doi.org/10.3390/en11030620>
20. Pisoni E, Farina M, Carnevale C, Piroddi L (2009) Forecasting peak air pollution levels using NARX models. *Eng Appl Artif Intell* 22:593–602. <https://doi.org/10.1016/j.engappai.2009.04.002>
21. Buitrago J, Asfour SS (2017) Short-term forecasting of electric loads using nonlinear autoregressive artificial neural networks with exogenous vector inputs. *Energies* 10:40. <https://doi.org/10.3390/en10010040>
22. Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 11:431–441. <https://doi.org/10.1137/0111030>
23. Le TT, Guillemot J, Soize C (2016) Stochastic continuum modeling of random interphases from atomistic simulations. Application to a polymer nanocomposite. *Comput Methods Appl Mech Eng* 303:430–449. <https://doi.org/10.1016/j.cma.2015.10.006>
24. Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res* 30:79–82. <https://doi.org/10.3354/cr030079>
25. Dao DV, Trinh SH, Ly H-B, Pham BT (2019) Prediction of compressive strength of geopolymers using entirely steel slag aggregates: novel hybrid artificial intelligence approaches. *Appl Sci* 9:1113. <https://doi.org/10.3390/app9061113>
26. Dao DV, Ly H-B, Trinh SH, Le T-T, Pham BT (2019) Artificial intelligence approaches for prediction of compressive strength of geopolymers. *Materials* 12:983. <https://doi.org/10.3390/ma12060983>
27. Devi SR, Arulmozhivarman P, Venkatesh C, Agarwal P (2016) Performance comparison of artificial neural network models for daily rainfall prediction. *Int J Autom Comput* 13:417–427. <https://doi.org/10.1007/s11633-016-0986-2>
28. Mahongo SB, Deo MC (2013) Using artificial neural networks to forecast monthly and seasonal sea surface temperature anomalies in the Western Indian Ocean. *The Int J Ocean Climate Syst* 4:133–150. <https://doi.org/10.1260/1759-3131.4.2.133>
29. Ouyang H-T (2017) Nonlinear autoregressive neural networks with external inputs for forecasting of typhoon inundation level. *Environ Monit Assess* 189:376. <https://doi.org/10.1007/s10661-017-6100-6>
30. Ang MRCO, Gonzalez RM, Castro PPM (2014) Multiple data fusion for rainfall estimation using a NARX-based recurrent neural network—the development of the REIINN model. *IOP Conf Ser: Earth Environ Sci* 17:012019. <https://doi.org/10.1088/1755-1315/17/1/012019>
31. Abou Rjeily Y, Abbas O, Sadek M, Shahroui I, Hage Chehade F (2017) Flood forecasting within urban drainage systems using NARX neural network. *Water Sci Technol* 76:2401–2412. <https://doi.org/10.2166/wst.2017.409>

32. Júnior J, Barreto GDA (2008) Multistep-ahead prediction of rainfall precipitation using the NARX network. Presented at the ESTSP'08, p 87
33. Noor HM, Ndzi D, Yang G, Safar NZM (2017) Rainfall-based river flow prediction using NARX in Malaysia. In: 2017 IEEE 13th international colloquium on signal processing its applications (CSPA), pp 67–72



# Demographic Analysis of World Population Using Various Predictive Algorithms



Ishan Bhatt and Kartik Ramaswamy

**Abstract** This project describes, in detail, the different regression models used in the analysis of population on various parameters like HDI, GDP, energy consumption and Internet usage. Population analysis can be useful in providing demographics of an area, help in targeted marketing and also help government to identify the areas of weakness of a country. Here, we have taken major countries and their various parameters like HDI, GDP, etc., and applied the regression models on them to find the most suitable one.

**Keywords** Population analysis · GDP · HDI · Internet usage · Regression analysis

## 1 Introduction

The United Nation collects variety of data from numerous countries generating a massive data set with millions of data points. The data sets, we have, consist of details from every country and contain information about health personnel, human development index, population growth, Internet usage, GDP, GDP in research and development, expenditure on health, expenditure in education, energy consumption, employment and education at all three levels. We are examining twenty countries, and these countries had the highest GDP in 2018. All these research fields are chosen in such a way that very useful information and insights can be derived from them, and also, we can predict the future growth of countries. All these data sets are very crucial to understand overall performance of a country and also to find out areas of improvement. These data sets are closely linked to people; that is, we can understand

---

I. Bhatt (✉) · K. Ramaswamy  
Computer Science and Engineering, Amity School of Engineering & Technology, Amity  
University, Noida, Uttar Pradesh 201303, India  
e-mail: [ishanbhatt.korba@gmail.com](mailto:ishanbhatt.korba@gmail.com)

K. Ramaswamy  
e-mail: [kartikrama007@gmail.com](mailto:kartikrama007@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication  
Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_23](https://doi.org/10.1007/978-981-15-2329-8_23)

overall growth of a country with the help of its citizens. Population analysis is very important in the field of data mining because change in population leads to the change of demography, social actors, consumer market, economy, etc. Few decades ago, population analysis was very complex process and was not simple to perform various calculations, but now with the help of advance computers the process became quite easy as compared to the past.

Population development is impacted by numerous variables that fall into the wide domains of statistic qualities, financial conditions, transportation framework, characteristic comfort, etc. It is seen that various factors have different effects on population change over time in rural, suburban and urban areas. It relies on the general pattern of population redistribution forms, nearby elements and geography. In general, an efficient analysis of population ought to think about the different variables and their impacts. The analysis ought to have the adaptability to recognize and join variables at a given point in time and space, not to hold to a particular set of drivers in all conditions. The discoveries have essential impact on population analysis. In the last few decades, population of the world is growing tremendously. Now it is necessary, and in fact it is very important to understand people. At present, estimated population is 2.5 billion and still counting. Understanding people in terms of caste, creed, religion, colour, country, sex, preferences and many more is difficult to anticipate.

## 2 Related Work Done

Human development index is a widely used tool by UNDP in their annual human development reports. The work of human development index was closely related to the famous Indian economist Amartya Sen. In earlier times, human development index was calculated by three basic parameters such as health, education and income of citizens of a country. From many years, countries are ranked with respect to human development index, and it can be seen in [1].

In [2], a close relationship between GDP and HDI is explained by author. In other words, we can say that HDI and GDP are coupled together and operated on similar grounds. But in [3], author explained in detail the basic difference between both GDP and HDI.

We can say that although they operate on similar grounds, they do have differences and basic and most important difference is in attitude. GDP views growth in terms of money, that is, import, export, government expenditure, but HDI is more complex tool that merges three fields together.

In [4], new-age HDI 2010 is mentioned in entire journal presented by UNDP in 2011. It includes poverty and several other areas which we should focus upon other than GDP and HDI. It also focuses on regional insights.

In [5], population growth is mentioned. A detailed study is conducted by publisher to understand population growth, and author also created an empirical view on population. In [5], heat maps and growth trends from eighteenth century are also shown. In [6], the generally adopted curve of population growth is explained with

brief explanation of environmental resistance. In [7], population growth is explained in economic point of view.

In [8], importance of Internet and daily uses of Internet are explained. In [9], interrelationship between human development and Internet usage is explained. We came to understand that Internet is now among the most used service, contributes a lot to human behaviours and is also closely linked to human development and country's economy.

In [10], authors explained in detail how literacy is closely linked to numerous aspects of a country. In [11], authors linked energy consumption to economic growth of a country. All the data sets are gathered from certified websites of UNDP from [12] to [13].

### 3 Methodology

The population is analysed on basis of HDI [3], GDP, Internet usage [7], literacy rate, employment, etc. We selected top 20 countries with highest GDP according to UN data [13]. We applied following regression models:

- **Simple Linear Regression:** It considers one independent variable and other one a dependent variable and finds a direct function which attempts to anticipate the reliance of the dependent variable qualities as an element of the independent factors. It is most normal and most generally utilized technique that enables us to comprehend and think about the connections between two things that can be said to be subject to one another.

$$Y = \beta_0 + \beta_1 X \tag{1}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \tag{2}$$

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \tag{3}$$

In given Eqs. (1), (2) and (3), the *betas* are said to be the coefficients. These coefficients are of need so as to make predictions with our model. To discover the parameters, we have to limit the total of squared errors.

In Fig. 1, the red dots are the genuine information and the blue line is linear model. The dark lines represent the errors between the anticipated and the genuine qualities. The blue line is hence the one that limits the whole of the squared length of the dark lines.

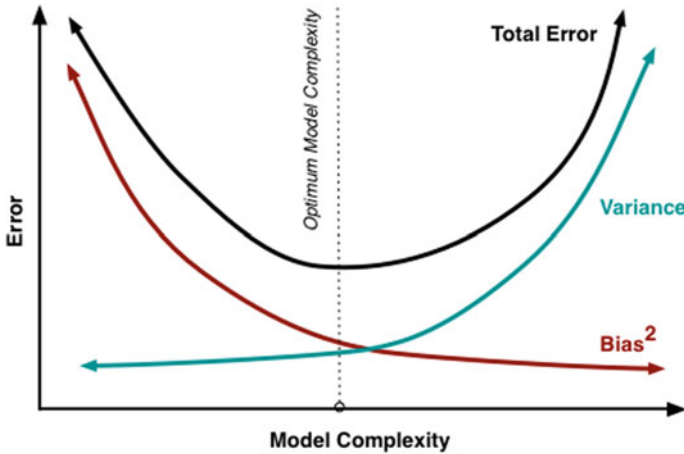


Fig. 1 Model complexity graph

- Polynomial Regression:** It is another method in which the relation or dependency between the variables that are dependent and independent is shown with help of curve or in simple language with help of degree of “ $n$ ”, where  $n$  is chosen as  $(y)-1$ .

It is visible that the linear graph is not able to show the patterns in the data. This shows that it is an instance of **under-fitting**. A statistical model or an algorithm is said to have under-fitting when it can't catch the hidden pattern of the information. Under-fitting demolishes the exactness of our model. In order to solve the issue of under-fitting, we are required to enhance the complexity of the model. To produce a higher-order condition, we can include powers of the initial features as new features. The linear model is modified to the following as shown in Eq. (4)

$$Y = \theta_0 + \theta_1 + \theta_2x^2 + \dots \tag{4}$$

With the help polynomial feature class which is a part of Scikit-learn, we will convert the original features into higher-degree terms.

Bias alludes to the mistakes because of the model's simplistic approach while fitting the information. A high bias implies that the model is unfit to catch the pattern in the information and this outcome in under-fitting.

Variance alludes to the mistakes because of the composite model endeavouring to fit the information. High variance implies that the model goes through the greater part of the information and this outcome in over-fitting the information.

We can see from the given below figure that with the increase in the complexity of the model, bias diminishes, variance rises and vice versa. In an ideal scenario, the model should have low variance and low bias. It is not possible to have both in practice. Hence, in order to attain a good, proper model that executes well on both the train and test data, a trade-off is made as shown in Fig. 2.

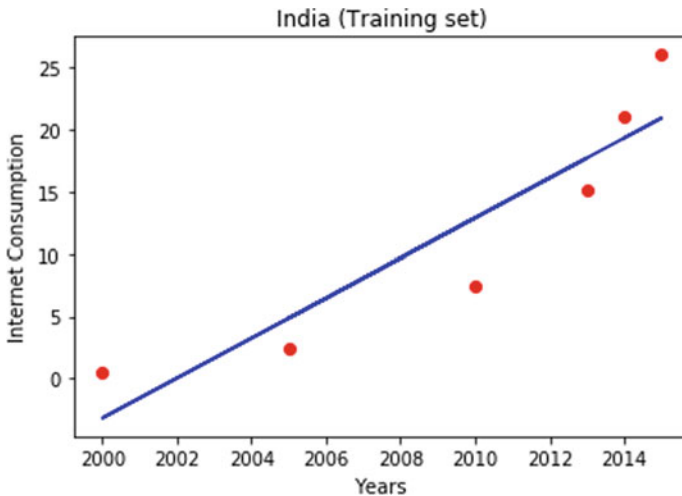


Fig. 2 Internet consumption of India to understand the model

- **Decision Tree:** It is a decision-making and predicting tool that uses a tree-like model of decisions. It contains root node and various other branches calculated using some algorithms.

It is a technique which is mainly used for classification and as regression model from decades. It provides the best predictive results in comparison with linear regression, polynomial regression and exponential regression. Decision tree model works by dividing the entire plot into several subparts, and then, these parts are considered as separate entities. This division of working area has to terminate when some criteria of the classifier are met or when the classes, it has divided into, are pure and cannot be divided further. There are 3 nodes in decision tree

- **ROOT NODE:** It has no incoming node and many outgoing nodes. It is the node where entire decision tree begins
- **LEAF NODE:** No outgoing nodes and many incoming nodes. It is where entire decision tree terminates
- **INTERNAL NODE:** Also known as intermediate nodes.
- **Random Forest Regression:** It builds many decision trees and merges them to create stable predictions. It is very complex to create; however, it can create highly useful and very accurate predictions. Random forest is the best-known regression technique as it is a flexible algorithm which gives good results in less time without much modification. The biggest problem with decision tree was that it over learns from the data and provides rigid solutions. Random forest, on the other hand, clubs many miss-fitting decision trees into one model creating a forest. It is also one of the most used algorithms because it is very easy in both classification and regression tasks. Coding for random forest is also very short, precise and simple.

The biggest advantage of random forest is that it performs both the works of classification and regression which are usually done separately in different machine learning algorithms. It is very easy to apply random forest on a raw data and get result in no time; however, in any other regression technique, first of all a coder has to simplify and classify the data and then has to apply any regression technique. Random forest not only skips the classification steps from coding but also creates more optimum result.

The best nature of random forest technique is that it is quite simple to quantify relative significance of each element, whereupon their decision trees are developed.

This should be possible utilizing Sklearn, it's incredible to which it gives capacities to measure relative significance of tree nodes, utilizing this package it is likewise simple to decrease in variations over every one of the trees. Sklearn package automatically computes scores of each feature and kill the result so that output will be between 0 and 1.0 means there is absolutely no importance and one mean that there is complete importance.

## 4 Implementation

Data mining is the process of discovering predictive information from the analysis of large databases. Following services offered by Python used were,

- NumPy: It is a tool which is used to perform mathematical functions. It consists of complex mathematical functions which help in performing functions with lesser lines of code.
- Pandas: It is a tool which is used to perform analysis and manipulation of data. It has various functions and data structures using which we can manipulate tables.
- PlotLy: It is used for creating visuals like graphs. It also provides statistical and analytical functions.
- Scikit-learn: Scikit-learn is the library in python which is just the tool which allows data scientist to tap into this and draw conclusions. It is told as the most used library. By the use of this library, programmers can not only build models but also do operations like scaling, normalizing, etc., using this package. They can draw conclusions about how the model works and the accuracy of the model.

## 5 Results

### 5.1 *Linear Regression*

See Fig. 3.

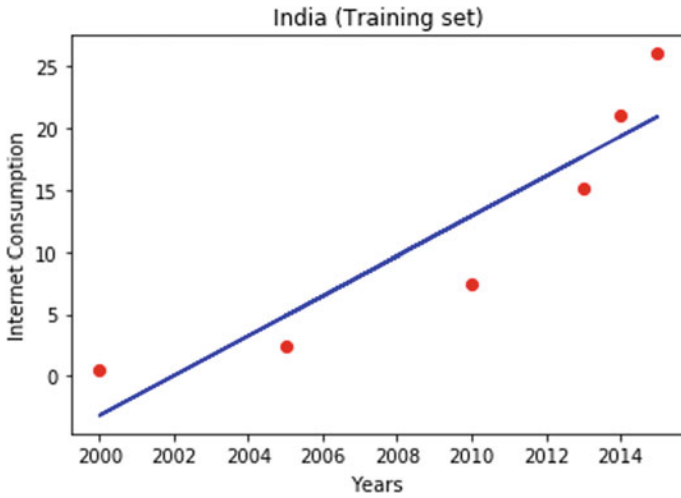


Fig. 3 Internet consumption of India (linear regression)

### 5.2 Polynomial Regression

See Fig. 4.

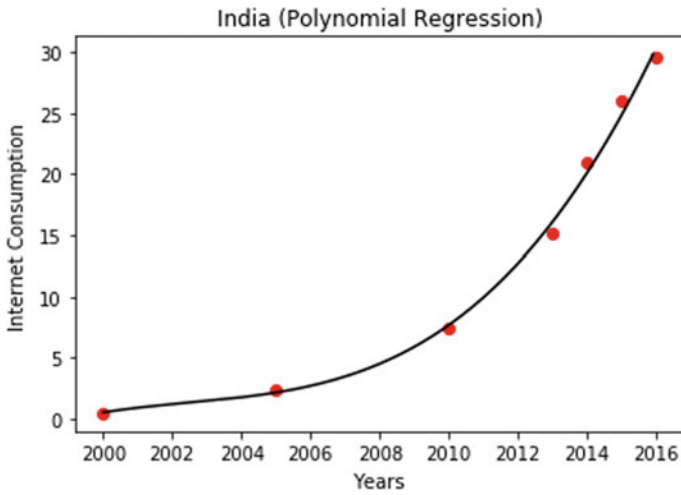
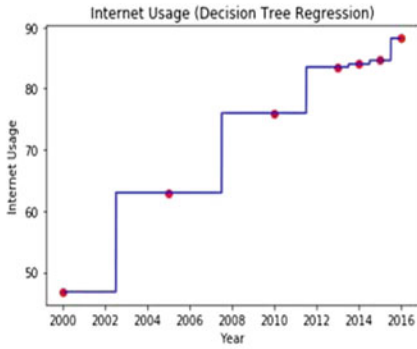


Fig. 4 Internet consumption of India (polynomial regression)



X	int64	(7, 1)	[[2000 2005]]
X_grid	float64	(1600, 1)	[[2000. ] [2000.01]]
X_test	float64	(1, 3)	[[ 84.6 5261. 9.4]]
X_train	float64	(5, 3)	[[ 76. 5431. 8.5] [ 63. 4762. 8. ]]
dataset	DataFrame	(7, 5)	Column names: T29, Internet U.
y	float64	(7,)	[46.8 63. 76. 83.5 84. 84.6 88.2]
y_pred	float64	(1,)	[76.]
y_test	int64	(1,)	[52238]
y_train	int64	(5,)	[58490 37674 65064 20841 62093]

Fig. 5 Internet consumption decision tree regression

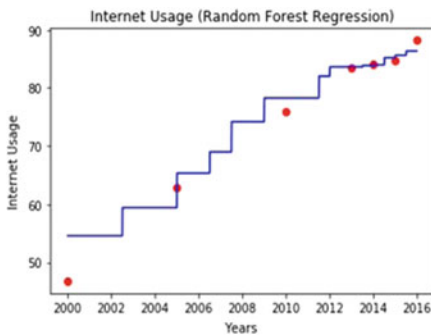
### 5.3 Decision Tree

The output for year 2011 was 76 and is less accurate as compared to random forest as shown in Fig. 5.

### 5.4 Random Forest

The output for year 2011 was 78.25 which was quite accurate, thus, indicating that random forest is much better than decision tree as shown in Fig. 6.

We can see from the above results that random forest gives the best result as compared to any other model. So we are going to examine every parameter using random forest regression model.



X	int64	(7, 1)	[[2000 2005]]
X_grid	float64	(1600, 1)	[[2000. ] [2000.01]]
X_test	float64	(1, 3)	[[ 84.6 5261. 9.4]]
X_train	float64	(5, 3)	[[ 76. 5431. 8.5] [ 63. 4762. 8. ]]
dataset	DataFrame	(7, 5)	Column names: T29, Internet U.
y	float64	(7,)	[46.8 63. 76. 83.5 84. 84.6 88.2]
y_pred	float64	(1,)	[78.25]
y_test	int64	(1,)	[52238]
y_train	int64	(5,)	[58490 37674 65064 20841 62093]

Fig. 6 Internet consumption in random forest regression



## 6 Conclusion and Comparison

We can conclude this study in two ways. First of all is technical where we can compare model's accuracy in prediction and we can also list out various advantages and disadvantages of all regression models. Second is fundamental where we can compare same various aspects such as energy consumption and GDP to find out overall growth of various countries. Fundamental conclusion can be used by lawmakers, policy makers, marketers and many other people to understand people and growth effectively.

Technical conclusions are as follows:

### (1) Linear regression

- a. Advantages:
  - i. It has low space complexity.
  - ii. It is easily understandable.
  - iii. Good interpretability.
- b. Disadvantages:
  - i. It is limited to linear relations only.
  - ii. It looks at the mean of the dependent variable.
  - iii. It is sensitive to outliers.
  - iv. The data used should be independent.

### (2) Polynomial regression

- a. Advantages:
  - i. It can fit a wide, big range of curvature.
  - ii. It provides many functions using which we can fit our data.
  - iii. It is quite flexible and easy to handle in cases where development of model is empirical in nature.
  - iv. It gives one of the best approximation relations between the variables.
- b. Disadvantages:
  - i. It is highly sensitive to outliers; i.e., even if there are one or two outliers, the result gets seriously affected.
  - ii. There are not many validation tools to check the data for outliers.

### (3) Decision tree regression

- a. Advantages:
  - i. It is quite easy to understand.
  - ii. It is easy for generating the rules.
  - iii. It can be used for categorical as well as numerical data inputs.
  - iv. It does not require lot of data preparation.
  - v. It helps in reducing the overall complexity of the problem.
  - vi. It is extremely fast in classifying the records that are not known.
- b. Disadvantages:
  - i. It might suffer from over-fitting.

**Table 1** India's growth in various fields

Field	Growth rate (%)
Energy consumption	6.925
Internet consumption	29.026
GDP	5.9904
Population growth	1.96
Human development index	1.968

- ii. In case, the decision tree is large, then pruning is necessary.
- iii. For a small change in set, it leads to selection of different attributes and decision trees.
- iv. It performs classification using rectangular partitioning, so it often fails to handle correlated data properly.
- v. There can be delay, and also, the technique may not reduce the amount of risk involved.

#### (4) Random forest regression

##### a. Advantages:

- i. Decision tree tends to overfit on data which reduces its prediction ability. Random forest on the other hand combines many such decision trees and reduces the problem of over-fitting.
- ii. Variance in random forest is also very less which means it is suitable for large amount of data.
- iii. Extremely flexible.
- iv. Highly accurate.
- v. It maintains accuracy even when some part of population is missing.
- vi. It requires no preparation in input of data.

##### b. Disadvantages:

- i. Complex coding.
- ii. Time-consuming.
- iii. It requires more computational resources than any other algorithm.

Fundamental conclusions are as follows (Table 1).

## References

1. En.wikipedia.org (2019) List of countries by human development index. [online] Available at [https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_Human\\_Development\\_Index](https://en.wikipedia.org/wiki/List_of_countries_by_Human_Development_Index). Accessed 20 Feb 2019
2. Khodabakhshi A (2011) Relationship between GDP and human development indices in India. SSRN Electron J
3. Metcalfe L (2019) Tracking development: GDP versus human development index. [online] NationMaster Blog. Available at <https://www.nationmaster.com/blog/?p=45>. Accessed 20 Feb 2019

4. Klugman J, Rodríguez F, Choi H (2011) The HDI 2010: new controversies, old critiques. *J Econ Inequality* 9(2):249–288
5. Roser M, Ortiz-Ospina E (2019) World Population Growth. [online] Our World in Data. Available at <https://ourworldindata.org/world-population-growth>. Accessed 20 Feb 2019
6. Archive.cnx.org (2019) Population growth curves. [online] Available at 732c4f286e0a@6/population-growth-curves. Accessed 20 Feb 2019
7. Cleland J (2013) World population growth; past, present and future. *Environ Resour Econ* 55(4):543–554
8. Kumar V (2019) Importance of Internet in our life—KLIENT SOLUTECH. [online] KLIENT SOLUTECH. Available at <http://www.klientsolutech.com/internet-inour-life/>. Accessed 20 Feb 2019
9. Pratama A, Al-Shaikh M (2012) Relation and growth of internet penetration rate with human development level from 2000 to 2010. *Commun IBIMA*:1–8
10. Coulombe S, Tremblay J (2006) Literacy and growth. *Topics Macroecon* 6(2)
11. Stern D, Cleveland C (2004) Energy and economic growth
12. Hdr.undp.org (2018) Human Development Data (19902017) | Human Development Reports. [online] Available at <http://hdr.undp.org/en/data>. Accessed 20 Feb 2019
13. Data.un.org (2019) UNdata. [online] Available at <http://data.un.org/>

# Attribute-Based Access Control Schemes in Cloud: Performance and Research Directions



S. Sabitha and M. S. Rajasree

**Abstract** Disclosure of sensitive data leads to identity theft and violation of privacy. The untrusted cloud service provider (CSP) may try to disclose/misuse the data. It is necessary to provide access control and security over the outsourced and shared data to hide it from the CSP and unauthorized users. Traditional access control schemes are prone to security threats in the cloud environment. Attribute-based access control schemes (ABAC) are well suited for the cloud environment. Attribute-based encryption (ABE) is a promising cryptographic solution to provide fine-grained access control over the shared data. It selectively shares the data among the users and hides data from the CSP and unauthorized users. It preserves the privacy of users and the security of data being shared. Users can decrypt the data only if their attributes are satisfied with the access policy associated in the ciphertext. This paper presents a comprehensive survey of the ABE schemes. Taxonomy, performance comparison, and applications of ABE schemes are dealt with. The taxonomy and performance comparison help the selection of the most suitable ABE scheme based on specific usage scenarios. Thus, the survey opens up very interesting avenues for further research in this area, which are also discussed.

**Keywords** Cloud computing · Attribute-based encryption · Access control · Data sharing · Access policy

## 1 Introduction

Cloud computing is gaining more attention due to its features such as multi-tenancy, virtualization, and elasticity. It attracts the IT business, organizations, and common people to outsource their data, application, and computation to the cloud and helps

---

S. Sabitha (✉)  
University of Kerala, Trivandrum, Kerala, India  
e-mail: [sabitha@cet.ac.in](mailto:sabitha@cet.ac.in)

M. S. Rajasree  
APJ Abdul Kalam Technological University, Trivandrum, Kerala, India  
e-mail: [rajasree40@gmail.com](mailto:rajasree40@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_24](https://doi.org/10.1007/978-981-15-2329-8_24)

235

them to reduce the initial investment cost. Cloud users may be competitors of the same business; they may try to get the data of their opponents to gain more benefits or to defend the opponents. Lack of security and access control are the main reasons behind this problem, which prevent users from widely adopting and utilizing the cloud without any fear.

The management of cloud computing is done by untrusted CSP. Due to this, organizations and IT business solution providers hesitate to outsource their data to the cloud. Data storage on cloud data center reduces the storage burden of user devices and improves data access convenience. But it restricts the user’s physical control over the data. Sensitive and confidential data in the cloud need to be protected from unauthorized access. So the users generally store their data in an encrypted form. Promising cryptographic access control schemes protect data from unauthorized access while keeping it confidential. These access control schemes enable the data owner to revoke users’ access permissions.

The ABE scheme is used to provide attribute-based access control as well as hiding data residing in the cloud [1]. The data owner can predetermine the recipient groups in this scheme [2, 3]. It is a fine-grained access control mechanism and better than the role-based access control scheme. Role-based access control is not that much suitable for fine-grained access control. In this scheme, permissions are associated with roles that are assigned to users. Roles may change depending on the situation and time. Setting up an initial role structure is difficult and inflexible in rapidly changing domains.

The paper is organized as follows. Section 2 describes the taxonomy of various attribute-based access control schemes. Performance analysis of ABE schemes is described in Sect. 3. Section 4 discusses the future research directions in ABE. Then, the last section concludes the survey.

## 2 Taxonomy of Attribute-Based Access Control Schemes

Attribute-based access control schemes are analyzed and classified based on the architecture and essential features required for ideal attribute-based encryption. Figure 1

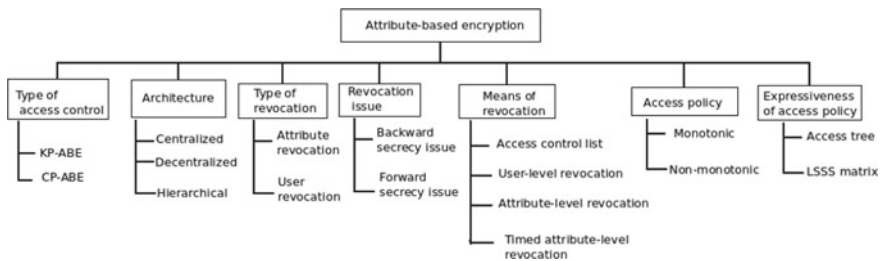


Fig. 1 Classifications of attribute-based access control

shows the taxonomy of attribute-based access control. The classification of ABE schemes is based on characteristics such as the type of access control, architecture, type of revocation, revocation security issue, means of revocation, access policy, and expressiveness of access policy. We have elaborated existing taxonomy by incorporating more additional characteristics such as means of revocation and expressiveness of access policy.

- (1) Type of access control: KP-ABE and CP-ABE are the general classifications of ABE schemes. (i) In KP-ABE, the user's secret key is determined by the access structure and ciphertext is associated with a set of attributes. Consequently, the users are allowed to decrypt the ciphertext only if their key's access structure is satisfied with the attributes associated in the ciphertext. (ii) In CP-ABE, access structure is associated with the ciphertext and the user's attributes decide their secret key. The user can decrypt the ciphertext only if their key's attributes satisfy the access structure associated with the ciphertext [2].
- (2) Architecture: ABE schemes are classified into three categories based on the architecture: Centralized, decentralized, and hierarchical. (i) Centralized ABE scheme has only one central attribute authority who is in charge of creating and distributing keys among users. (ii) Decentralized ABE scheme has multiple attribute authorities who are in charge of creating and distributing keys to the users [4]. Users can get their key from any attribute authority. It eliminates the risks involved in centralized ABE schemes. (iii) Hierarchical ABE scheme consists of root authority and domain authorities arranged in a hierarchical manner, which is responsible to distribute the keys to the users [5].
- (3) Type of revocation: Revocation prevents users from accessing data once their attributes are revoked. There are two types of revocation methods in an ABE scheme: attribute revocation and user revocation. (i) In attribute revocation, the user cannot access the data after their attributes get revoked. (ii) In the case of user revocation, the ABE scheme prevents the user from accessing data if he/she gets revoked.
- (4) Revocation security issue: ABE scheme suffers from two security issues: backward security issue and forward security issue. (i) Backward secrecy issue means that the user who newly joins the system and has sufficient attributes is able to decrypt the ciphertext which was encrypted and shared before he/she joined the system. (ii) Forward secrecy issue means that the user who revoked the attributes is able to decrypt any new ciphertext that needs the dropped attribute to decrypt [6].
- (5) Means of revocation: There are four means of revocation: access control list (ACL), user-level revocation, attribute-level revocation, and timed attribute-level revocation. (i) In the access control list, the authorized user details are maintained in the ACL. Only those users who are in the access control list are able to access the data. (ii) In user-level revocation, authorization of users

is dropped based on their identity. (iii) In attribute-level revocation, authorization of users is dropped based on their attributes. (iv) In timed attribute-level revocation, a time-based attribute is used to revoke the authorization of users [7].

- (6) Access policy: The access policy is represented using an access tree with attributes as leaves and threshold gates as internal nodes. The internal nodes are represented in two types: monotonic and non-monotonic. A monotonic access structure is constructed from “AND logic,” “OR logic,” and other threshold gates. But a non-monotonic access structure supports “NOT gate” also. However, monotonic policies do not support this feature. This problem is resolved by defining the negation of all individual attributes.
- (7) Expressiveness of access policy: Access policy expressiveness is of two types: access tree and LSSS matrix.

The taxonomy of ABE schemes based on the features supported in various literatures such as [2, 4, 5, 8–15] is shown in Fig. 2. It gives a clear understanding of various developments in ABE.



**Fig. 2** Taxonomy based on features of ABE

### 3 Performance Analysis of Attribute-Based Access Control Schemes

Table 1 compares the characteristics of existing attribute-based access control schemes, viz. access structure representation, type of access control, policy expressiveness, type of revocation, and means of revocation. It reveals the method used for user revocation. (1) Access structure representation: It can be a monotonic boolean function, conjunctive normal form (CNF), and disjunctive normal form (DNF). (2) Type of access control: KP-ABE and CP-ABE are the encryption schemes used for attribute-based access control. (3) Policy expressiveness: Access policy is expressed using either access tree or LSSS matrix. In access tree, leaf nodes represent attributes while internal nodes are threshold gates like AND, OR, etc. Linear secret sharing scheme (LSSS) is a matrix-like representation. Each row of matrix maps to an attribute. (4) Type of revocation: One of the major challenges in ABE scheme is to prevent the users from accessing the ciphertext even after their attributes and user permission gets revoked. User revocation and attribute revocation are types of revocations. (5) Means of revocation: Different means of revocation exists, namely user-level, attribute-level, timed attribute-level, access control list, and boolean function.

Table 2 describes the notations used for the following comparisons. Table 3 compares the performance in terms of computation time of existing ABAC schemes in the cloud. It reveals the overhead exerted by the scheme at the data owner's side as well as at the user's side due to the encryption and decryption process.

**Table 1** Comparison of characteristics of attribute-based access control schemes

Scheme	Access policy representation	Type of access control	Expressiveness of policy	Type of revocation	Means of revocation
RNS [16]	Monotonic boolean function	CP-ABE	LSSS matrix	User revocation	Attribute-level
BSW [2]	Monotonic boolean function	CP-ABE	Access tree	Key revocation	Timed attribute
YWRL [17]	Monotonic boolean function	CP-ABE	Access tree	User revocation	Access control list (ACL)
LYRL [18]	CNF boolean function	KP-ABE	Access tree	User revocation	Attribute-level
LYZRL [19]	CNF with wildcard	KP-ABE	Access tree	User revocation	Attribute-level

(continued)



**Table 1** (continued)

Scheme	Access policy representation	Type of access control	Expressiveness of policy	Type of revocation	Means of revocation
SW [20]	Monotonic boolean function	KP-ABE	Access tree	No	No
GPSW [21]	Monotonic boolean function	KP-ABE	Access tree	No	No
CN [22]	DNF	CP-ABE	Access tree	User revocation	Boolean function
GHW [23]	Monotonic boolean function	CP-ABE	LSSS matrix	No	No
W [1]	Monotonic boolean function	CP-ABE	LSSS matrix	No	No
HUR [6]	Monotonic boolean function	CP-ABE	Access tree	User revocation	Attribute-level
RSN [4]	Monotonic boolean function	CP-ABE	LSSS matrix	User revocation	Attribute-level
YJR [24]	Monotonic boolean function	CP-ABE	LSSS matrix	Attribute revocation	Attribute-level
ZNS [25]	Monotonic boolean function	CP-ABE	Access tree	No	Attribute-level

## 4 Research Directions

Research gaps with respect to attribute-based access control scheme are identified. Various research scopes and open problems still exist in the attribute-based access control scheme. Some of them are listed below:

1. Design an efficient attribute-based access control scheme with less communication and computation complexity while preserving forward and backward secrecy.
2. Efficiently manage dynamic updates of attributes and shared ciphertext.
3. Efficiently manage user and attribute revocation without redistributing the keys.

**Table 2** Description of notations used

Notations used	Purpose
$ G_1 ,  G_T $	Bit length of an element in group $G_1, G_T$ , respectively
$ Z_p $	Bit length of an element in $Z_p$
$r_1, r_2$	Set of attributes associated with ciphertext and secret key, respectively
$ msg $	Size of the message
$n$	Number of elements
$m$	Number of attributes in access structure
$u$	Size of an attribute universe
$N'$	Total number of possible statements of attributes
$l$	Number of attribute type in public domain (PUD)
$N$	Number of attribute authority (AA) in public domain (PUD)
$T_p$	Time taken for one pairing operation
$T_H$	Time to hash using function $H$
$n_u$	Number of users in the system
$E_0, E_T$	Exponentiation in group $G_0, G_T$ respectively

**Table 3** Performance comparison in terms of computation time of ABE schemes in cloud

Scheme	Encryption	Decryption
SW [20]	$r_1G_1 + 2G_T$	$r_1T_p + (r_1 + 1)G_T$
GPSW[21]	$r_1G_1 + 2G_T$	$r_1T_p + (n + 1)G_T$
CN [22]	$(n + 1)G_1 + 2G_T$	$(n + 1)T_p + (n + 1)G_T$
BSW [2]	$(2r_1 + 1)G_1 + 2G_T$	$2r_1T_p + (2r_1 + 2)G_T$
NYO [26]	$(2N' + 1)G_1 + 2G_T$	$(3n + 1)T_p + (3n + l)G_T$
W [1]	$(3r_1n + 1)G_1 + 2G_T$	$(l + n + n)T_p + (3r_1 - 1)G_1 + 3G_T$
KAKM [27]	$(n + 1)G_1 + 2G_T$	$2T_p + 2G_T$
YWRL [17]	$(u + 1) G_1  +  G_T $	$(u + 1)T_p + r_2G_l$
LYRL [18]	$O(r_1)$	$O(mT_p)$
RNS [16]	$O(r_1)$	$O(mT_p)$
HUR [6]	$(2m + 1) G_1  +  G_T $	$(2r_2 + 2)T_p + r_2k G_1  + \log m$
HN [28]	$O(r_1 + \log n_u)$	$O(r_2)$
YJR [24]	$O(r_1)$	$O(1)$
RSN [4]	$(3m + 1)E_0 + 2mE_T + T_p + (2l + 2)E_1 + 2tE_2 + T_H$	$2mT_p + T_H + O(mh)$

## 5 Conclusion

A comprehensive literature survey has been done on ABAC schemes. We have analyzed the ABE schemes based on the essential requirements of an efficient access control mechanism. The ABE schemes are classified into three major categories based on the architecture of the schemes. A taxonomy of available ABAC schemes is presented. Various ABE schemes are compared based on important features, performance, etc. Compared to the traditional access control schemes, ABAC schemes are more suitable for the cloud computing environment. Most of the schemes are able to ensure privacy, security, and access control for the data storage and sharing in the cloud. The comprehensive analysis of ABE schemes thus presented open up several interesting and challenging research directions. These future research directions are also discussed.

## References

1. Waters B (2011) Ciphertext-policy attribute-based encryption: an expressive, efficient, and provably secure realization. In: *Cryptography-PKC 2011*. Springer Berlin Heidelberg, pp 53, 70
2. Bethencourt J, Sahai A, Waters B (2007) Ciphertext-policy attribute-based encryption. In: *Proceedings of the IEEE symposium on security and privacy*, pp 321–334
3. Xu S, Yang G, Mu Y, Deng RH (2018) Secure fine-grained access control and data sharing for dynamic groups in cloud. *IEEE Trans Inf Forensics Secur*
4. Ruj S, Stojmenovic M, Nayak A (2014) Decentralized access control with anonymous authentication of data stored in clouds. *IEEE Trans Parallel Distrib Syst* 25(2):384–394
5. Deng H, Wu Q, Qin B, Domingo-Ferrer J, Zhang L, Liu J, Shi W (2014) Ciphertext-policy hierarchical attribute-based encryption with short ciphertexts. *Inf Sci* 275:370–384
6. Hur J (2013) Improving security and efficiency in attribute-based data sharing. *IEEE Trans Knowl Data Eng* 25(10)
7. Huang Q, Yang Y, Shen M (2017) Secure and efficient data collaboration with hierarchical attribute-based encryption in cloud computing. *Future Gener Comput Syst* 72:239, 249
8. Hur J (2013) Attribute-based secure data sharing with hidden policies in smart grid. *IEEE Trans Parallel Distrib Syst* 24(11):2171–2180
9. Teng W, Yang G, Xiang Y, Zhang T, Wang D (2015) Attributebased access control with constant-size ciphertext in cloud computing. *IEEE Trans Cloud Comput* 99:1–11
10. Chase M, Chow SSM (2009) Improving privacy and security in multi-authority attribute-based encryption. In: *Proceedings of the 16th ACM conference on computer and communications security*, pp 121–130
11. Muller S, Katzenbeisser S (2008) Distributed attribute-based encryption. *ICISC, Lect Notes Comput Sci Springer* 5461:20–36
12. Lewko AB, Waters B (2011) Decentralizing attribute-based encryption. In: *Proceedings of the Annual international conference on the theory and applications of cryptographic techniques (EUROCRYPT)*, pp 568–588
13. Ruj S, Stojmenovic M, Nayak A (2012) Privacy preserving access control with authentication for securing data in clouds. In: *12th IEEE/ACM international symposium on cluster, cloud and grid computing*, pp 556–563
14. Wan Z, Liu J, Deng RH (2012) HASBE: a hierarchical attribute-based solution for flexible and scalable access control in cloud computing. *IEEE Trans Inf Forensics Secur* 7(2):743–754

15. Wang G, Liu Q, Wu J (2010) Hierarchical attribute-based encryption for finegrained access control in cloud storage services. In: Proceedings of the 17th ACM conference on computer and communications security (CCS), pp 735–737
16. Ruj S, Nayak A, Stojmenovic I (2011) DACC: distributed access control in clouds. In: IEEE 10th international conference on trust, security and privacy in computing and communications (TrustCom), pp 91–98
17. Yu S, Wang C, Ren K, Lou W (2010) Attribute based data sharing with attribute revocation. In: Proceedings of the 5th ACM symposium on information, computer and communications security (ASIACCS), pp 261–270
18. Li M, Yu S, Ren K, Lou W (2010) Securing personal health records in cloud computing: patient-centric and fine-grained data access control in multi-owner settings. In: Proceedings of the sixth international ICST conference on security and privacy in communication networks (SecureComm), pp 89–106
19. Li M, Yu S, Yao Z, Kui R, Lou W (2013) Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. *IEEE Trans Parallel Distrib Syst* 24(1):131–143
20. Sahai A, Waters B (2005) Fuzzy identity-based encryption. In: International conference on advances in cryptology (EUROCRYPT), pp 457–473
21. Goyal V, Pandey O, Sahai A, Waters B (2006) Attribute-based encryption for fine-grained access control of encrypted data. In: ACM conference on computer and communication security, pp 89–98
22. Cheung L, Newport C (2007) Provably secure ciphertext-policy attribute-based encryption. In: ACM conference on computer and communication security, pp 456–465
23. Green M, Hohenberger S, Waters B (2011) Outsourcing the decryption of ABE ciphertexts. In: Proceedings of the USENIX security symposium, vol 3
24. Yang K, Jia X, Ren K (2012) Dac-macs: effective data access control for multiauthority cloud storage systems. In: IACR cryptology ePrint Archive, pp 419–429
25. Zhao F, Nishide T, Sakurai K (2011) Realizing fine-grained and flexible access control to outsourced data with attribute-based cryptosystems. In: Seventh international conference on information security practice and experience (ISPEC), pp 83–97
26. Nishide T, Yoneyama K, Ohta K (2008) Attribute-based encryption with partially hidden cryptor-specified access structures. In: ACNS, pp 111–129
27. Keita Emura, Atsuko Miyaji, Akito Nomura, Kazumasa Omote, Masakazu Soshi (2009) A ciphertext-policy attribute-based encryption scheme with constant ciphertext length. *Lect Notes Comput Sci* 5451:13–23
28. Hur J, Noh DK (2011) Attribute-based access control with efficient revocation in data outsourcing systems. *IEEE Trans Parallel Distrib Syst* 22(7):1214–1221

# Junctionless Gaussian Doped Negative Capacitance SOI Transistor: Investigation of Device Performance for Analog and Digital Applications



Hema Mehta  and Harsupreet Kaur 

**Abstract** In this work the performance of Junctionless Gaussian Doped Negative Capacitance Silicon-on-Insulator (JLGDNC SOI) transistor has been explored to examine the suitability of device for various analog and digital applications. The Negative Capacitance phenomenon of ferroelectric layer along with vertical Gaussian doped channel significantly enhances the performance of JL devices. To explore the electrical characteristics of JLGDNC SOI transistor TCAD models along with Landau-Khalatnikov equation which takes into account properties of Hafnium oxide based ferroelectric layer such as coercive field and remanent polarization have been used. It has been demonstrated that device exhibits substantially improved transfer characteristics, output characteristics, transconductance generation factor, output conductance and unity gain frequency.

**Keywords** Ferroelectric · Gaussian doped · Junctionless · Negative capacitance · Unity gain frequency

## 1 Introduction

Over the last few years, the scaling of CMOS devices has led to remarkable growth of microelectronic industry [1]. However, further reduction of device dimensions is limited due to various critical challenges such as complex fabrication procedure ‘for sharp S/D junctions, poor gate controllability, high leakage currents leading to significant power dissipation etc. [2–5]. Therefore, to overcome these issues, various new device designs such as Silicon-on-Insulator (SOI) technology, Junctionless (JL) device (without source-channel and drain-channel junctions) etc. have been researched extensively [5–7]. It has been well reported in literature that JLSOI

---

H. Mehta · H. Kaur (✉)

Department of Electronic Science, University of Delhi South Campus, Benito Juarez Road, New Delhi 110021, India

e-mail: [harsupreetkaur@gmail.com](mailto:harsupreetkaur@gmail.com)

H. Mehta

e-mail: [mehta\\_h14@yahoo.in](mailto:mehta_h14@yahoo.in)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106, [https://doi.org/10.1007/978-981-15-2329-8\\_25](https://doi.org/10.1007/978-981-15-2329-8_25)

245

devices exhibit superior performance at nanoscale dimensions [8, 9]. However, in bulk JLSOI devices it is difficult to completely deplete the channel, thereby, resulting in high leakage currents [9].

## 1.1 Literature Review

In order to suppress the leakage currents Mondal et al. [9] demonstrated JLSOI device with vertically Gaussian doping (GD) profile in channel and it was reported that device exhibits significantly high  $I_{ON}/I_{OFF}$  ratio along with improved gate control and steeper transfer characteristics [10, 11]. However, JLGDSOI device cannot achieve sub-60 mV/decade due to Boltzmann limit [12]. Hence, in order to further elevate the performance of JLGDSOI devices, in our previous work [13], we had incorporated gate insulator of ferroelectric (FE) material and the impact of Negative Capacitance (NC) phenomenon possessed by FE layer along with Gaussian doping profile parameters (projected range, straggle) was explored on Junctionless devices. It was demonstrated that due to FE layer and Gaussian doped channel, device exhibits voltage amplification capability along with sub-60 mV/dec subthreshold swing [14, 15], thereby, signifying its potential for ultra low power applications. Further, recently, Seo et al. has experimentally investigated FE FinFET for neuromorphic applications [16]. Also, Agarwal et al. has demonstrated negative differential resistance in NCFETs for analog applications [17]. In addition, Liang et al. has analysed the performance of NCFET based on a compact SPICE model [18]. Various other studies have been carried out [19, 20] focusing on NCFETs to examine the advantages of incorporating FE layer in nanoscale devices and circuits.

Therefore, in this work the suitability of Junctionless Gaussian Doped Negative Capacitance SOI transistor (JLGDNC SOI) for various digital and analog applications has been explored by obtaining several key figure of merits such as drain current ( $I_d$ ), transconductance ( $g_m$ ), transconductance generation factor ( $g_m/I_d$ ), output conductance ( $g_d$ ) and unity gain frequency ( $f_T$ ). The characteristics have been obtained for a range of doping profile parameters (projected range, straggle) and bias voltages by using Landau-Khalatnikov equation [21] with TCAD models [22]. Doped Hafnium oxide is considered as FE material in this analysis since these materials exhibit high coercive field ( $E_c$ ), low remanent polarization ( $P_r$ ) along with high scalability [23, 24]. It has been demonstrated that JLGDNC SOI exhibits substantially elevated performance in terms of steep transfer characteristics, high  $g_m$  and  $g_m/I_d$  values, high saturation current values and  $f_T$ , thereby, indicating the potential of device for a variety of electronic circuit applications.

## 2 Device Structure and Simulation Approach

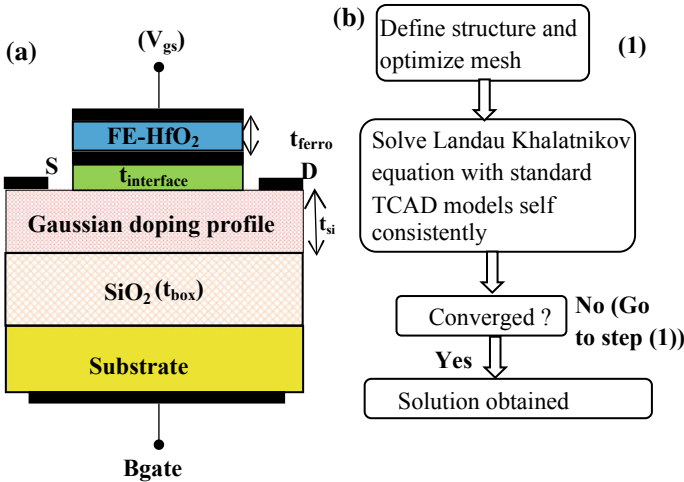
The schematic of Junctionless Gaussian Doped Negative Capacitance SOI Transistor (JLGDNC-SOI) is shown in Fig. 1a where,  $t_{ferro}$  is thickness of Hafnium oxide based ferroelectric layer,  $t_{interface}$  denotes thickness of interfacial layer between FE layer and silicon channel and  $t_{si}$  denotes silicon channel thickness. The Gaussian profile distribution in channel is expressed as [9–11]:

$$N_d(x) = N_{pk} \exp\left(-\left(\frac{x - R_p}{\sqrt{2}\sigma}\right)^2\right) \tag{1}$$

where,  $N_d(x)$  is the doping concentration in channel in vertical direction,  $N_{pk}$  denotes peak doping concentration,  $R_p$  denotes projected range and  $\sigma$  denotes straggle. The electrical characteristics of JLGDNC-SOI for different values of  $R_p$ ,  $\sigma$  and bias voltages is obtained by following simulation method shown in Fig. 1b. The device parameters have been obtained using TCAD models [17] *consrh*, *auger*, *bgn*, *conmob*, *fldmob* and *prpmob* along with Landau Khalatnikov equation which is expressed as [16]:

$$V_{ferro} = 2 t_{ferro} Q_g (a + 2bQ_g^2) \tag{2}$$

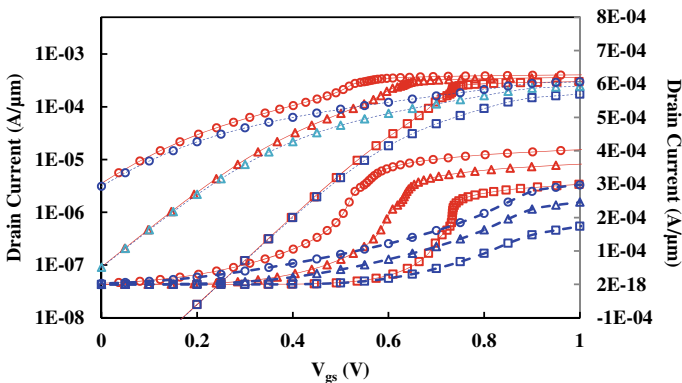
where,  $V_{ferro}$  is voltage drop across FE-HfO<sub>2</sub> layer,  $Q_g$  is the total charge density,  $a = 2a_0 t_{ferro}$  and  $b = 4b_0 t_{ferro}$ ,  $a_0$  and  $b_0$  are Landau parameters of FE-HfO<sub>2</sub> layer [18–20].



**Fig. 1** a Schematic OF Junctionless Gaussian doped negative capacitance SOI transistor (JLGDNC-SOI). b Numerical simulation method

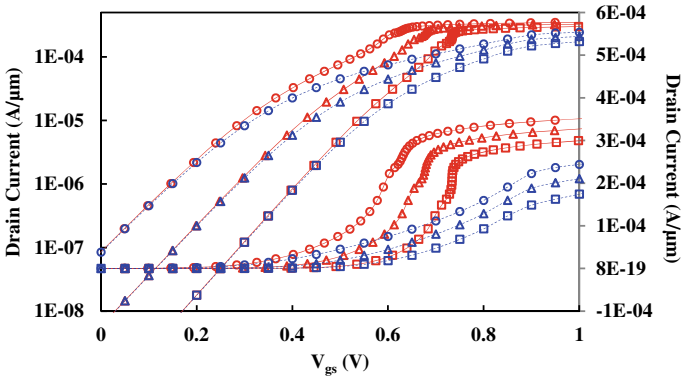
### 3 Results and Discussion

The characteristics of JLGDNC SOI [13] have been obtained for device parameters peak doping ( $N_{pk}$ ) =  $10^{19}$  cm $^{-3}$ ,  $t_{interface}$  = 1 nm,  $t_{si}$  = 10 nm, dielectric constant of interfacial layer = 3.9, metal workfunction = 5.1 eV, gate length = 20 nm, thickness of buried oxide layer = 20 nm and drain bias voltage = 0.1 V unless otherwise specified. The impact of doping profile parameters on various figure of merits have been explored for JLGDNC SOI transistor and the results have been compared with JLGDSOI. Figure 2 shows transfer characteristics of JLGDNC SOI and JLGDSOI in both log and linear scale (secondary axis) for different values of straggle. It can be noted from Fig. 2 that curves for JLGDNC SOI are steeper than JLGDSOI due to NC phenomenon which results in improved switching characteristics. Also, it can be observed that OFF current increases significantly with increase in straggle since it is difficult to deplete the channel with higher  $\sigma$  due to more number of carriers. Further, it can be noted that the peak observed in  $I_d-V_{gs}$  characteristics of JLGDNC SOI due to NC effect decreases with increase in  $\sigma$ . This is attributed to suppression of impact of NC effect on device due to increase in average number of carriers. Also, it can be noted from linear scale that ON current for JLGDNC SOI increases about 1.5 times as compared to JLGDSOI, thereby, signifying that JLGDNC SOI exhibits superior performance. In addition, ON current is maximum for  $\sigma = 6$  nm, since the average number of carriers will be high as compared to  $\sigma = 3.5$  nm. However, OFF current increases significantly for  $\sigma = 6$  nm and NC effect suppresses, thereby, indicating that an optimum value of  $\sigma$  is required to achieve maximum enhancement in device characteristics. In Fig. 3 impact of projected range on transfer characteristics of JLGDNC SOI and JLGDSOI have been shown on both log and linear scale. It can be noted that characteristics are steeper for JLGDNC SOI with  $R_p = 0$  nm i.e. when peak doping is at surface and decreases gradually in vertical direction. This is due to the fact that for device with  $R_p = 0$  nm, channel is easily depleted as number of



**Fig. 2** Drain Current with gate voltage for different values of straggle ( $\sigma$ ),  $t_{ferro} = 8$  nm,  $R_p = 0$  nm,  $N_{pk} = 10^{19}$  cm $^{-3}$ ,  $\square \sigma = 3$  nm,  $\triangle \sigma = 4.5$  nm,  $\circ \sigma = 6$  nm

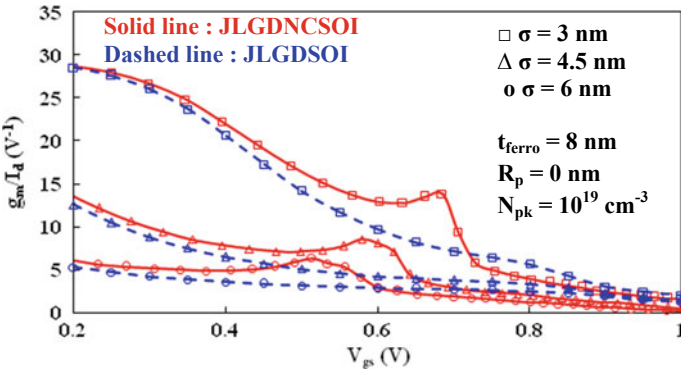




**Fig. 3** Drain Current with gate voltage for different values of projected range ( $R_p$ ),  $t_{\text{ferro}} = 8 \text{ nm}$ ,  $\sigma = 3 \text{ nm}$ ,  $N_{\text{pk}} = 10^{19} \text{ cm}^{-3}$ ,  $\square - R_p = 0 \text{ nm}$ ,  $\triangle - R_p = 1 \text{ nm}$ ,  $\circ - R_p = 2 \text{ nm}$

carriers decrease significantly in vertical direction which aids in depletion of channel and consequently results in enhancement of impact of NC phenomenon on device. Also, the peak observed decreases with increase in  $R_p$  indicating that the voltage amplification capability of device decreases as projected range increases. Further, it is clearly evident that ON current for JLGDNCSOI is significantly higher than JLGDSOI for all values of projected range and maximum current is obtained for JLGDNCSOI with  $R_p = 2 \text{ nm}$ . But for higher values of  $R_p$  leakage current increases and NC effect decreases. Therefore, it is of crucial importance to optimize doping profile parameters to achieve maximum capacitance matching, thereby, resulting in elevated device performance.

Figure 4 shows variation of transconductance generation factor ( $g_m/I_d$ ) with  $V_{\text{gs}}$  respectively for different values of straggle for  $V_{\text{ds}} = 0.5 \text{ V}$ . The peak observed in  $I_d - V_{\text{gs}}$  characteristics translates into a peak in  $g_m/I_d$  characteristics of JLGDNCSOI FET



**Fig. 4** Transconductance generation factor ( $g_m/I_d$ ) with gate voltage for different values of straggle ( $\sigma$ )

as can be clearly seen in Fig. 4. It can be observed that  $g_m/I_d$  values for JLGDNCSOI are much higher than JLGDSOI FET. However, the magnitude of peak observed decreases with increase in straggle due to suppression of NC effect as was also observed in Fig. 2.

In Fig. 5 impact of projected range has been demonstrated on  $g_m/I_d$ . It can be noted that transconductance efficiency for JLGDNCSOI is significantly high in comparison to JLGDSOI FET for all values of projected range and maximum value is observed for JLGDNCSOI with  $R_p = 0$  nm. Also, the magnitude of peak observed in  $g_m/I_d$  decreases with increase in  $R_p$  as was also observed in Fig. 3. Figure 6a, b shows output characteristics of JLGDNCSOI and JLGDSOI FETs at different gate bias voltages. It can be seen from Fig. 6a that saturation current of JLGDNCSOI FET is significantly higher as compared to JLGDSOI FET. Also, saturation current increases with increase in applied gate bias voltage for both JLGDNCSOI and JLGDSOI FET,

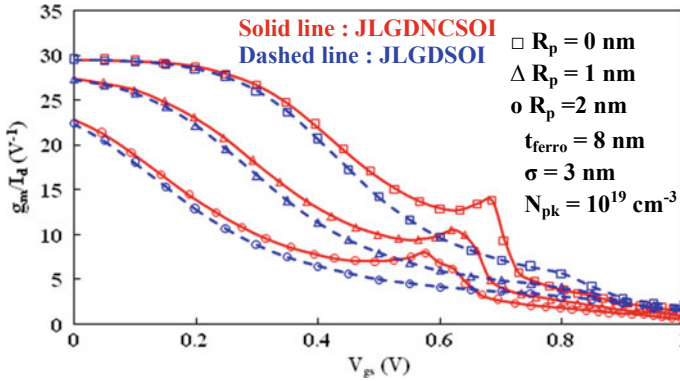


Fig. 5 Transconductance generation factor ( $g_m/I_d$ ) with gate voltage for different values of projected range ( $R_p$ )

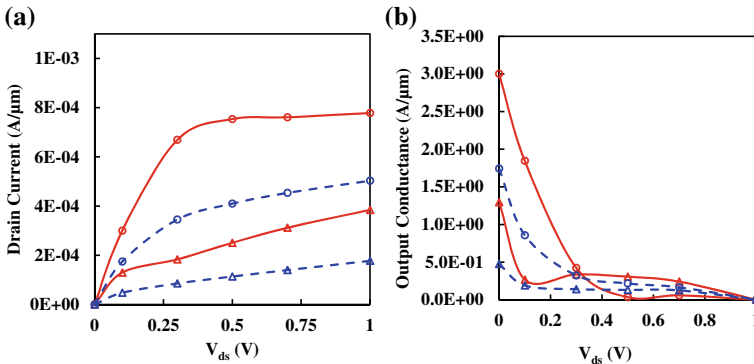
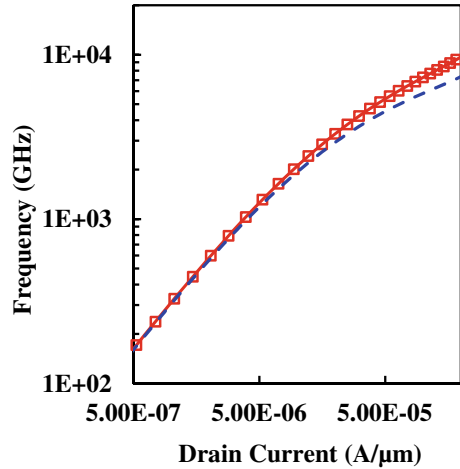


Fig. 6 a Drain current. b Output conductance with drain voltage for different values of gate bias voltage.  $t_{ferro} = 8$  nm,  $N_{pk} = 10^{19}$   $cm^{-3}$ ,  $\sigma = 3$  nm,  $R_p = 0$  nm

**Fig. 7** Unity gain frequency with drain current for both JLGDNCSOI and JLGDSOI devices



but increase in current for JLGDNCSOI is significantly high, thereby, indicating superior performance of JLGDNCSOI FET for a wide range of bias voltages and device parameters. In addition, it can be noted from Fig. 6b that output conductance for JLGDNCSOI is also high in comparison to JLGDSOI FET and the values increase further with increase in gate voltage. Figure 7 shows variation of unity gain frequency with drain current for both JLGDNCSOI FET and JLGDSOI FET and it can be noted that unity gain frequency for JLGDNCSOI FET increases due to increase in  $g_m$  values for the device. Hence, it can be observed that JLGDNCSOI exhibits substantial improvement in key figure of merits such as switching characteristics, transconductance efficiency, saturation current and output conductance and is suitable for future CMOS analog circuit applications.

## 4 Conclusion

In the present work performance of Junctionless Gaussian Doped Negative Capacitance Silicon-on-Insulator (JLGDNCSOI) transistor has been investigated for a wide range of doping profile parameters and bias voltages. The study primarily focuses on evaluation of analog parameters such as drain current characteristics, transconductance generation factor, output conductance and unity gain frequency. The electrical parameters have been obtained by self consistently solving Landau-Khalatnikov equation with TCAD models. The gate stack consists of hafnium oxide based ferroelectric layer along with an interfacial layer of silicon dioxide. It has been demonstrated that incorporation of Gaussian doping profile and ferroelectric layer significantly elevates the performance of JLGDNCSOI FET in terms of various electrical parameters. Also, it is crucial to obtain optimum values of doping profile parameters i.e. straggle, projected range to obtain enhanced impact of NC phenomenon on

device performance. The results obtained clearly indicate suitability of JLGDNCSOI for various analog and digital circuit applications.

## References

1. Borkar S (1999) Design challenges of technology scaling. *IEEE Micro* 19(4):23–29
2. Cavin RK, Zhirmov VV, Hutchby JA, Bourianoff GI (2005) Energy barriers, demons and minimum energy operation of electronic devices. *Fluctuations Noise Lett* 5(4):C29–C38
3. Kuhn KJ (2012) Considerations for ultimate CMOS scaling. *IEEE Trans Electron Devices* 59(7):1813–1828
4. Frank DJ, Dennard RH, Nowak E, Solomon PM, Taur Y, Wong H-S (2001) Device scaling limits of Si MOSFETs and their application dependencies. *Proc IEEE* 89:259–288
5. Colinge J-P, Lee C-W, Afzalain A, Dehdashti Akhavan N, Yan R, Ferain I, Razavi P, O'Neill, Blake A, White M, Kelleher A-M, McCarthy B, Murphy R (2010) Nanowire transistors without junctions. *Nat Nanotechnol* 5(3):225–229
6. Su KW, Kuo JB (1997) Analysis of current conduction in short-channel accumulation-mode SOI PMOS devices. *IEEE Trans Electron Devices* 44(5):832–840
7. Suzuki K, Tanaka T, Tosaka Y (1993) Scaling theory for double-gate SOI MOSFET's. *IEEE Trans Electron Devices* 40(12):2326–2329
8. Raksharam, Dutta AK (2017) A unified analytical drain current model for double-gate junctionless field-effect transistors including short channel effects. *Solid-State Electron* 130:33–40
9. Mondal P, Ghosh B, Bal P (2013) Planar junctionless transistor with non-uniform channel doping. *Appl Phys Lett* 102:133505
10. Zhang G, Shao Z, Zhou K (2008) Threshold voltage model of short-channel FD-SOI MOSFETs with vertical Gaussian profile. *IEEE Trans Electron Devices* 55(3):803–809
11. Dubey S, Tiwari P, Jit S (2010) A two-dimensional model for the potential distribution and threshold voltage of short-channel double-gate metal-oxide-semiconductor field-effect transistors with a vertical Gaussian-like doping profile. *J Appl Phys* 108:034518
12. Salahuddin S, Datta S (2008) Use of negative capacitance to provide voltage amplification for low power nanoscale devices. *Nano Lett* 8(2):405–410
13. Mehta H, Kaur H (2018) Impact of ferroelectric HfO<sub>2</sub> and non-uniform doping on Nanoscale planar SOI junctionless transistor. In: *Proceedings IEEE international conference on computer, power and communication technologies GUCON-2018*, pp 940–943, India
14. Wang C, Ye Y, Cao Y (2012) The potential of Fe-FET for robust design under variations: a compact modeling study. *Microelectron J* 43:898–903
15. Ko E, Lee JW, Shin C (2017) Negative capacitance FinFET With Sub-20-mV/decade subthreshold slope and minimal hysteresis of 0.48 V. *IEEE Electron Device Letters* 38(4):418–421
16. Seo M, Kang MH, Jeon SB, Bae H, Hur J, Jang BC, Yun S, Cho S, Kim WK, Kim MS, Hwang KM, Hong S, Choi SY, Choi YK (2018) First demonstration of a logic-process compatible junctionless ferroelectric FinFET synapse for neuromorphic applications. *IEEE Electron Device Lett* 39(9):1445–1448
17. Agarwal H, Kushwaha P, Duarte JP, Lin YK, Sachid AB, Kao MY, Chang HL, Salahuddin S, Hu C (2018) Engineering negative differential resistance in NCFETs for analog applications. *IEEE Trans Electron Devices* 65(5):2033–2039
18. Liang Y, Li X, Gupta SK, Datta S, Narayanan V (2018) Analysis of DIBL effect and negative resistance performance for NCFET based on a compact SPICE model. *IEEE Trans Electron Devices* 65(12):5525–5529
19. Oh S, Song J, Yoo IK, Hwang H (2019) Improved endurance of HfO<sub>2</sub>-based metal-ferroelectric-insulator-silicon structure by high-pressure hydrogen annealing. *IEEE Electron Dev Lett* 40(7):1092–1095

20. Mehta H, Kaur H (2019) Study on impact of parasitic capacitance on performance of graded channel negative capacitance SOI FET at high temperature. *IEEE Trans Electron Devices* 66(7):2904–2909
21. Kittel C (1976) *Solid state physics*. Wiley Eastern Limited, New York
22. Silvaco International (2014) *Atlas user's manual*. Silvaco International
23. Muller J, Polakowski P, Mueller S, Mikolajick T (2015) Ferroelectric hafnium oxide based materials and devices: assessment of current status and future prospects. *ECS J Solid State Sci Technol* 4(5):N30–N35
24. Böscke TS, Müller J, Bräuhäus D, Schröder U, Böttger U (2011) Ferroelectricity in hafnium oxide thin films. *Appl Phys Lett* 99(10):102903

# Development of a Down-Converter for a Software-Defined Radio Environment



Bruce Alistair van Niekerk, Tariq Ekeramodien Kahn and Vipin Balyan

**Abstract** The development of a down-converter is investigated that can be used for a software-defined radio (SDR) environment. The paper also compares down-conversion receiver architectures, and an alternative front-end receiver architecture is proposed. The proposed front-end receiver will extend the range of the typical SDR to the X band which will help reduce the complexity and cost of higher-frequency SDRs.

**Keywords** Front-end · Down-conversion · Software-defined radio

## 1 Introduction

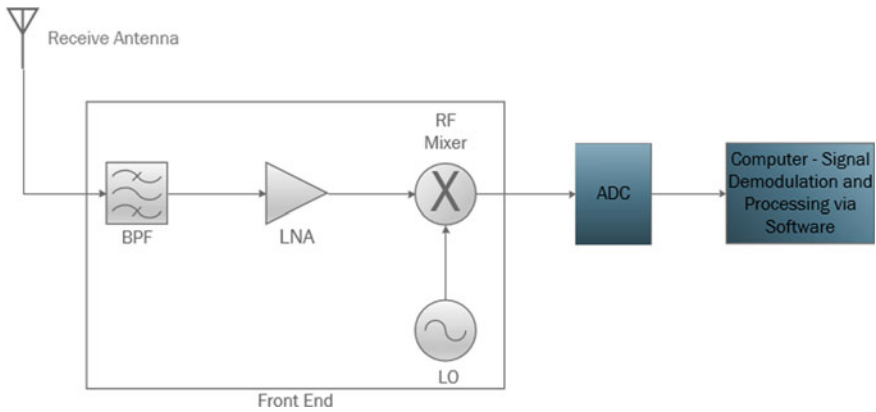
In 1991, Joseph Mitola conceived the name software-defined radio (SDR). The intention was to have most of the receiver/transmitter sections implemented and processed through software into a binary format, through a converter for receivers termed analog-to-digital (ADC) converter and for transmitters termed digital-to-analog (DAC) converter. The ideal configuration is to have the SDR placed directly after the antenna. The limitations of the ADC prevent this, and consequently, the desired receive signal has to be frequency shifted to lower practical frequency standard termed the intermediate frequency (IF) where the lower practical frequency is transferred to the ADC and digitized to a baseband signal for processing as in [1]. The receiver front-end is the portion of the receiver that translates the desired input signal to the lower in frequency IF signal. The receiver front-end as in Fig. 1 is also known as the analog-to-baseband part of a receiver [2]. The general components of this subsystem as in Fig. 1 that follow directly after the antenna [3] are:

- Band-pass filter (BPF): component that limits the input signal to what is known as a band-limited signal. It also isolates the mixer from the antenna and prevents the mixer from radiating through to the antenna.

---

B. A. van Niekerk (✉) · T. E. Kahn · V. Balyan  
Department of Electrical Engineering,  
Cape Peninsula University of Technology, Cape Town 7535, South Africa  
e-mail: [bruce8424@gmail.com](mailto:bruce8424@gmail.com)

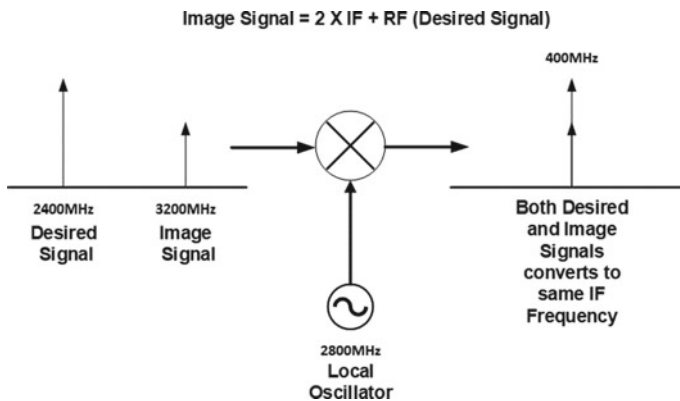
© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_26](https://doi.org/10.1007/978-981-15-2329-8_26)



**Fig. 1** General front-end

- Low-noise amplifier (LNA): component that amplifies the band-limited signal by adding minimal noise.
- Local oscillator (LO): component which is used as a reference signal.
- Radio frequency (RF) mixer: component that mixes the band-limited signal with the reference signal to a lower usable frequency that the ADC can sample and correctly digitize for processing.

There are many receive architectures that can provide the IF signal to the ADC. The concern with most of the receiver architectures is that there is a signal at a specific frequency that will always interfere with the desired frequency after down-conversion is done. This specific frequency is known as the image frequency as seen in Fig. 2 and cannot be filtered out by any type of filter once it is down-converted with the desired frequency [4]. The receiver architectures that suppress this image frequency are the:



**Fig. 2** Image problem

- Homodyne/direct-conversion/zero-IF receiver [5].
- Superheterodyne/heterodyne receiver [6].
- Low-IF receiver [7].
- Hartley image-reject receiver [6].
- Weaver image-reject receiver [6].

All these different receiver architectures have their methods of dealing with the image frequency, but only one receive architecture is best suited for the mobile SDR environment.

## 2 Receiver Architectures

### 2.1 Homodyne/Direct-Conversion/Zero-IF Receiver

This category of frequency translation was first implemented in 1924 in a single vacuum tube receiver, then again in 1947 for a carrier-based telephony system. It was only until 1980 that this receiver architecture was used in a radio paging receiver. This architecture is distinctive of having the LO in sync with the desired RF to create a zero-frequency baseband signal [5]. This process eliminates the need for a BPF filter before and after the mixer as seen in Fig. 3. The image frequency is mixed down to zero. Consequently, the image and desired RF can be distinguished by their amplitude [8] (Fig. 4).

Also, in Fig. 3, the quadrature of the LO is used. This is to detect frequency and phase modulation schemes [6] and to prevent negative-frequency half-channels

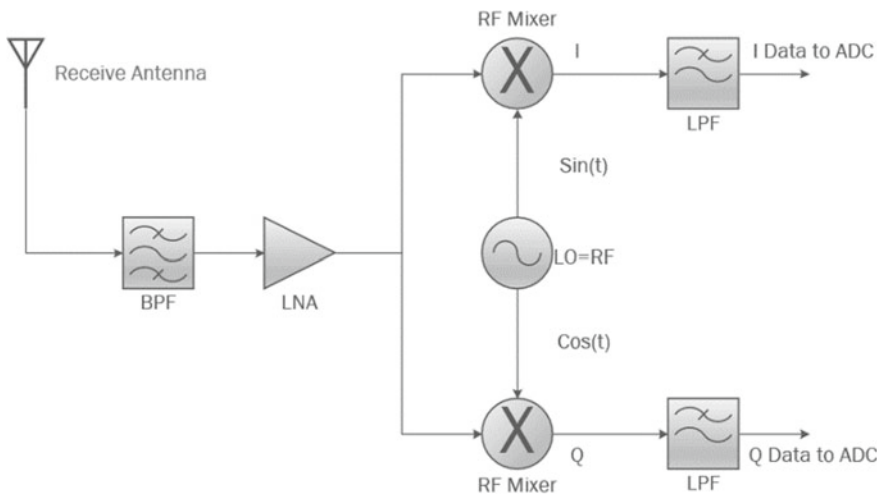


Fig. 3 Homodyne receiver



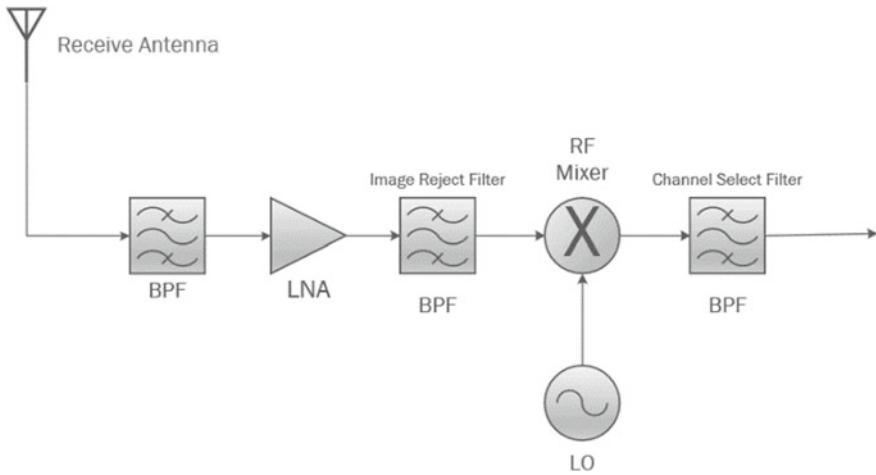


Fig. 4 Superhetrodyne receiver

from folding over to the positive-frequency half-channels after the down-conversion [8]. The negative-frequency half-channels are not used and filtered out with the LPF centred on 0 Hz [8]. This receiver architecture therefore has an advantage of eliminating the BPFs before and after the mixers. Namely, the image-reject filter and channel-select filter, shown in Fig. 5, are expensive and difficult to implement [9].

This receiver architecture also has its disadvantages. The desired frequency is frequency translated to a lower baseband signal. This presents a great deal of interference on the desired signal at baseband level, even though the image is completely non-existence after the LPF. Other disadvantages are listed below:

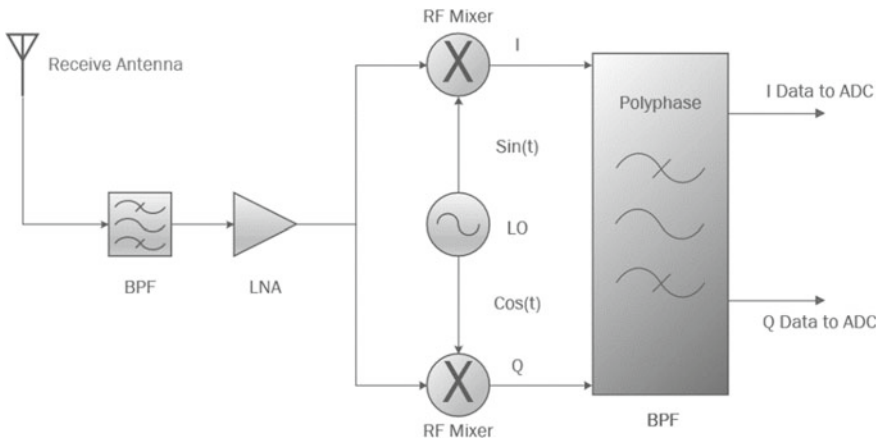


Fig. 5 Low-IF receiver

- LO-RF feedback: the LO signal interferes with the RF port when the LO-RF isolation is not satisfactory. The LNA will then amplify this interfering signal called direct current (DC) which consequently saturates the LNA and therefore restricts the desired RF signal from being enhanced [6].
- Even-order distortions: since RF mixers are nonlinear devices, they produce intermodulation distortion (IMD) products when the input RF port and LO port signals mix, giving a range of even and order harmonic products. The third-order products are only odd-order products of high concern, but only regarding IP3 of the receiver chain. On the other hand, all the even-order products can add to the baseband signal at DC and is known as the IP2 of DC receivers [10].
- Flicker noise/ $1/f$  noise: this is a type of noise that exists in all electronic components adds the DC offset and is inversely proportion with frequency [6].

## 2.2 SuperHeterodyne/Heterodyne Receiver

In 1918, Edwin Howard Armstrong first implemented this receiver architecture [9]. This receiver architecture in Fig. 4 initiates by band-limiting the desired RF signal into the LNA which amplifies and adds minimum noise to the band-limited signal. The image signal is then attenuated from the band-limited signal by the image-reject BPF. The band limited signal is mixed with the LO through the RF mixer, which results in the wanted IF and many other unwanted IMD products. The channel-select BPF removes the IMD products after the RF mixer. The stages to follow are usually the LO quadrature stage to allow phase and frequency modulation detection as in the homodyne receiver [11]. The advantage of having this arrangement is that the receiver has a very high ability to select the desired RF signal from a band-limited signal (receiver selectivity) and has a high ability to detect weak signals (receiver sensitivity). Both high receiver selectivity and high receiver sensitivity cannot be achieved simultaneously; therefore, a trade-off must be made [11].

High receiver sensitivity and thus good image rejection are achieved by allowing the IF signal to be as high as possible in frequency, away from the desired RF signal. To achieve this, the channel-select BPF after the RF mixer as in Fig. 4 requires a high Q factor because of the higher frequency. A high Q-factor filter is expensive, difficult to implement and has a higher insertion loss. High receiver selectivity is achieved by allowing the IF signal as low as possible and close to the desired RF. This creates relaxed requirements for the channel-select BPF. The downside to this is the high Q factor needed for the image-reject BPF Fig. 4 before the mixer which has a narrow bandwidth requirement. High Q-factor BPFs are difficult to implement and result in bad image suppression. Towards satisfying both high sensitivity and high selectivity simultaneously, the double superheterodyne receiver was created. The first LO frequency translates the desired RF signal to a higher IF. This satisfies the sensitivity parameter, and the second LO frequency translates the higher IF signal to a lower IF signal which satisfies the sensitivity parameter [12].

### 2.3 *Low-IF Receiver*

The concept of this receive architecture was to eliminate the image-reject filter like the homodyne receiver, but still have the selectivity and sensitivity as the superheterodyne receiver [13]. The process as in Fig. 5 initiates by filtering the input signal to a band-limited signal that is inputted to the LNA. This band-limited signal is provided to the LO quadrature to provide phase and frequency modulation detection. The IF is lower than normal in the range of kilohertz, therefore the use of a poly-phase filter Fig. 5 is needed for image rejection and channel selection. The poly-phase BPF has an asymmetric frequency response. This characteristic filters out all negative frequencies [9].

The advantage of having this architecture is that there is no LO self-mixing, DC offset and flick or  $1/f$  noise influences [13]. The disadvantage of using this architecture is that the wanted IF signal in the kHz is very close to the image signal and therefore needs a complex poly-phase BPF as in [14] to filter out the unwanted image signal. This consequence is that there is now very little room for I/Q balance mismatch [13].

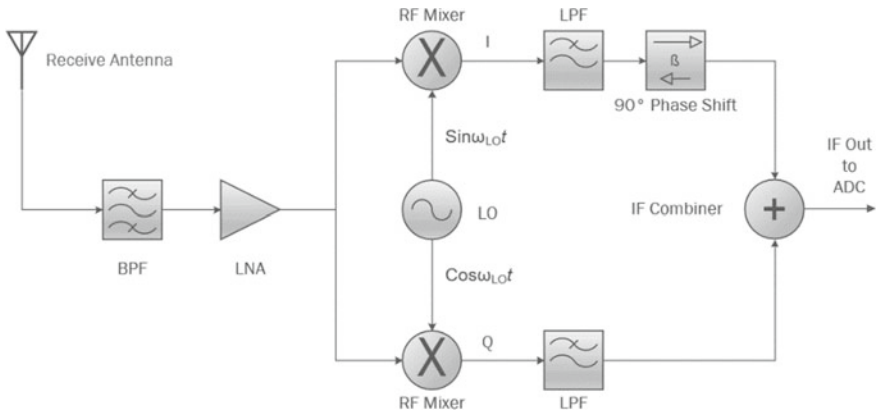
### 2.4 *Image-Reject Receivers*

The previous receive architectures all rely on passive components such as filters to remove the image frequency prior to the frequency translation process by the RF mixer to the respective IF and baseband signals. The following receive architectures rely on changing the phase of the band-limited signal so that when they combine, the wanted signals that are in-phase increases in power, while the unwanted image signals that are out-of-phase cancel out each other [15].

#### **Hartley Image-Reject Receiver**

The process as in Fig. 6 initiates by filtering the input signal to a band-limited signal that is inputted to the LNA. This amplified band-limited signal then becomes down-converted by the quadrature of the LO for phase and frequency modulation detection. This down-converted signal consisting of I and Q signals gets filtered by an LPF to limit higher frequencies from entering the stages to follow. Then, the I signal is phase shifted by  $90^\circ$ , while the Q signal is not shifted at all. Lastly, the I and Q signals combine to form a suppressed unwanted signal and an unambiguous wanted signal.

The advantage of using this architecture is that the component count is reduced with no image-reject BPFs. The disadvantage of this architecture is that the phase shift section and the LO must have decent phase and amplitude stability to sufficiently suppress the unwanted signal. Because of analog component tolerance, it is difficult to implement an accurate  $90^\circ$  phase shift; thus, a different approach is achieved by implementing  $45^\circ$  phase change on both the I and Q signal paths via a resistor capacitor-capacitor resistor (RC-CR) network configuration. This RC-CR



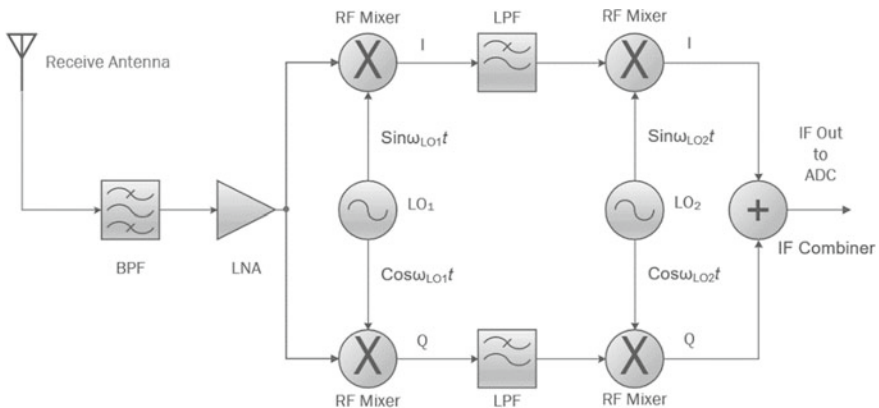
**Fig. 6** Hartley image-reject receiver

configuration though is affected by temperature variations which impact the phase and amplitude of the phase shifting [15].

**Weaver Image-Reject Receiver**

D. K. Weaver invented this receive architecture Fig. 7 which was to improve on the Hartley image-reject receiver by replacing the RC-CR network with an additional quadrature of the second LO. This receiver architecture operates the same as the former but with the second LO stage that converts the IF signal from the first LO into I and Q components. These components are added and subtracted to cancel the unwanted and promote the wanted signals.

The benefit of this architecture over the former is more stability under process and temperature variations. The disadvantage is still the same as the former where the LO needs to be phase and amplitude stable to ensure proper image rejection. The



**Fig. 7** Weaver image-reject receiver

addition of the down-conversion stage also requires more mixers and therefore more power to operate [15].

## 2.5 Summary and Conclusion

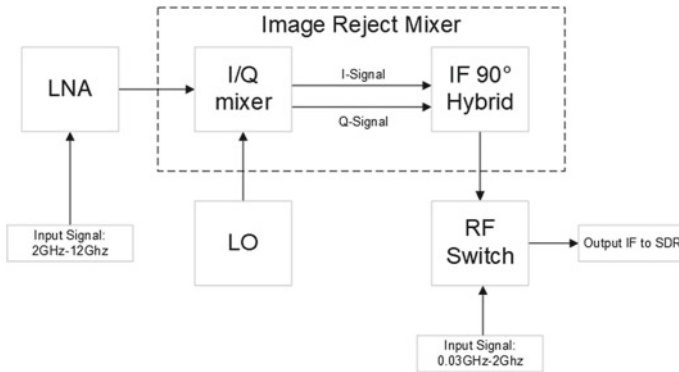
All receiver architectures use BPFs, which require high Q values for adequate filtering. This increases the cost and design complexity of the down-converter. Therefore, the least amount of BPFs needs to be allocated to decrease the cost and difficulty of the design. All receiver architectures use LOs, which is essential in translating the higher desired signal into a lower usable frequency. This unfortunately advances the current required for the translation process. Therefore, the least amount of LOs are needed for lower current usage and consequently lowering the cost. Poly-phase filters are used in the low-IF architecture to reduce adjacent and channel inference. These filters are more complex and inadvertently consume more power which rules out the consideration of this receive architecture. In Table 1, the Hartley architecture is seen to use the average amount of components.

The intention of this investigation is to deliver a receiver front-end for the mobile SDR environment that is simple to develop, uses the minimum amount of current and is less costly to implement. Research was completed in [16] where a front-end was added to the RealtekRTL2832U chipset SDR to extend the range to 6 GHz. The proposed design is different where it will as in Fig. 8:

- Detect signals up to X Band.
- Include a high IP3 LNA, which will allow the BPF before the LNA to be removed.

**Table 1** Number of components

Receiver architecture	LNA	LO	BPF	LPF	Mixer	Phase shifter	Combiner	Components
Homodyne	1	1	1	2	2	0	0	7
Superheterodyne	1	1	3	0	1	0	0	6
Double superheterodyne	1	2	4	0	2	0	0	9
Low IF	1	1	2	0	2	0	0	6
Image-reject Hartley	1	1	1	2	2	1	1	9
Image-reject Weaver	1	2	1	2	4	0	1	11



**Fig. 8** Conceptual front-end for the software-defined radio

- Include a MMIC I/Q mixer combined with MMIC IF quadrature hybrid combiner to form a Hartley image-reject mixer.

This reduces the number of components down to 4 compared to 9 in Table 1. Lastly, a RF switch is added, increasing the component count to 5 to allow the switching of the low- and high-frequency signals to the SDR.

## References

1. Krishnan R, Babu RG, Kaviya S, Kumar NP, Rahul C, Raman SS (2018) Software defined radio (SDR) foundations, technology tradeoffs: a survey. In: IEEE International Conference on Power, Control, Signals and Instrumentation Engineering ICPCSI 2017, pp 2677–2682
2. Li RC-H (2008) RF circuit design. Wiley, Hoboken, NJ, USA
3. Carr JJ (2000) The technician's radio receiver handbook, 2001st edn. Butterworth-Heinemann
4. Phang K (2002) Ece 1352F Rf image-reject receivers. Toronto, Ontario
5. Azzouni S, Khitouni N, Bouhleb MS (2019) Direct-conversion receiver front-end for LTE wireless network. In: 2019 19th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), pp 479–484
6. Razavi B (1996) Challenges in portable RF transceiver design. IEEE Circuits Devices Mag 12(5):12–25
7. Carrera A (2007) Design methodology for image-reject low-power receivers for wireless communications
8. Abidi AA (1995) Direct-conversion radio transceivers for digital communications. IEEE J Solid-State Circuits 30(12):1399–1410
9. Spiridon S (2016) Toward 5G software defined radio receiver front-ends. Springer International Publishing, Cham
10. John IM, Rogers WM, Plett C (2013) Radio frequency system architecture and design. Artech House Publishers
11. Cruz P, Gomes H, Carvalho N (2010) Receiver front-end architectures—analysis and evaluation. In: Advanced microwave and millimeter wave technologies semiconductor devices circuits and systems. InTech, pp 495–521
12. Razavi B, Behzad R (1998) RF microelectronics, 2nd edn. Prentice Hall

13. Puvaneswari M, Sidek O (2004) Wideband analog front-end for multistandard software defined radio receiver. In: 2004 IEEE 15th international symposium on personal, indoor and mobile radio communications (IEEE Cat. No. 04TH8754), vol 3, no I, pp 1937–1941
14. Khumsat P (2018) Multi-band low-IF receiver utilizing complex filter I/Q switching technique. In: TENCON 2018—2018 IEEE Region 10 Conference, October, pp 7–11
15. Idachaba FE, Orovwode HE (2011) Analysis of a Weaver, Hartley and Saw-filter based, image reject architectures for radio receiver design. *Adv Mater Res* 367:199–204
16. Kleber N et al (2017) RadioHound: a pervasive sensing network for sub-6 GHz dynamic spectrum monitoring. In: 2017 IEEE international symposium on dynamic spectrum access networks (DySPAN), pp 1–9

# Analysis for Time-Synchronized Channel Swapping in Wireless Sensor Network



M. Prathap, Gashaw Bekele, Melkamu Tsegaye and P. Karthika

**Abstract** The Internet of Things (IoT) is not just a promising inquire about subject yet in addition a sprouting modern pattern. In spite of the fact that the essential thought is to bring things or articles into the Internet, there are different methodologies, in light of the fact that an IoT framework is exceptionally application situated. This paper displays a remote sensor network (WSN)-based IoT stage for wide territory and heterogeneous detecting applications. The stage, comprising of one or different WSNs, entryways, a Web server, and a database, gives a dependable association between sensors at fields and the database on the Web. The WSN is fabricated dependent on the IEEE 802.15.4e time opened channel jumping convention, since it has the advantages such as multi-jump transmission, impact-free transmission and high vitality proficiency. Notwithstanding the plan of at equipment for range expansion, synchronization conspires and a burst transmission highlight is likewise displayed to support the system limit and decrease the vitality squander. Accordingly, the proposed stage can satisfy the high throughput necessity for high rate applications and the prerequisite of long battery life for low-rate applications in the meantime. It has built up a testbed in our grounds to approve the proposed framework.

**Keywords** Mobile ad hoc network · AODV (Ad hoc on demand distance vector) · DSDV (Destination-sequenced distance vector)

---

M. Prathap (✉)

Research & Development, Bharathiar University, Coimbatore, Tamil Nadu, India  
e-mail: [prathaparthi@yahoo.co.in](mailto:prathaparthi@yahoo.co.in)

G. Bekele

Department of Information System, Kombolcha Institute of Technology, Wollo University, Dessie, Ethiopia

M. Tsegaye

Department of Computer Science, Kombolcha Institute of Technology, Wollo University, Dessie, Ethiopia

P. Karthika

Kalasalangam Academy of Research and Education, Krishnankoil, Tamilnadu, India  
e-mail: [karthikasivamr@gmail.com](mailto:karthikasivamr@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_27](https://doi.org/10.1007/978-981-15-2329-8_27)

265



## 1 Introduction

The Mobile Ad hoc Network (MANET) is an accumulation of remote hubs speaking with one another in the nonattendance of any foundation. Study halls, combat zones and catastrophe help exercises are a couple of situations where MANETs can be utilized. MANET explore is making progress because of the omnipresence of little, cheap remote conveying gadgets. Since, very few MANETs have been sent, the vast majority of this examination is recreation based. These recreations have a few parameters counting the portability model and the conveying traffic design. In this paper, it has considered the center around effect of versatility models on the presentation of MANET steering conventions. It recognizes that the conveying traffic design likewise has a noteworthy effect on the directing convention execution and merits an examination all alone.

In any case, as in many investigations in this territory, so as to confine the impact of portability, fix the imparting traffic example to comprise of haphazardly picked source-goal sets with sufficiently long session times. Numerous past in versatility design works was randomly accepted to be waypoint. In the present test system (ns-2) dissemination, the portability model executed as per the following: At each moment, in a hub haphazardly picks a goal and consistently to moves toward with a speed picked arbitrarily from  $[0, V_{\max}]$ , where the most extreme of  $V_{\max}$  reasonable speed for the each portable hub [1]. The utilizing of the majority reproductions the random waypoint model depends on this standard usage. For the remainder of the paper, allude to this fundamental execution as the random waypoint model. Later on, MANETs are relied upon to be sent in heaps of situations having complex hub portability and availability elements. For instance, in MANET on a front line, the development of the fighters will be impacted by the authority. In city-wide MANET, the hub development is confined by hindrances or maps. The hub portability attributes are the application explicit. Generally, differing portability attributes are required to significantly affect the execution of the directing conventions like DSR [2], DSDV [3] what is more, and AODV [4]. Irregular a waypoint is very much structured model be that as it may, it is deficient to catch the accompanying qualities:

1. The development hubs among the spatial reliance.
2. Dependence on the temporal of development of a hub after some time.
3. Obstructions of existence or snags obliging versatility.

In Fig. 1, investigation is center around the effect of the previously mentioned portability attributes on convention execution. While doing in this way, propose a nonexclusive structure to efficiently breakdown the effect of versatility on the exhibition of directing conventions for MANETs.

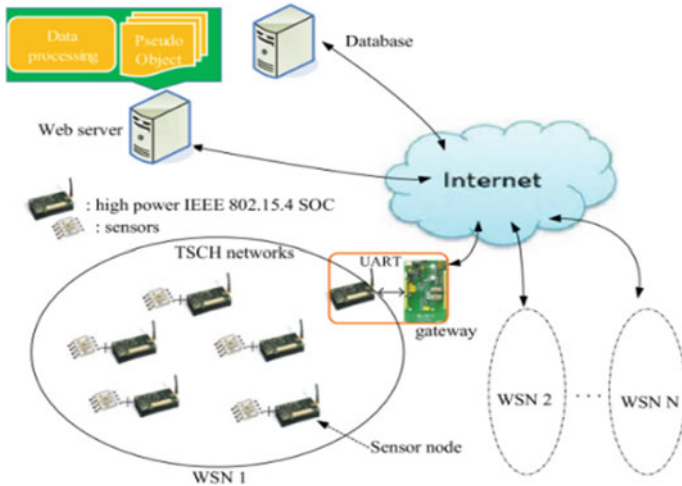


Fig. 1 Channel swapping in wireless sensor network using MANET framework

## 2 Related Work and Comparison

The past work on the exhibition investigation of the versatile specially appointed steering conventions is diagrammed. It is seen that a few papers think about under four conventions in their examination while some others do not consider. A few papers change portability however do not think about speed as a significant variable. System burden or shifting quantity of the hubs is a major factor which is the directing presentation of the impromptu conventions; however, not many papers have made a far-reaching examination over it. For occasion, Payal [5] makes an exhibition correlation of the steering conventions for specially appointed systems with the number of hubs to be fixed. They are looking for standard Dijkstra calculation with AODV, OLSR and DSR. As indicated by CBR sources, the exhibition of parcels effectively conveyed is very high (over 90%). In spite of the fact that this is valid, in any case, increment in rush hour gridlock load altogether diminishes the general parcel conveyance proportion. This is mimicked and contemplated in our examination.

In Fig. 2, the point considers to the versatility model and random waypoint model, the ideal model for managing MANET systems hubs can move toward an any path. Ganesh Babu and Amudha [2] give execution assessment looking at three steering conventions—AODV, DSR and OLSR. This paper basically talks about interactive media transmission more than 50 hubs and breaks down the execution of the directing conventions. Dissimilar to our recreation situation it gives an examination over a broad scope of hubs, this paper [5] has a number of hubs consistent and pursues a Manhattan Grid model contradicts the arbitrary waypoint model considered in this exploration. Execution likes the parameters bundle conveyance proportion, start to finish delay and steering overhead have been broke down yet for the most part these parameters plotted are against the number of associations which considered

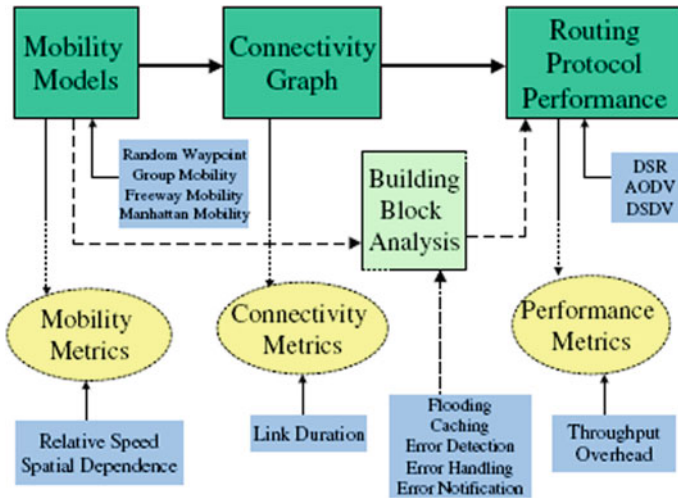


Fig. 2 MANET framework

the primary variable by Payal [5]. As per [6], delay time fundamentally decides the versatility rate of the model on the grounds that as respite time increment, versatility increments. Along these lines delay time considered as the standout among for most significant variable dissecting portability rate of exploration. A generally amazing correlation between the DSR, DSDV and AODV is made in [7], and the system load is expanded by five hubs for each reenactment. Ganesh Babu and Amudha [7] look at just three conventions dissimilar to this exploration which thinks about all the four. It has thought about a limit of 20 hubs not at all like this examination which analyzes the system investigation up to 70 hubs. Moreover, Ganesh Babu and Amudha [7] do not plainly state what number of sources and recipients are conveying and TCP specialist is utilized. It is additionally not clear, the greatest speed of the hubs is when examination of the steering conventions is made. Just two specially appointed directing conventions DSR and AODV are examined in [8] contrasted with four conventions which have been done in this examination.

It for the most part examined the between layer associations between the physical and the MAC layer and their presentation suggestions. The vast majority of the recreation parameters and execution parameters were comparatively contrasted with this exploration. They have additionally incorporated the random waypoint model and the spine model, keep running for 50–100 hubs with various reenactment time for each measurement. Speed has fluctuated between the 0–20 m/s and the number of sources between as 10–30. It has discovered DSR showed altogether low-level steering, load than the AODV and equivalent to the consequences of this examination. The paper sees AODV beats DSR as far as parcel conveyance proportion notwithstanding when the system load is expanded which is tantamount to the outcomes that drawn by this examination. Be that as it may, this paper carefully follows to correlation of just two specially appointed directing conventions and not gives examination with

proactive steering calculations like the OLSR and the DSDV. An exhibition examination between all the four MANET directing conventions DSDV, AODV, DSR and OLSR gives in [9]. Anyway for the paper depends on hypothetical investigation and does not give anything identified with recreation and the result examination. This paper does not talk anything explicit about execution investigation identified with the OLSR convention.

### 3 Proposed Model and Performance Parameters

A model is planned dependent on a portion of significant presumptions that not considered by these papers referenced in the past segment. The whole research is partitioned into the three stages—portability, system burden and speed to make a far-reaching execution examination of the portable specially appointed steering conventions. Three noteworthy execution parameters—normal throughput of the system, bundle conveyance portion and number of parcels dropped—are considered to decide the presentation of the impromptu directing conventions. In the first and second stage, is reenacted to confirm precision of the conventions. Be that as it may, OLSR is added to break down its exhibition against different conventions. In the first stage, versatility is viewed as keep the quantity of hubs and CBR sources steady. In the second stage, greatest speed for the fundamental variable and presentation parameters are plotted against the speed. The third stage manages system thickness to decide the specially appointed steering conventions perform against different system loads. The CBR sources are generally taken to 33% the quantity of hubs to look after consistency.

$$LC(i, j) = \sum_{t=1}^T C(i, j, t) \tag{1}$$

where

LC → Link Change

$i, j \rightarrow i$  and  $j$  is the number of times the link between them transitions

$C(i, j, t)$  is an indicator random variable.

If the nodes between link  $i$  and  $j$  is down at the is time  $t - 1$ , the time comes up is  $t$

$$LC = \frac{\sum_{i=1}^N \sum_{j=i+1}^N C(i, j)}{P} \tag{2}$$

where

$N \rightarrow$  It is the value of  $LC(i, j)$  averaged over the number of node

$P$  is number of the pairs  $I, j$  is such that  $X(i, j) = 0$ .

This model decides the exhibition examination of the steering convention in a changed scale of systems it incorporates little, medium and vast systems. Delay time is considered as one of the fundamental variables for examination since most extreme versatility change is a standout among the most significant factor is the moving nodes [6]. The total bundle investigation given the quantity of parcels sent, dropped, got and sent is broke down to check exhibition of the versatile impromptu steering conventions under different traffic condition.

$$PA(i, j) = \frac{\sum_{t=\text{start}(i,j)}^T A(i, j, t)}{T - \text{start}(i, j)} \tag{3}$$

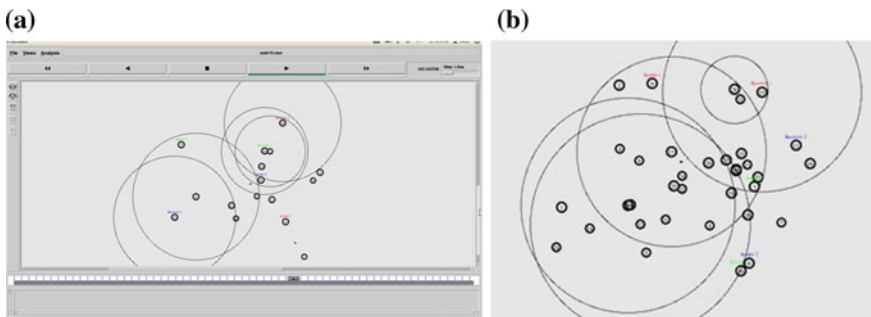
where

PA is Average Path Availability  
 If  $T - \text{start}(i, j) > 0$  otherwise 0

$$PA = \frac{\sum_{i=1}^N \sum_{j=i+1}^N PA(i, j)}{P} \tag{4}$$

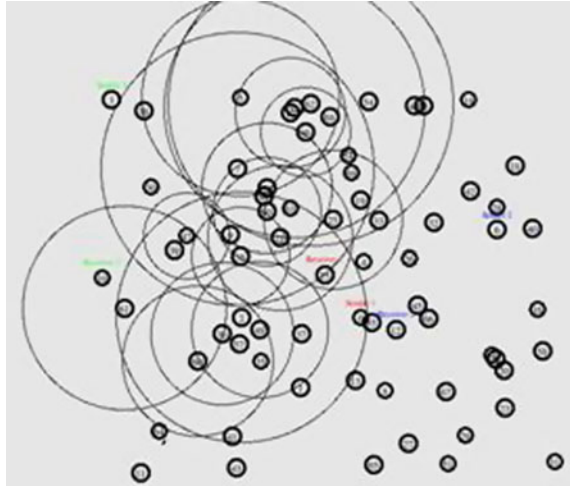
Figure 3a, b demonstrates the situations with three unique sizes of system . It is expected that anything underneath 25 hubs is a little system, 25–50 is medium, and 50–100 hubs are an extensive system.

In a little system situation as appeared in Fig. 3a, b, the quantity of hubs framing the arrange is normally less. There are 16 hubs and the hubs are haphazardly moving in various ways. The circles around the hubs are remote transmission run and dependent on the scope of the diverse portable hubs, the goal for the information transmitted from the source. In a medium of system situation as appeared in Fig. 3 the quantity of hubs making the organize are 35. They are bigger than the little system situation and diverse steering conventions act in various ways attributable to the difference in the system conditions. Figure 4 demonstrates an expansive system situation which



**Fig. 3** a Small network scenario using of 16 nodes; b medium network scenario using of 35 nodes

**Fig. 4** Large network scenario consisting of 70 nodes



comprises of 70 hubs. In such expansive systems, it is hard to accomplish 100% parcel conveyance division. This is checked in the inevitable areas of this examination.

#### 4 Performance Parameters

In the three primary executions, parameters that considered for this exploration—normal throughput, parcel conveyance division of the quantity—are bundles dropped. Normal throughput decides the steadiness system in various rush hour gridlock conditions. Bundle conveyance portion records to the level of parcels conveyed when arrange is exposed to various traffic rules. Number of bundles dropped is the considered if to watch the quantity of bundles got is influenced more sent by the parcels or the dropped bundles. These three parameters assessed through in three periods of examination it makes exhibition investigation of the specially appointed steering conventions. It is giving the portion of channel limit utilized for valuable transmission for (data bundles accurately conveyed to the goal) characterized as all out number of the bundles gotten by the goal. It is truth a proportion of viability of a directing convention estimated in bits/second.

$$\text{Throughput} = (\text{Number of bundles sent} * 8 * 512) / \text{Simulation Time} \quad (5)$$

It is the proportion of information bundles got to parcels sent. It informs us regarding the division of the parcels conveyed from source to goal when the system is exposed to various traffic conditions. It likewise gives a thought regarding the quantity of parcels dropped or sent by the steering convention.

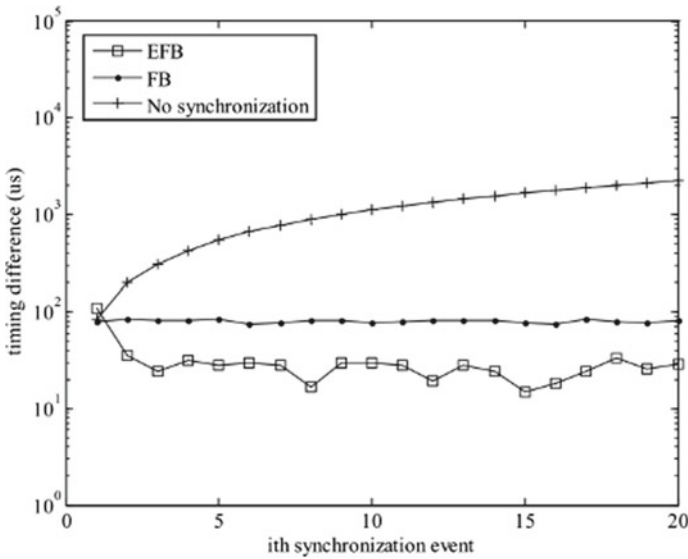


Fig. 5 Performance analysis by varying speed

$$\text{Parcel Delivery Fraction} = \text{Number of bundles Received} / \text{Number of parcels sent} \tag{6}$$

The parameter which is essential significance is stop time. In Fig. 5, interruption time fundamentally decides the portability rate of the model, as delay time builds the versatility rate decrease [6]. Respite time for measure it was taken by everyone of moving hubs before they are begin transmitting parcels. At the point g when the respite time is huge and hang tight time of the hubs is huge, the versatility is low on the grounds that the hubs are not persistently sending parcels. At the point of respite time is to low, the hang tight time to the hubs is to low and subsequently the versatility is huge. It implies the hubs are continually sending parcels with no hold up time.

## 5 Conclusion

A standout among the most encouraging system that has risen up out of the innovation world is the versatile specially appointed system or MANET. It is a sort of multi-bounce arrange. Remote naturally, MANETs do not have a particular system framework. It is a gathering of remote cell phones that speak with one another without the assistance of any outsider spine like to base station or switch. It tends to difficult the envision for the every hub in sort of system speaks to each other without having the switch. The MANETs, hubs with changes for area time, arrange themselves to

get data transmitted from the source to goal without assistance it any switch or the base station. Subsequently, are effective information transmission, it is basic to comprehend the kind of steering that is being utilized by these systems. Since it has no particular switches are deal with the undertakings, it is very well may be grand errand for the hubs to proficiently decide a way to advance and course their parcels and they are steady movement. The future work it makes a far-reaching execution investigation of the different versatile specially appointed directing conventions. More than 160 reenactment situations have been directed and upwards of six execution parameters are used and looked at in three changed sizes of system to make it an exhaustive investigation.

## References

1. Ehsan H, Uzmi AZ (2004) Performance comparison of adhoc wireless network routing protocols. In: 8th IEEE international multi-topic conference (INMIC), pp 457–465, Lahore, Pakistan
2. Ganesh Babu R, Amudha V (2018) Allow an useful interference of authenticated secondary user in cognitive radio networks. *Int J Pure Appl Math* 119(16):3341–3354
3. Talooki VN, Ziarati K (2006) Performance comparison of routing protocols for mobile adhoc networks. In: IEEE Asia-Pacific conference on communications (APCC), pp 1–5, Busan, South Korea
4. Ganesh Babu R, Karthika P, Aravinda Rajan V (2019) Secure IoT systems using raspberry Pi machine learning artificial intelligence. In: Smys S, Senjyu T, Lafata P (eds) Proceedings of second international conference on computer networks and inventive communication technologies. Lecture notes on data engineering and communications technologies. Springer, Singapore, in press
5. Payal JSK (2013) CBR traffic based performance investigations of DSDV, DSR and AODV routing protocols for MANET using NS2. *Int J Soft Comput Eng* 3(4):12–16
6. Ganesh Babu R, Karthika P, Elangovan K (2019) Performance analysis for image security using SVM and ANN classification techniques. In: Third IEEE international conference on electronics, communication and aerospace technology, Coimbatore, India, in press
7. Ganesh Babu R, Amudha V (2014) Spectrum sensing techniques in cognitive radio networks: a survey. *Int J Sci Eng Res* 5(4):23–32
8. Gupta P, Gupta S (2013) Performance evaluation of mobility models on MANET routing protocols. In: Third international conference on advanced computing and communication technologies (ACCT), pp 248–253, Rohtak, India
9. Ganesh Babu R, Amudha V (2019) A survey on artificial intelligence techniques in cognitive radio networks. In: Abraham A, Dutta P, Mandal J, Bhattacharya A, Dutta S (eds) Emerging technologies in data mining and information security, vol 755. *Adv Intell Syst Comput*. Springer, Singapore, pp 99–110
10. Gouda BS, Dass AK, Narayana KL (2013) A comprehensive performance analysis of energy efficient routing protocols in different traffic based mobile ad-hoc networks. In: IEEE international multi-conference on automation, computing, communication, control and compressed sensing (iMac4s), pp 306–312, Kottayam, India
11. Rajeswari M, Maheswari P, Bhuvanawari S, Gowri S (2012) Performance analysis of AODV, DSR, TORA and OLSR to achieve group communication in MANET. In: 4th international conference on advanced computing (ICoAC), pp 1–8, Chennai, India (2012)
12. Karthika P, Vidhya Saraswathi P (2017) Content based video copy detection using frame based fusion technique. *J Adv Res Dyn Control Syst* 9(17):885–894



# Adiabatic Design Implementation of Digital Circuits for Low Power Applications



Bhavika Mani, Shaloo Gupta and Hemant Kumar

**Abstract** This paper presents the comparative analysis of average power dissipation for conventional CMOS and different adiabatic logic techniques like efficient charge recovery logic (ECRL) and positive feedback adiabatic logic (PFAL) based digital circuits like inverter, NAND, NOR, 2:1 MUX, EXOR, and full adder. These circuits are based on reversible logic that works on AC power supply which can be trapezoidal or sinusoidal voltage source. The analysis of average power and delay is carried out at 180, 90, and 45 nm technology files for different frequencies. The result shows the significant reduction in power dissipation up to 26, 36, 16, 59, 73, 99% for inverter, NAND gate, NOR gate, EXOR gate, 2:1 MUX, and full adder circuits, respectively with adiabatic logic comparatively CMOS within specified frequency range of 1 kHz to 1 MHz and also manifests till what extent the power can be reduced so as to avoid degradation in performance. The design and simulation are performed on cadence virtuoso EDA tool.

**Keywords** CMOS · Adiabatic logic · Four-phase power clock · ECRL · PFAL

## 1 Introduction

In the area of integrated circuit design, the growing technology and demand necessitate the immediate effort in the development of low power VLSI design circuit. Several approaches have been established to minimize power dissipation in which adiabatic logic is proposed as a novel class of logic circuits that is based on AC (i.e., time-varying) power supply in place of constant DC supply used in conventional CMOS technology [1]. In today's context, researchers are taking effort to acquire

---

B. Mani (✉) · S. Gupta · H. Kumar  
Department of Electronics, Banasthali Vidyapith, Newai, Rajasthan 304022, India  
e-mail: [bhavikamani25@gmail.com](mailto:bhavikamani25@gmail.com)

S. Gupta  
e-mail: [Shaloo8196@gmail.com](mailto:Shaloo8196@gmail.com)

H. Kumar  
e-mail: [erhemantkumar1987@gmail.com](mailto:erhemantkumar1987@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_28](https://doi.org/10.1007/978-981-15-2329-8_28)

several optimization techniques and principle of energy conservation to design VLSI circuit that led to develop a new approach of switching logic compare to conventional CMOS technique which is adiabatic switching logic. Adiabatic logic circuits are circuits with low power significance, and to conserve energy, it uses reversible logic. In the load capacitor, stored energy is recycled and retrieved to time-varying (pulsed) power supply instead of dissipating the stored energy to ground [2].

In 2016, K. Srilakshmi compared the performance of ECRL, PFAL, 2PC2AL, and CAL adiabatic logic families based on bulk CMOS and FinFET by designing a Brent-Kung adder of 4-bit. The result illustrates that by replacing bulk CMOS with FinFET, up to 76–97% of power savings can be attained [3]. Later in 2017, the work proposed by Mei Han used LT Spice tool to simulate  $4 \times 4$  multiplier based on two-phase clocked CMOS (2PC2AL). It shows power dissipation can be reduced comparatively CMOS circuit in the range of frequencies that is from 100 Hz to 100 MHz [4]. The work recommended by Samik Samanta in 2017 calculates the power dissipation and compares the DFF, TFF, and BCD counter based on cascaded adiabatic logic. Power dissipation of DFF based on CMOS and various energy recovery logics is also calculated. The result specifies that better performance is obtained in PFAL-based DFF [5]. The proposed work focuses on designing various digital circuits by using conventional CMOS and different adiabatic techniques for comparative analysis of power dissipation and delay. Comparison between two adiabatic logic styles—ECRL and PFAL—is also described in this paper. It also demonstrates the reversible logic concept.

In Sects. 2 and 3, we discussed the conventional CMOS and four-phase adiabatic logic, respectively. Section 4 presents the discussion of results along with the design parameters used in implementation of digital circuits and also some graphical representation to show relation between different parameters. The whole work is concluded in Sect. 5.

## 2 Conventional CMOS

Conventional CMOS logic uses constant DC power supply while adiabatic logic is based on the usage of AC power clock or sinusoidal power supply, due to this adiabatic circuit is capable to recover the energy to power supply from the node capacitor. AC power clock consists of four phases of clock, i.e., evaluate, hold, recovery, and wait with phase difference equals to  $T/4$  where  $T$  is clock period.

Energy transfer takes place in conventional CMOS logic circuits from power supply to output and from output to ground with every switching. During the transition from 0 to  $V_{DD}$  at output, from power supply the energy is drawn which is equal to  $C_{LOAD}V_{DD}$ . Output node capacitance is charged up to  $V_{DD}$ , i.e., half of the energy equal to  $(C_{LOAD}V_{DD}^2/2)$  is stored in load capacitor and in the PMOS network and the rest half of the energy is dissipated. During the transition from  $V_{DD}$  to ground at output, no energy is consumed from the power supply and while during the charging

phase, energy in the load capacitor will be dissipated in NMOS network [6]. Therefore, there is no possibility to recover the energy in conventional CMOS circuits. Primarily, there are two kinds of power dissipation:

### Dynamic Power Dissipation

Due to transition from logic 0 to logic 1 at output and vice versa, switching activity occurs in device and because of such phenomenon switching (dynamic) power dissipation occurs. Operation of circuit is defined with consideration that the circuit consists of NMOS and PMOS and input switches from logic 0 to 1 and logic 1 to 0, and load capacitor is placed at output. Dynamic power can be defined as

$$P_{\text{dyn}} = C_L V_{\text{DD}}^2 f \quad (1)$$

where  $C_L$  is representing load capacitance,  $V_{\text{DD}}$  denotes the power supply, and  $f$  is frequency.

### Static Power Dissipation

In this type of power dissipation, the current flows in the circuit even when the power supply is turned off, i.e., it depends only on logical state of circuit and not on switching activity. Static power dissipates due to leakage current. It can be defined as

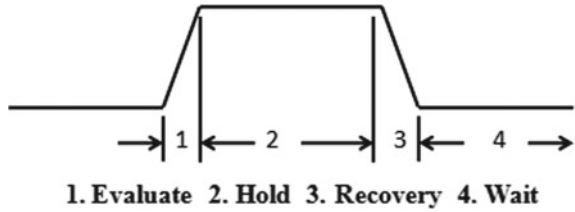
$$P_{\text{static}} = I_{\text{static}} V_{\text{DD}} \quad (2)$$

where  $I_{\text{static}}$  denotes the current flowing in the circuit when no switching activity occurs and  $V_{\text{DD}}$  is the DC power supply.

## 3 Adiabatic Logic

Instead of employing a constant DC power supply, adiabatic logic utilizes time-varying sinusoidal power supply or constant current source that enables it to recover the stored energy at output node in load capacitor to the power supply [6]. By using adiabatic technique dissipated power in circuit is conserved as energy and reused back. According to desired applications and system, this approach is used for minimizing power dissipation. The principle of adiabatic switching states is that there is need of power supply during transition from logic 0 to 1 but no need of power supply in case of transition from logic 1 to logic 0 due to use of sinusoidal power supply as input. So power dissipation can be reduced. Adiabatic technique does not allow instantly switching from logic high to low and vice versa. By using this approach, true and complement output both will be obtained in same circuit. Adiabatic circuits use four-phase power clock or combined power supply and clock that make it capable to retrieve the stored energy in load capacitor to power source. Four-phase

**Fig. 1** Four-phase power clock



power clock has a phase difference which is equal to one-fourth of period. Operation contains four phases: evaluate, hold, recovery, and wait (Fig. 1).

In **Evaluate** Phase, power supply goes slowly to  $V_{dd}$  from 0 and evaluation of output will be done. In complete **Hold** Phase, the output is kept reasonable when power clock remains high and provides constant input signal. In **Recovery** Phase, power supply goes slowly to 0 from  $V_{dd}$ , and energy is recovered from charged load capacitor to power source. In **Wait** Phase, power clock remains low that maintains the output at low value [7].

Figure 2 represents RC model for adiabatic switching, where constant current source or time-varying (sinusoidal) power supply is used to charge the load capacitor in place of constant voltage source used in conventional CMOS approach. We are assuming here that the period  $T$  of ramp signal is much greater than time constant  $RC$  of the circuit, i.e.,  $T \gg RC$ . Therefore, the voltage across the load capacitor  $V_c$  is nearly following the supply voltage that results in the difference of potential across resistor  $R$ , which is almost zero [8].

Voltage across capacitor,

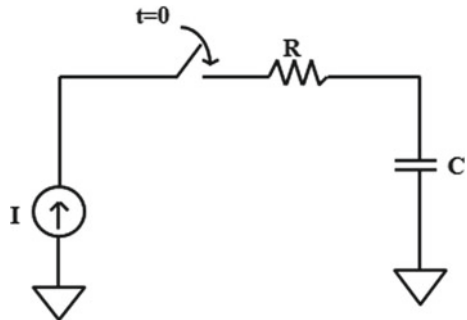
$$V_C(t) = 1/C \cdot I(t) \cdot t \tag{3}$$

During 0 to  $t$ , average current value

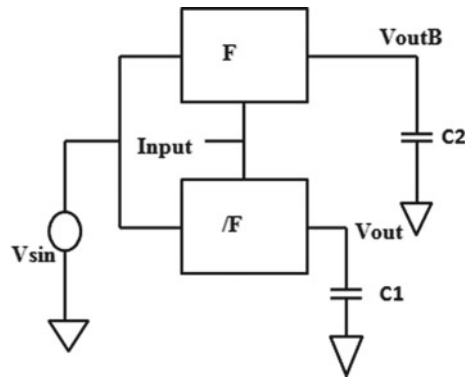
$$I(t) = C \cdot V_C(t)/t \tag{4}$$

From 0 to time  $T$ , total energy in resistor,

**Fig. 2** Principle of adiabatic switching



**Fig. 3** Adiabatic logic gate implementation topology



$$E_{diss} = RC/T(CV^2) \tag{5}$$

where  $V$  is supply voltage swing and  $T$  is period of power clock.

Figure 3 represents a general circuit topology that implements adiabatic function. For charging and discharging process the output node capacitance in adiabatic logic circuit pull-up network denoted by functional block  $F$  and pull-down network denoted by functional block of inverted  $F$  are used that ensures at the end of every cycle stored energy at output node can be recycled and recovered back by power supply.  $V_{pwr}$  is used in place of  $V_{dc}$  [9].

Adiabatic circuit is characterized by two fundamental classes:

**Fully Adiabatic Circuit**

The circuit in which complete charge stored in output node capacitance is retrieved back by power supply is termed as fully adiabatic circuit. The circuits are as follows:

- Split rail charge recovery logic (SCRL),
- Pass transistor adiabatic logic (PAL).

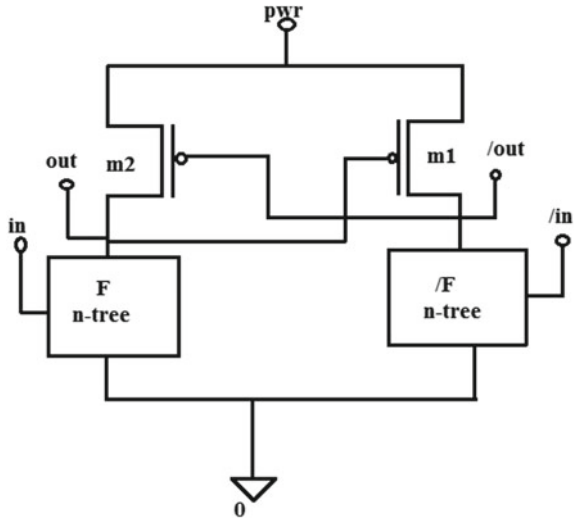
**Partially Energy Recovery Adiabatic Circuit**

The circuit in which only some amount of charge is carried back to power supply and remaining transferred to ground is termed as partially adiabatic circuit. The circuits are as follows:

- 2N-2N2P adiabatic logic,
- Efficient charge recovery logic (ECRL),
- Source-coupled adiabatic logic (SCAL),
- Positive feedback adiabatic logic (PFAL),
- NMOS energy recovery logic (NERL).

**ECRL** The structure of standard CMOS-based ECRL is similar to differential signaling with cascode voltage switch logic (CVSL). To evaluate the functions, this structure uses pair of NMOS devices, and to hold state, a latch combination of PMOS devices is used. It is called a quasi adiabatic logic as it is not possible to

**Fig. 4** Schematic for ECRL logic block



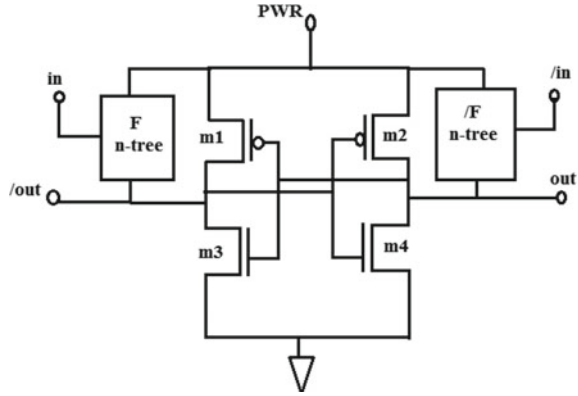
completely recover the power clock using the PMOS devices. Based on a latch of PMOS transistors, we have ECRL circuits where the power clock is directly supplied to the source terminals and the gate of PMOS devices is coupled to the drain of other and these nodes result in the output signals which are complementary in nature. A series of NMOS devices are there to evaluate the function [10–12] (Fig. 4).

**PFAL** PFAL is the abbreviation for positive feedback adiabatic logic. Against variations in technological parameter, it has a good robustness. An adiabatic amplifier is the fundamental element of PFAL gate and consists of a latch which is the combination of two PMOS P1-P2 and two NMOS N1-N2 which prevents the degradation of logic level at nodes on the output, i.e., out and /out. The logic function is implemented using the two  $n$ -trees. PFAL uses four-phase clock  $\Phi$ . At high frequency during the charging phase of the capacitor, a minimization in the dissipation of energy can be attained by placing the functional block in parallel combination with the pull-up transistors of the adiabatic amplifier which results in decrement in equivalent resistance. This is the major advantage of PFAL logic over ECRL [13] (Fig. 5).

## 4 Results and Discussion

Table 1 describes the design parameter used in implementation of combinational circuits. Constant pulsed power supply (square pulse) used in CMOS technique while sinusoidal power supply is used in different adiabatic techniques like ECRL and PFAL. Various input combinations are used to design circuits for which required specifications like time period, rise time, fall time, and delay time are provided.

**Fig. 5** Schematic of PFAL logic block



**Table 1** Design parameters

Type		CMOS			Adiabatic logic		
		45 nm	90 nm	180 nm	45 nm	90 nm	180 nm
NMOS	(Width)	120 nm	120 nm	2 $\mu$ m	120 nm	120 nm	2 $\mu$ m
PMOS	(Length)	45 nm	100 nm	180 nm	45 nm	100 nm	180 nm
Supply voltage		1.1 V DC power supply	1.2 V DC power supply	1.8 V DC power supply	1.1 V AC power supply at different frequencies	1.2 V AC power supply at different frequencies	1.8 V AC power supply at different frequency

Table 2 concludes the result for given combinational circuits with different technology (45, 90, 180 nm) using different techniques at different frequencies of 100 kHz. It shows the significant power reduction in ECRL and PFAL adiabatic technique comparatively CMOS.

Table 3 draws the parallels between power dissipation in ECRL and PFAL adiabatic technique using different frequencies. From the table we can observe, for a specified range of frequency from 1 kHz to 100 MHz, ECRL contributes less power dissipation comparatively PFAL. PFAL proves to be better power reduction technique than ECRL above certain frequency.

**Graphical Representation to Represent the Relationship Between Different Parameters**

By increasing the channel length of inverter designed with different methodology as shown in Fig. 6, it is examined that power dissipation is also increased almost linearly. But this power dissipation is less with adiabatic techniques when compared to that of CMOS technique.

The graphical representation of power dissipation and supply voltage is shown in Fig. 7 for the case of inverter circuit, and it concludes that when the supply voltage

**Table 2** Comparison of power dissipation at different technology (180, 90, and 45 nm)

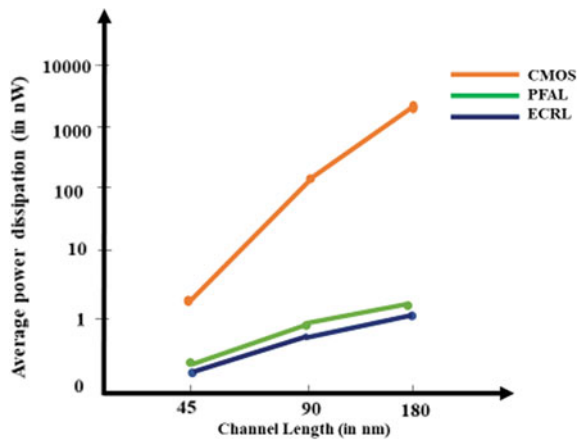
Circuits	Conventional circuit			Adiabatic circuit					
	180 nm (1.8 V)	CMOS 90 nm (1.2 V)	45 nm (1.1 V)	180 nm (1.8 V)	PFAL 90 nm (1.2 V)	45 nm (1.1 V)	180 nm (1.8 V)	ECRL 90 nm (1.2 V)	45 nm (1.1 V)
Inverter	2.394 $\mu$ W	266 nW	4.32 nW	1.620 nW	367.9 pW	27.13 pW	1.325 nW	109.2 pW	14.92 pW
2 input NAND	1.4016 $\mu$ W	316.4 nW	11.70 nW	1.238 nW	1.098 nW	69.39 pW	12.869 nW	40.05 pW	131.5 pW
2 input NOR	2.3881 $\mu$ W	254 nW	13.64 nW	2.41 nW	770.2 pW	45.09 pW	44.76 nW	110.2 pW	67.50 pW
2 input EXOR	7.580 $\mu$ W	988.7 nW	53.26 nW	3.8 nW	1.092 nW	81.83 pW	1.181 nW	99.12 pW	13.94 pW
2:1 MUX	8.197 $\mu$ W	1.009 $\mu$ W	7.73 $\mu$ W	5.870 nW	939.7 pW	73.75 pW	3.949 nW	225.8 pW	42.36 pW
Full adder	11.468 $\mu$ W	1.247 $\mu$ W	90.07 nW	15.06 nW (sum)	1.707 nW (sum)	185.6 pW (sum)	13.41 nW (sum)	8.03 pW (sum)	173.5 pW (sum)
				19.49 nW (carry)	1.957 nW (carry)	202.0 pW (carry)	8.918 nW (carry)	181.2 pW (carry)	189.5 pW (carry)



**Table 3** Comparison of power dissipation between different adiabatic techniques at different frequencies (180 nm technology)

Circuits	Adiabatic		Frequency				
	Logic		100 MHz	10 MHz	100 kHz	10 kHz	1 kHz
Inverter	ECRL		7.04 mW	1.865 $\mu$ W	1.325 nW	17.98 pW	982.3 fW
	PFAL		44.55 mW	246.6 $\mu$ W	1.620 nW	116.6 pW	14.92 pW
NAND	ECRL		9.664 mW	2.275 $\mu$ W	12.87 nW	13.69 pW	3.18 pW
	PFAL		48.60 mW	77.5 $\mu$ W	1.238 nW	318.9 pW	38.10 pW
NOR	ECRL		6.79 mW	2.80 $\mu$ W	44.76 nW	305.9 pW	5.203 pW
	PFAL		65.99 mW	79.84 $\mu$ W	2.415 nW	824.1 pW	86.35 pW
EXOR	ECRL		1.47 mW	178.6 $\mu$ W	1.181 nW	13.64 pW	63.90 fW
	PFAL		84.3 mW	207.3 $\mu$ W	3.8 nW	439 pW	51.77 pW
2:1 MUX	ECRL		1.345 mW	2.978 $\mu$ W	3.946 nW	209.4 pW	23.03 pW
	PFAL		84.67 mW	144.7 $\mu$ W	5.870 nW	267.6 pW	34.64 pW
Full adder	ECRL	Sum	149.1 mW	607.8 $\mu$ W	13.41 nW	388.7 pW	59.0 pW
		Carry	99.10 mW	385.6 $\mu$ W	8.918 nW	122.0 pW	22.12 pW
	PFAL	Sum	153.5 mW	210.5 $\mu$ W	15.06 nW	342.3 pW	57.47 pW
		Carry	103.4 mW	199.4 $\mu$ W	19.49 nW	136.2 pW	26.24 pW

**Fig. 6** Relation between average power dissipation and channel length for different techniques



across any circuit is reduced, the power dissipation in the circuit will also reduce, so we can say that power consumption is directly proportional to supply voltage.

From Fig. 8, it is seen that power consumption is minimum in case of ECRL and maximum in PFAL for the particular range of frequency, i.e., 1 kHz to 100 MHz. The applications of both the techniques depend on the range of frequency in which the circuit is to be operated.

Figure 9 demonstrates the trade-off between average power and delay for both CMOS and adiabatic techniques at 100 kHz. It concludes that average power is

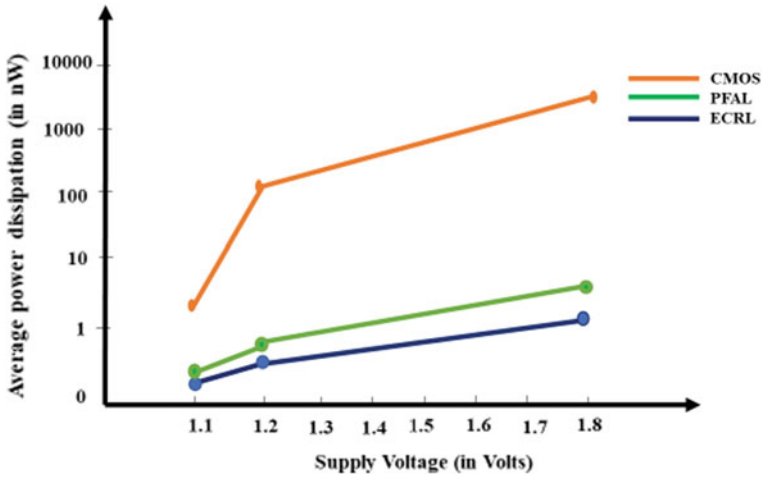


Fig. 7 Comparison of average power dissipation at different supply voltages by using different techniques

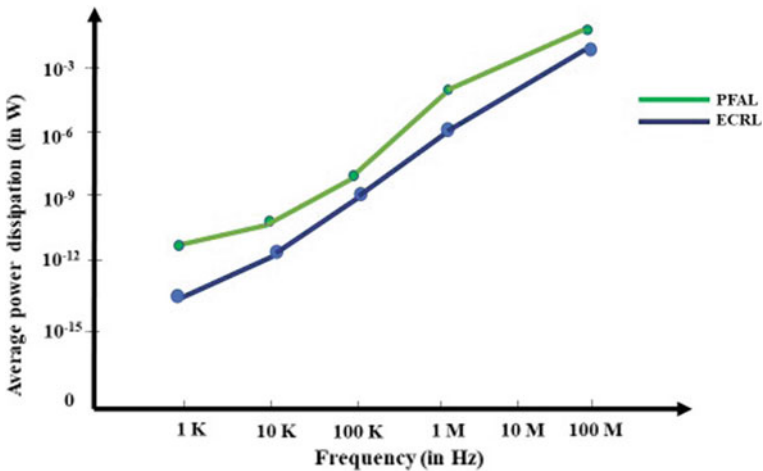


Fig. 8 Comparison of power dissipation between different adiabatic techniques at different frequencies for inverter circuit

indirectly proportional to the delay of the circuit, i.e., when the power is reduced, correspondingly delay will increase.

Figure 10 shows that the number of transistors used in adiabatic techniques is greater than that of CMOS technique [14, 15]. This increase in transistor count, i.e., area overhead in ECRL and PFAL, is compensated with reduction in power dissipation.

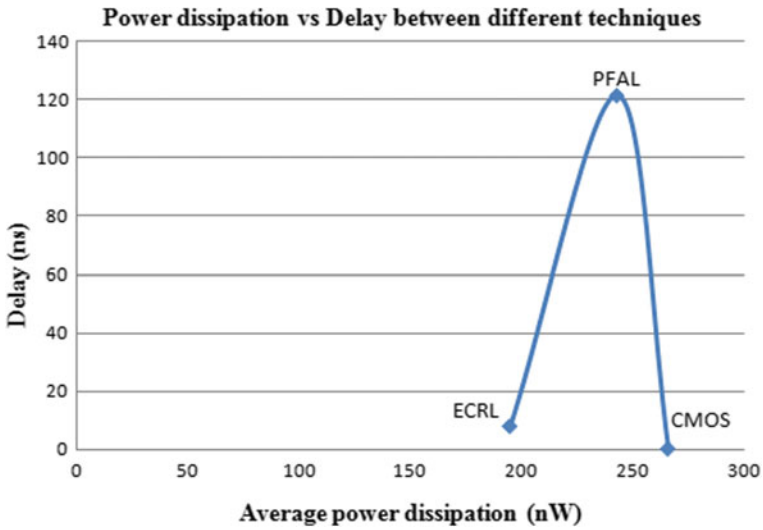


Fig. 9 Trade-off between average power and delay in CMOS and different adiabatic techniques at 100 kHz

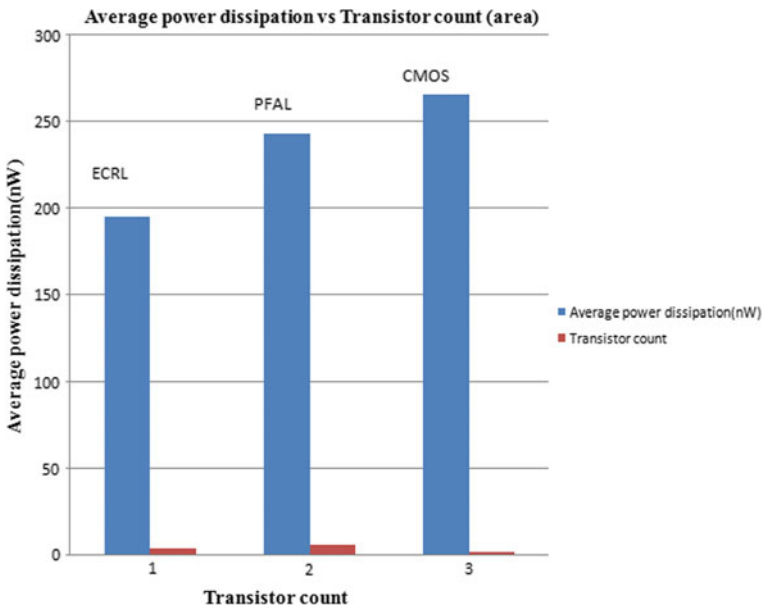


Fig. 10 Average power dissipation (nW) against transistor count (area)

## 5 Conclusion

From the implementation and simulation of various combinational circuits, we conclude that adiabatic logic circuit has more advantage over conventional CMOS circuit. The proposed work shows the simulation result of inverter, NAND, NOR, EXOR, 2:1 MUX, and full adder using conventional CMOS and ECRL, PFAL adiabatic techniques with different technology like 45, 90, and 180 nm. All the designed circuit with adiabatic technique shows significant power reduction over conventional CMOS. Inverter, NAND gate, NOR gate, EXOR gate, 2:1 MUX, and full adder circuits designed with ECRL techniques save 26%, 36%, 16%, 59%, 73%, and 99%, respectively, comparatively CMOS within specified frequency range of 1 kHz to 1 MHz. PFAL shows significant power reduction comparatively ECRL above certain frequency. The entire efficiency of adiabatic circuit largely depends upon the power clock efficiency. This work can be extended if the energy efficiency of resonant drivers for the power clock generation can be increased to drive adiabatic circuits, so that they can have a good scope to become an integral part of large digital systems. Research on MEMS-based resonators is being carried out by some researchers from the field of mechatronics.

## References

1. Sanadhya M, Kumar MV (2015) Recent development in efficient adiabatic logic circuits and power analysis with CMOS logic. *Procedia Comput Sci* 57:1299–1307
2. Keote ML, Karule PT (2015) Design and Implementation of energy efficient adiabatic ECRL and basic gates. In: *International conference on soft computing techniques and implementations (ICSCTI)*, pp 87–91
3. Srilakshmi K, Tilak AVN, Srinivasa Rao K (2016) Performance of FinFET based adiabatic logic circuits. In: *IEEE region 10 conference (TENCON)*, pp 2377–2382
4. Han M, Takahashi Y, Sekine T (2017) Low power adiabatic logic based on 2PC2AL. In: *IEEE international conference on IC design and technology (ICICDT)*, pp 1–4
5. Samanta S (2017) Sequential adiabatic logic for ultra-low power applications. In: *Devices for integrated circuit (DevIC)*, pp 821–824
6. Gupta Y, Sasamal TN (2015) Implementation of reversible logic gates using adiabatic logic. In: *IEEE power, communication and information technology conference (PCITC)*, pp 595–598
7. Teichmann P (2012) Adiabatic logic. In: *Springer series in advanced microelectronics*, vol 34, chapter 2, pp 9–10
8. Pindoo IA et al (2015) Power dissipation reduction using adiabatic logic techniques for CMOS inverter circuit. In: *Sixth IEEE international conference on computing communication and networking technologies (ICCCNT)*
9. Chaudhuri A et al (2015) Implementation of circuit in different adiabatic logic. In: *IEEE sponsored 2nd international conference on electronics and communication system (ICECS)*, pp 353–359
10. Teichmann P (2012) Fundamentals of adiabatic logic. In: *Springer series in advanced microelectronics*
11. Moon Y, Jeong D-K (1996) An efficient charge recovery logic circuit. *IEEE J Solid-State Circ* 31:514–522

12. Kumar A, Sharma M (2013) Design and analysis of MUX using adiabatic techniques ECRL and PFAL. In: IEEE international conference on advances in computing, communications and informatics (ICACCI), pp 1341–1345
13. Amirante E et al (2002) Adiabatic 4-bit adders: comparison of performance and robustness against technology parameter variations. In: Proceedings of the 2002 45th Midwest symposium on circuits and systems, vol III, pp 644–647
14. Kumar H, Srivastava S, Singh B (2019) Comparative analysis of 6T, 7T conventional CMOS and CNTFET based SRAM memory cell design. J Adv Sci Eng Med 11(1/2):3–10. <https://doi.org/10.1166/asem.2019.2301>
15. Saun S, Kumar H (2019) Design and performance analysis of 6T SRAM cell on different CMOS technologies with stability characterization. IOP Conf Ser: Mater Sci Eng 561(012093):1–9. <https://doi.org/10.1088/1757-899X/561/1/012093>

# A Review: Emerging Trends of Big Data in Higher Educational Institutions



**Raza Hasan, Sellappan Palaniappan, Salman Mahmood, Vikas Rao Naidu, Aparna Agarwal, Baldev Singh, Kamal Uddin Sarker, Ali Abbas and Mian Usman Sattar**

**Abstract** Universities/higher educational institutions are finding ways to increase the student-faculty interactions beyond the traditional classroom, helping institutions to gather the information to enhance the student learning experiences with the help of learning analytics. These interactions are captured using the virtual learning environment through which institutions learn from the student interactions and behavioral patterns within those systems. This helps the institutions for better retention rate, prediction of the results and focus on weak students. Many institutions have placed an early detection system for management and faculty to engage with the students and figure out the problems faced by the students and provide a remedy to improve

---

R. Hasan (✉) · S. Palaniappan · S. Mahmood  
Malaysia University of Science and Technology, Selangor, Malaysia  
e-mail: [raza.hasan@pg.must.edu.my](mailto:raza.hasan@pg.must.edu.my)

S. Palaniappan  
e-mail: [sell@must.edu.my](mailto:sell@must.edu.my)

S. Mahmood  
e-mail: [salmanm2015@yahoo.com](mailto:salmanm2015@yahoo.com)

V. R. Naidu · B. Singh  
Vivekananda Global University, Jaipur, India  
e-mail: [vikasrn@gmail.com](mailto:vikasrn@gmail.com)

B. Singh  
e-mail: [baldev.vit@gmail.com](mailto:baldev.vit@gmail.com)

A. Agarwal · K. U. Sarker · A. Abbas  
Department of Computing, Middle East College, Muscat, Oman  
e-mail: [aparna@mec.edu.om](mailto:aparna@mec.edu.om)

K. U. Sarker  
e-mail: [ku\\_sarker@yahoo.com](mailto:ku_sarker@yahoo.com)

A. Abbas  
e-mail: [aabbas@mec.edu.om](mailto:aabbas@mec.edu.om)

M. U. Sattar  
University of Management and Technology, Lahore, Pakistan  
e-mail: [usman.sattar@umt.edu.pk](mailto:usman.sattar@umt.edu.pk)

for the faculty members. Most of the institutions rely mainly on one system such as the learning management system to capture the student interactions thus creating a gap. The Internet gives an edge to its users for practicing, learning, by doing, this leads to the emergence of video-based learning technologies that are practiced and used in several ways, such as flipped classrooms. Student faces a doubt often in their phase of learning, to clear their doubts they refer to multiple sources to get the information and knowledge. These videos provide complete skill sets, students due to lack of skill set they use these sources for their specific problems. This paper discusses literature and background studies on the big data used in institutions of higher education. It establishes a framework based on the latest trends in this area that can help stakeholders to predict their business needs.

**Keywords** Learning analytics · Big data · Virtual learning environment · Learning management system · Education data mining · Prediction

## 1 Introduction

Learning is always a challenge for students and teachers in a monotonous environment in which a conventional/traditional teaching method is involved. Ubiquitous devices and smartphone culture are increasing among students, create an environment where knowledge flows in both directions and not in one way. Nowadays the technologies and infrastructure for the ubiquitous devices pave way to get benefit from these technologies in higher educational institutions (HEI) for better teaching and learning. This also creates an opportunity from making the classroom away from traditional to smart classrooms. Due to the fast-changing methods adopted by the institution for teaching and learning, learning management system (LMS) will not be enough to handle the needs of the learners. Video analytics (VA) along with the LMS and learning analytics (LA) can help the institutions with better analysis of their learners and decision making. This study focuses on proposing a framework based on the work that has been done in the field and how to integrate different technologies to get the maximum output. The aim of the study is to enhance the teacher and student education system. Learning analytics focuses on the data that came from the learning management system (LMS) and other technological systems [1].

## 2 Literature Review

### 2.1 Education

Social, economic and technological advances in response evolved the education systems. Education is progressing through society and the changes made by it. Education has gradually progressed and transformed from person-to-person learning system to

a more focused one-to-many system, providing additional opportunities to learners in catering them with the use of technology. ICT has played a vital role in changing the world rapidly from agricultural society to an industrial society that is named as Industry 4.0 era. Learners in response to these rapid developments and changes must acquire the necessary skills and ability to adapt to the social changes, redefining Industry 4.0 to Education 4.0.

Industry 1.0 was a completely agrarian society with limited tools and machinery. Informal methods of education which steadily established into schools, work distribution in this era were merely on the skills set possess by an individual. Education 1.0 responded to Industry 1.0 needs where knowledge is comprehensively transferred from the teacher to the learner (one-to-one). Industry 2.0 was with industrial revolution where energy resources were utilized, and improvements were made in working processes. Mass education has emerged (one-to-many) due to the invention of the printing press. This created the culture of scientific inquiry but lacks in creativity. Education 2.0 responded to the needs of Industry 2.0 where learning was more focused on learning the technologies and how to apply it in the work as a tool. Industry 3.0 was economically fast changing due to the communication networks and fast exchange of information, thus creating the world as a global village. Education 3.0 addressed the need for a technological society where technology supported for self-learning and creating knowledge. Interactive learning enhanced the student learning experience with a wide variety of Internet-enabled tools such as LMS. Various free and open-source tools have added further enhancement in the development of various teaching and learning aids by means of authoring and customization as per the requirements [2]. Industry 4.0 is blending new technologies creating innovation knowledge. This era caters to the needs of institutions with eLearning tools and creating new ways to advance delivery methods through virtual means of communications [3–6].

## ***2.2 Understanding Big Data***

In the era of education 4.0, HEIs are using innovative producing technology, eLearning tools, availability of virtual communication and advanced delivery methods. Creating an opportunity to gather a large volume of data not only about their learners and their educational system itself [6]. Different systems produce the information either structured or unstructured through the number of educational systems they use. Thus, creating a lot of attention from academia and the IT sector the need of big data. The information generating from these systems is at a rapid pace and exceeds the boundary to cope up with the needs and demands of a HEIs [7]. Student information systems (SIS) store the data, which in nature is in large amount; includes interactions between learners and different education technologies and in different databases [8, 9]. The data extracted from technology-based systems are useful for the collection, analysis and deduction of data that has attracted the attention of university administrators and academics, researchers and authorities. Implementing the streaming



server, which aims to aid the video learning mechanism as there is a lack of systems using technology which provide capturing and analysis of video interactions. Due to the high volume of raw data on students produced by these systems, the selection for analysis and the availability of qualitative data are created in a complex way [10]. Education data mining (EDM) is used to evaluate this information, which is widely used to analyze the performance of students [11].

### ***2.3 Educational Data Mining***

EDM is also referred to the discovery of Knowledge in Databases (KDD), finding new patterns that can be useful for analysis and concluding from HEI systems. The approaches that encourage KDD are divided into five categories which require analysis. Prediction refers to forecasting and understanding the behavior of the student. Consequently, the results of education draw from a single data aspect a combination of other related data. Classification, regression and density estimation are common techniques used for prediction. Clustering is to find the data set that can be grouped into complete data sets in other categories. Relationship, finding relationship between the number of variables in the data sets is determined. The association rules mining, casual data extraction, sequential pattern and correlation mining are the methods used in a relationship. Distillation recognizes and classifies the data characteristics for human inferences in the visualization form. Model discovery is a model validation technique, used in another analysis as a component. Model methods of discovery can be a prediction, clustering or mining relationships [12].

### ***2.4 Learning Analytics***

Analytics is referred to as data utilization, quantitative analysis, explanatory, predictive models and to use tools to deal with complex issues. LA deals with data analysis on the learner and their activities to enhance the student learning experience [13]. Implementing LA allows HEI to recognize their learners and the barriers to their learning thus ensuring institutional success retaining a large and diversified student population, operational facilities, fundraising and admissions. Student success is a key factor in accelerating academic institution resources management. LA measures, collects, analyzes and reports learner data, understand, optimize learning and the settings in which the learning takes place [14].

LA provides the opportunity not only for teacher to provide information to the students but assist the teacher to understand the behavior of the student in acquiring those skills and knowledge in a better manner. LMS is used as an aid in a suitable learning environment for the learning and teaching. LA becomes as important when the data is based on distributed sources [15]. LA plays a vital role in students learning

behavior and their perspective toward learning. LA relies on the digital footprint left behind by the students on the LMS; this data is available to the teacher and the administrator of the system.

Different theories have revealed that there exists a relationship between LA and LMS. Some suggest that learning takes place within the LMS systems by the students. However, this leads to low interaction as most LMS only caters to this type of data from different activities performed within the LMS such as quizzes and discussions some suggest that student behavior on the LMS can be useful for LA but this totally ignores the learning taken place outside the system.

LA aims to assist the teacher to use and improve the skills in teaching and learning. Providing different activities is required for student in their course of studies. Data from systems such as Moodle or SIS, which includes statistics on students who have been logged in time, assessment of course requirements by students, material reading time spent on the system, time taken to complete the activity, time taken to complete the module assessments, participation in the forum discussions and requests by students, but these are only a few [16].

## **2.5 Video Analytics**

Many platforms are available to cater the needs of visual learners such as YouTube. Video analytics with the help of streaming server can add more benefits alongside with LMS. It will provide stakeholders for better prediction of students, improving student's success factor and enhancing the learning and learners experience. Blended learning will boost the two analytics and be moving toward research-based learning. Real-time visualization using the dashboard helps improve the performance, effort and time of students by providing information on their learning trends and the knowledge they gain. Tools and technology give information quicker access to enable learners and educators to assess their progress, performance and ways of achieving better results. Therefore, the learner's goal is achieved. LA's advantage helps teachers enhance teaching and learning to understand students better [17]. LA also supports and assists educational institutions to achieve their student's performance, enhancing good knowledge is based on the forms of education tailed by the LA and data mining techniques in educational institutions [18]. The idea of using LA is to retain students. Providing personalized learning experiences, helping to predict the performance of learners, suggestions for learners increasing the reflection and awareness of learners with relevant learning resources, detecting unwanted learning behavior. Video learning analytics allows understanding and improving the effectiveness of video-based learning as a tool and use practices [18]. Exploring student interactions with video-based learning, extraction of valuable information about students and predicting their success based on their interactions.

### 3 Proposed Model

The literature review shows that two categories must be considered to predict the performance of students, namely academic information and student activity. Both academic information for students and student activities have a good relationship with VLE to understand student’s learning trends and patterns [19, 20]. It can be investigated the life cycle where the first step is to identify the datasets, after the process it can either be used for processing the datasets and apply the DM techniques which can later be used for visualization purposes. Furthermore, VLE can be used to gather student’s activity and behaviors by using LMS and VSS. This gives the opportunity for the stakeholders to understand the true picture in the learning process. This is an iterative process and output can be used as input for future decision making as shown in Fig. 1.

The proposed framework which an institution can use to apply business intelligence (BI) and DM used to forecast and predicted the problems they might face. As for HEI’s they normally have the student information system (SIS) which provides the student information, VLE such as Moodle can be used to get the student activity on the VLE and student behaviors can be acquired from the VSS. Classification is the most appropriate technique for predicting student performance and DM. Visualization can be used which is for all the stakeholders of the HEI’s.

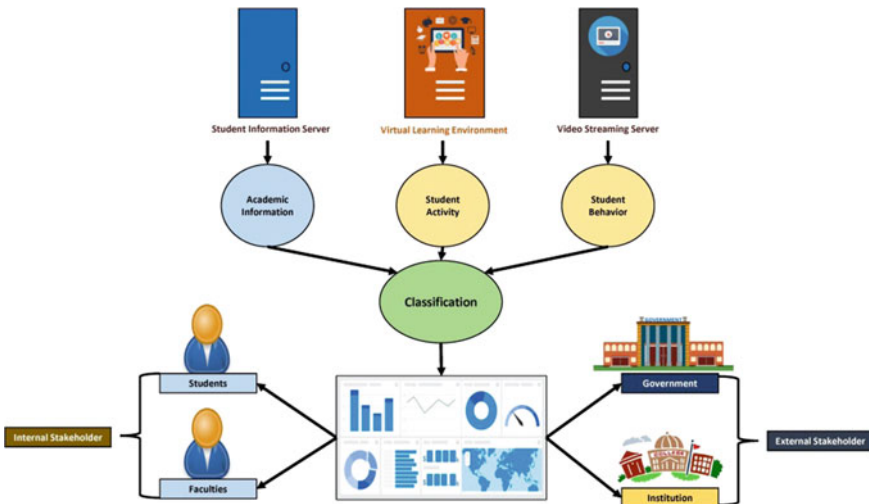


Fig. 1 Proposed framework

## 4 Implementation

Middle East College (MEC) is a private institution in the Sultanate of Oman, having around 5000 students enrolled in different disciplines of science and engineering. MEC uses SIS to store information related to students enrolled, Moodle as VLE for teaching and learning activities. Shifting from Education 3.0 to Education 4.0 integrating knowledge with innovation such as innovative eLearning tools is a need; college initiative for the flipped classroom was fruitful. The need of eLearning tool was missing which was solved by an in-house video streaming server. Delivery of video content over multiple connected devices solves the problem of delivering flipped content to the classroom. Figure 2 shows the Moodle where the link for the streaming videos is uploaded for students to go through before the commencement of the class. Figure 3 shows the streaming video from the streaming server which

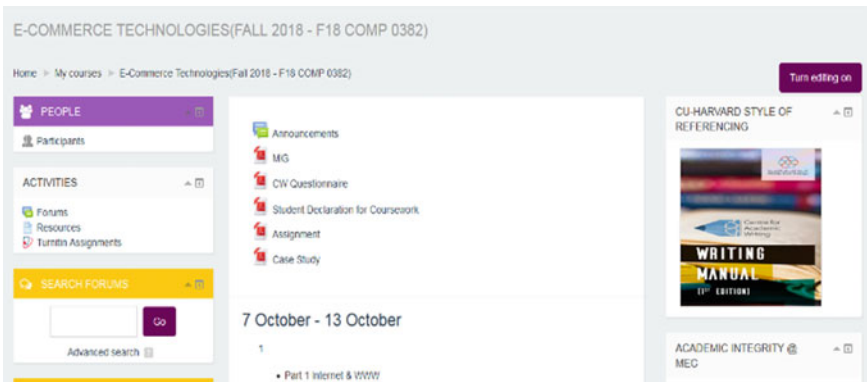


Fig. 2 Moodle course fall 2018



Fig. 3 Streaming video

is made by module leader teaching the module. The video can be accessed not only within the campus but outside the campus. Thus, giving an opportunity to students not only they can access within the campus but also at home or while traveling.

## 5 Conclusion

Big data along with the technological tools, it can have not only for prediction of students performance but also forecasting the course beforehand the commencement of the classes. These possibilities aid to provide a better learning environment for the students and for teachers a plan to conduct the lectures based on the learners. Blended or flipped learning can take place which enables students to better understand the concept not only in the class but beyond the class. This will equip students prepared for the automated world [21]. With the help of this framework, it will meet the requirements of the future.

## 6 Future Works

The need for the Education 4.0 era is to integrate these technologies on a common platform. SIS data is structured data in nature, Moodle data is structured in nature and VSS data is unstructured in nature. Integrating these on a common platform with the use of big data can help the educators to better understand the learners. Importing the data in real time and analysis on the go is a challenging task. Efficiency and efficacy of the system will be tested before moving ahead.

## References

1. Hasan R, Palaniappan S, Raziff ARA, Mahmood S, Sarker KU (2018) Student academic performance prediction by using decision tree algorithm. In: 2018 4th international conference on computer and information sciences (ICCOINS), Kuala Lumpur, Malaysia, 2018, pp 1–5
2. Naidu VR, Singh B, Al Harrasi RA, Al Balushi HH (2017) Technology enhanced learning assisted by free and open source software. *IJAEDU Int E J Adv Educ* 447–452. <http://dergipark.gov.tr/doi/10.18768/ijaedu.338515>
3. Baena F, Guarín A, Mora J, Sauza J, Retat S (2017) Learning factory: the path to industry 4.0. *Procedia Manuf* 9:73–80
4. Puncreobutr DV (2016) Education 4.0: new challenge of learning. *St. Theresa J Hum Soc Sci* 2(2):93–97
5. Arthur AMH (2008) Leapfrog principles and practices: core components of education 3.0 and 4.0. *Futures Res Q* 21(1):19–31
6. Naidu VR, Balushi HA, Bhatia S (2017) Effectiveness of free & open source tools to enhance game based learning experience in school education. In: International conference on education and new learning technologies

7. Ben D (2015) Big data and analytics in higher education: opportunities and challenges. *Br J Educ Technol* 46(5):904–920
8. Khan N, Yaqoob I, Hashem IAT, Inayat Z, Ali WKM, Alam M, Shiraz M, Gani A (2014) Big data: survey, technologies, opportunities, and challenges. *Sci World J* 18
9. Atif A, Bilgin RDA, Marrone M (2013) A panorama of learning analytics featuring the technologies for the learning and teaching domain. In: *Electric dreams—2013 ASCILITE conference*, Sydney
10. Tair MA, El-Halees A (2012) Mining educational data to improve students' performance: a case study. *Int J Inf Commun Technol Res* 2(2):140–146
11. Qasem M, Qaddoura R, Hammo B (2017) Educational data mining (EDM): a review. *New Trends Inf Technol*
12. Shahir AM, Husain W, Rashid NA (2015) A review on predicting student's performance using data mining techniques. *Procedia Comput Sci* 72:414–422
13. Baker RS (2010) Data mining for education. *Int Encycl Educ* 7:112–118
14. Prakash BR, Hanumanthappa D, Kavitha V (2014) Big data in educational data mining and learning analytics. *Int J Innov Res Comput* 2(12):7515–7520
15. Siemens G, Gasevic D (2012) Guest editorial-learning and knowledge analytics. *J Educ Technol Soc* 15(3):1–2
16. Chatti M, Schroeder U, Jarke M (2012) Convergence of knowledge management and technology-enhanced learning. *Trans Learn Technol* 5:177–189
17. Eckerson W (2011) *Performance dashboards*, 1st edn. Wiley, New York
18. Hadhrami G (2017) Learning analytics dashboard to improve students' performance and success. *IOSR J Res Method Educ (IOSRJRME)* 7(1):39–45
19. Giannakos M, Chorianopoulos K, Chrisochoides N (2015) Making sense of video analytics: lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course. *Int Rev Res Open Distrib Learn* 16(1)
20. Hasan R, Ali S, Hayat M (2015) Enhancing student's learning experience at middle east college by using blended learning. In: *Science and information conference (SAI)*, UK
21. Kazmi SIA, Hayat MS, Hasan R, Dattana V (2017) Network simulation tool enhances learning and understanding of computer network protocol concepts in middle east college. *Airo Int Res J* 13

# Realization of a Dual Stop Band for ‘S’ and ‘X’ Bands From the Complimentary Geometry of the Dual Pass Band (S and X Band Application) FSSs



Amanpreet Kaur and Komalpreet Kaur

**Abstract** Interference from the nearby wireless communication bands may weaken the desire frequency signals. Hence, to overcome or minimize the problem of interference from nearby wireless communication bands, a dual stop band unit cell structure of FSS that realized from the pass band geometry of FSSs, is proposed. Complimentary structure of the FSS unit cell offers transmission characteristics which are complimentary of the original transmission characteristics. For example original FSS unit cell have single square shaped structure that imprinted on substrate offers stop band transmission characteristics. In addition complimentary of the square loop that is square shaped slot in substrate offers pass band transmission characteristics. Therefore, the stop band transmission characteristics from pass band can be realized by simply replacing the metal elements of FSS unit cell structure by slots/aperture of the same size and geometry (i.e. complimentary of the original FSS unit cell structure). But here, it is important to mention that the resultant frequency response of the complimentary structure will not be exactly opposite to that of frequency response of original structure. Therefore, the FSS unit cell is optimized to get the desired stop band response for S and X bands.

**Keywords** Frequency-selective surfaces (FSS) · Transmission parameters · Complimentary structure · Angle of incidence (AOI) · Polarization mode

## 1 Introduction

Wireless communication technology has covered many progressive paths and has upgraded performance and proficiency in the communication environment in terms of removal of unwanted signal from the desired ones [1, 2]. To overcome or minimize the problem of interference of the unwanted signals from nearby frequency bands,

---

A. Kaur (✉) · K. Kaur  
Thapar Institute of Engineering and Technology, Patiala, India  
e-mail: [amanpreet.kaur@thapar.edu](mailto:amanpreet.kaur@thapar.edu)

K. Kaur  
e-mail: [preet.komal0094@gmail.com](mailto:preet.komal0094@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_30](https://doi.org/10.1007/978-981-15-2329-8_30)

microwave special filters have been proposed that are called FSS [3]. FSS allows the specific frequency bands to pass through and block or stop the unwanted electromagnetic signals from unwanted frequency band. FSSs are also used to diminish the volume and expand the capability of anti-jamming of some multiband electronic systems [4]. FSS is an array of periodically arranged metallic patches or apertures on the dielectric substrate that shows total reflection and transmission, respectively [5, 6]. However, the transmission response of the FSS structures are basically a function of frequency, angle of incidence (AOI) [7] and polarization [8] of the incoming waves. In addition to this, the inter-element spacing, periodicity, size and geometry of the unit cell structure of FSS also affect its performance [9]. It is always desirable to use FSS structures with multi-resonance bands so as to have their employability for multiple wireless applications at the same time. There are a number of different geometries of FSS unit cell structures which have been introduced by different researchers such as fractal geometry-based FSS structure [10], double [11] and multiband FSS structures and multilayer [12] FSS structures. It is always appreciable to design a FSS structure which offers independency for both polarization [13] and angle of incident. Their inherent features that result from their specific design make them a potential candidate to use as spatial filters in a variety of microwave applications such as radomes and satellite communication. Therefore, with the purpose of design a FSS unit cell structure with stop band characteristics for the 'S' and 'X' wireless communication bands; The proposed FSS unit cell structure (flower shaped design with outer square loop) offers dual stop band frequency response for both S and X bands.

## 2 Unit Cell Design

In this section, the design procedure and optimization of the complimentary FSS unit cell structure for dual pass band to realize the dual stop band characteristics for S and X band is presented. Step-by-step changes in the complimentary structure to obtain the dual stop band frequency response are shown in Fig. 1.

Figure 1 depicts the different steps for the realization of required stop band FSS from the pass band geometry of FSS as depicted in Fig. 1a. The steps include replacing the metal part of the pass band unit cell element with slot/aperture and slotted part with metal one to get the complimentary structure as depicted in Fig. 1b. For obtaining the required stop band transmission response, the centre square of the complimentary unit cell (b) is made wider as depicted in Fig. 1c. In addition to it, a wider slot is cut from centre square that offers better results for X band without affecting the S band transmission response. Then, one single square loop element on the back side of the substrate is added that increases the bandwidth of both S and X wireless bands, as depicted in Fig. 1d, e, respectively. Finally, one small patch of dimensions  $0.5 \text{ mm} \times 0.5 \text{ mm}$  is added at the centre of the square slot that offers smoothed transmission response for both S and X bands. Final optimized FSS unit cell for stop band characteristics is also depicted in Fig. 2. The optimized unit cell metallic elements are imprinted on the FR4 dielectrics substrate with thickness 1.57 mm and



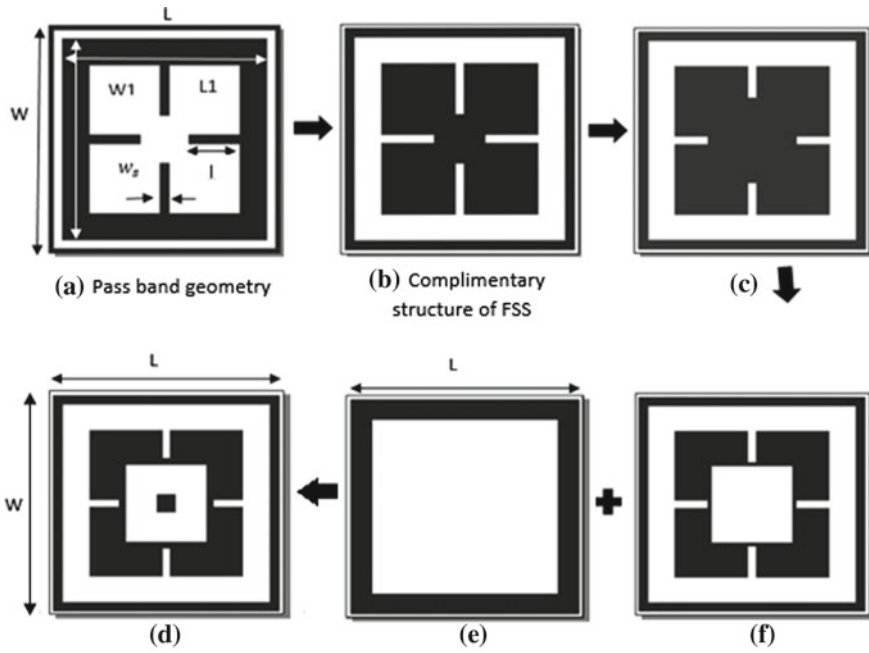


Fig. 1 Step-by-step development of the final unit cell structure of FSS

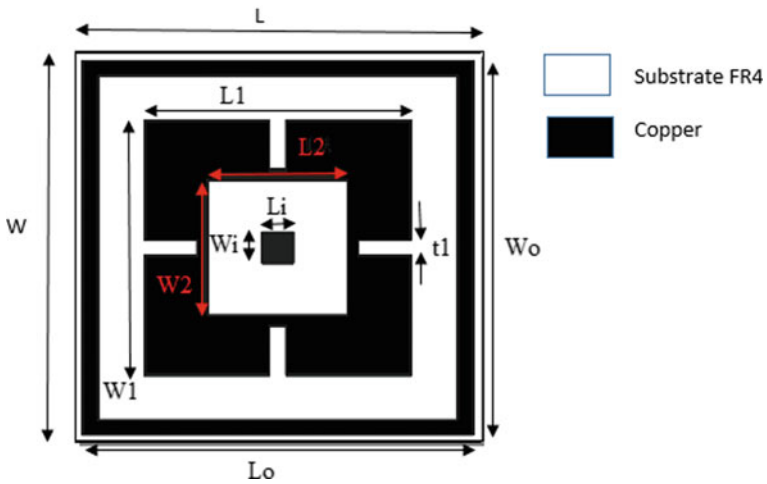


Fig. 2 Top view of the unit cell structure of FSSs

**Table 1** Specified and optimized parameters of unit cell of FSSs

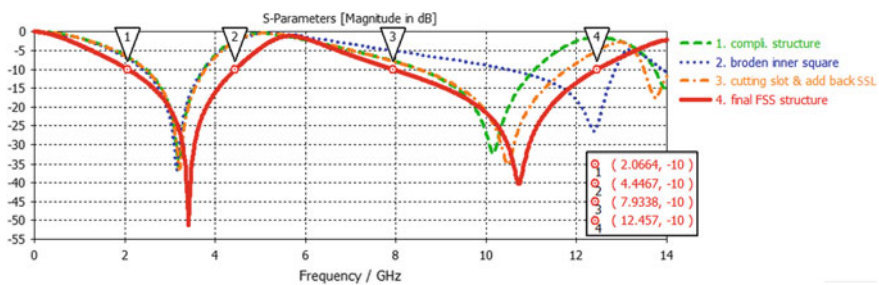
Specification	Dimension (mm)
Thickness of outer loop	0.25
Length of flower-shaped structure (L1)	8.44
Width of flower-shaped structure (W1)	8.44
Slot thickness (t1)	0.44
Length of inner patch (Li)	1
Width of inner patch (Wi)	1
Thickness of substrate	1.57
Thickness of copper	0.035

dimensions ( $L \times W$ )  $13 \times 13 \text{ mm}^2$  and dielectric constant 4.4. Outer metallic square loop ( $L_o \times W_o$ ) having dimension of  $12.5 \text{ mm} \times 12.5 \text{ mm}$ . In addition to this, the optimized parameters of the unit cell structure of FSS are mentioned in Table 1.

### 3 Simulation Results

The transmission characteristics of the required stop band have been shown in Fig. 3. From here, it is observed that to realize the band stop transmission characteristics from pass band, unit cell of FSSs has to go through different stages, as shown in Fig. 1, where every parameter of the unit cell structure affects the transmission characteristics. Figure 3 shows the first stage complimentary structure of the unit cell that is shown in Fig. 1a.

It is seen that the transmission characteristics are complimentary of the pass band but do not cover the entire S and X bands. Therefore, optimized complimentary structure of FSS is obtained for stop band characteristics that covers the entire S and X bands. At the second stage, the broadened centre square of the inner flower-shaped structure gives the transmission characteristics that do not alter the stop band response for ‘S’ band but shifts the resonance frequency of ‘X’ band towards the



**Fig. 3** Transmission characteristics (stop band) of FSSs

higher frequency as shown in Fig. 3 (dotted blue curve). At the third stage, a slot is cut from the inner square, and single square loop of metal pattern on the back side of substrate is added. It is seen that it again does not alter the ‘S’ band frequency response but broadens the bandwidth of ‘X’ band frequency response. With the addition of square patch at the center of FSS unit cell, the bandwidth becomes wider and offers the transmission stop band characteristics for the entire S(2.06–4.44 GHz) and X (7.9–12.4 GHz) bands.

### 3.1 Parametric Study

To obtain the required stop band characteristics from complimentary structure of the proposed pass band FSS structure, various parameters of the unit cell structure of FSS such as length (L1) of inner flower-shaped structure, its slots thickness (t1), length (L2) of the inner slot and small inner patch (Li) have been optimized, and effect of each parameter on the performance has been explained in this section.

**Variation in the length of inner flower-shaped structure.** Figure 4 depicts the effect of variation in length of the flower-shaped structure of FSS unit cell element.

It is observed that smaller length (L1) of inner flower-shaped structure offers more effects on the S band of the transmission characteristics without much affecting the X band. As the length L1 goes on increasing from 8.42 to 8.46 mm, change in the transmission characteristics of both S and X band is observed. It is observed that best results are found at L1 = 8.44 mm and is therefore selected.

**Variation in the length (Li) of the small inner metallic patch of FSS structure.** From Fig. 5, it is observed that the inner patch affect the S band frequency response more without much affecting the frequency response of X band. Best results are found at Li = 1 mm and is therefore a selected parameter.

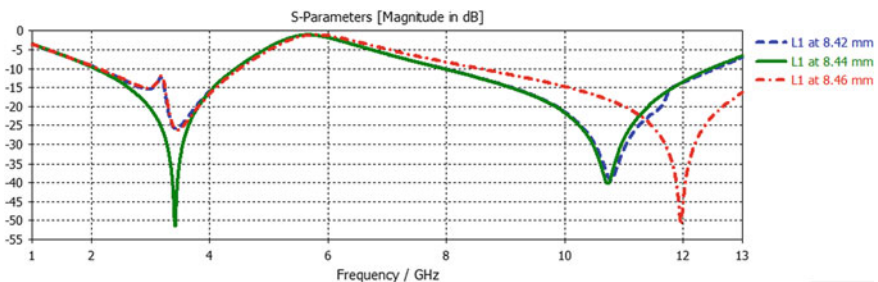
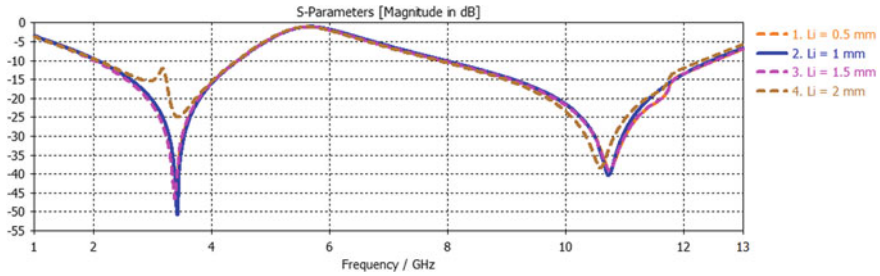


Fig. 4 Variation in the length of inner flower-shaped structure



**Fig. 5** Variation in the length ( $L_i$ ) of the small inner metallic patch of FSS structure

## 4 Conclusion

In this article, a complimentary and optimized structure of dual pass band, i.e. S (1.44–4.64 GHz) and X (6.32–12.17 GHz) is proposed that offers dual stop band transmission characteristics for S (2.06–4.4 GHz) and X (7.93–12.45 GHz) bands. From the simulated results, it is verified that the outer square loop of the unit cell structure excites the S band while inner flower-shaped geometry excites the stop band for X band applications. In addition to this, the inner small metallic patch is responsible for overall smoothness and enhancement in bandwidth. A single dielectric substrate (FR-4) has been used to generate the dual band (S and X band) characteristics. The proposed FSS structure is miniaturized in size and have simple geometry that is easy to fabricate. It can be used in many potential applications such as military applications, defence, wireless communications, air surveillance radar and many other S and X band applications.

## References

1. Rappaport TS (2012) *Wireless communication and practice*. 2nd edn
2. Goldsmith A (2005) *Wireless communications*. Cambridge University Press
3. Li W, Suo Y, Chen J-Z, Chou H-T (2017) A FSS of hybrid combined elements for dual-band operations. *IEICE Electron Expr* 14:1–6
4. Schennum GH (1973) Frequency Selective surfaces for multiple frequency antennas. *Microwave J* 16:55–57
5. Yahya R, Nakamura A, Itami M (2016) Compact UWB frequency selective surfaces with angular stability. *IEICE Electron Expr* 5:39–43
6. Munk BA (2000) *Frequency selective surface—theory and design*. Wiley, New York
7. Zhao Z, Li W, Yang Y, Huo H, Li J, Zhang A (2017) An incident angle insensitive band-pass frequency selective surface with flat top. *IEEE* 273–274
8. Orr R, Fusco V, Zelenchuk D, Goussetis G, Saenz E, Simeoni M, Drioli S (2015) Circular polarisation frequency selective surface operating in Ku and Ka band. *IEEE Trans Antennas Propag* 5194–5197
9. Wu TK (1995) *Frequency selective surfaces and grid array*. Wiley, New York
10. Zhang JC, Yin YZ, Ma JP (2009) Frequency selective surfaces with fractal four legged elements. *Progress Electromagnet Res Lett* 8:1–8

11. Zhao Z-Z, Li W, Chen X-M (2018) Broad-band angle insensitive frequency-selective surface with multiple resonant modes. *Microw Opt Technol Lett* 60:2660–2664
12. Moy-Li HC, Sánchez- Escuderos D, Antonino-Daviu E, Ferrando- Bataller M (2019) Dual-polarized planar lens antenna designed with a quad-ridged frequency selective surface. *Microw Opt Technol Lett* 61:479–484
13. Yadav S, Jain CP, Sharma MM (2018) Polarization independent dual-band pass frequency selective surface for Wi-Max applications. *Int J RF Microw Comput Aided Eng* 28:1–7

# Defect Prediction of Cross Projects Using PCA and Ensemble Learning Approach



Lipika Goel, Mayank Sharma, Sunil Kumar Khatri and D. Damodaran

**Abstract** Cross-project defect prediction (CPDP) is a technique of detecting defects in software modules in which the training and the testing projects for the classification model are different. The effective prediction leads to a more reliable software. The merging of dataset from varying sources results to an imbalanced dataset. The complex structure and the imbalance data make it a challenge for an effective cross-project defect prediction. To overcome these issues, in this paper, we propose a cross-project defect prediction framework. In the first stage of this framework, PCA is applied for dimensionality reduction of the dataset into two components. In the second phase, SMOTE technique of data sampling is applied to handle the class imbalance problem. Then the ensemble classifiers random forest and XGBoost are applied for an effective defect-prediction model. We have conducted the experiments on eight open source software projects. The results are compared with few baseline techniques. The results indicate that the proposed framework gave comparable performance of cross-project defect prediction to some baseline methods.

**Keywords** SMOTE · Cross-project defect prediction (CPDP) · Ensemble learning · Within-project defect prediction (WPDP) · PCA · Class imbalance learning

---

L. Goel (✉)  
Amity University, Noida, India  
e-mail: [lipika.bose@gmail.com](mailto:lipika.bose@gmail.com)

AKG Engineering College, Ghaziabad, India

M. Sharma · S. K. Khatri  
Amity Institute of Information Technology Amity University, Noida, India  
e-mail: [msharma22@amity.edu](mailto:msharma22@amity.edu)

S. K. Khatri  
e-mail: [skkhatri@amity.edu](mailto:skkhatri@amity.edu)

D. Damodaran  
Centre for Reliability, Thiruvananthapuram, Chennai, India  
e-mail: [damodarand@gmail.com](mailto:damodarand@gmail.com)

## 1 Introduction

In software industry, the list of software modules is prioritized for testing. This limits the testing resources required and detects the higher number of defects with minimal efforts [1]. The quality and the reliability of the software are enhanced, thereby improving the software maturity and process [2]. In the past few decades, much of the study is done on within-project defect prediction. In WPDP, the defect-prediction model is trained on the old releases of the same project [3]. In many practical circumstances, data of the old releases of the projects are unavailable. This leads to the motivation for research of CPDP [4].

In CPDP, the projects which do not have historical data of releases can use the data from the other projects to build the defect-prediction model. In this, the training and the testing are done on two different projects [5]. The limitation of WPDP can be overcome by CPDP. CPDP is now much under study in this decade, because this paper proposes a framework for CPDP. In this, PCA is applied for dimensionality reduction of the dataset into two components. In the second phase, SMOTE technique of data sampling is applied to handle the class imbalance problem. Then the ensemble classifiers random forest and XGBoost are applied for an effective defect-prediction model.

The main objectives of this paper are to infer the following:

- RQ1. Whether principal component analysis method of feature extraction effective in CPDP?
- RQ2. How does CPDP perform after suitable proposed framework as compared traditional defect-prediction process?

The paper is divided into six sections. The literature survey is summarized in Sect. 2. The brief about the datasets, dimensionality reduction, data sampling techniques, the performance measures used in the experiments, and the models are stated in Sect. 3. Section 4 describes the proposed methodology. The results are analyzed and discussed in the Sect. 5. The conclusion is stated in Sect. 6.

## 2 Literature Review

Turhan et al. [6] implemented nearest neighbor filter for reducing the difference in distribution of the data. He then applied NB classifier for prediction of defects. Zimmermann et al. [7] gave a novel direction to the research in this field. He proposed cross-project defect prediction and used decision tree and logistic regression for defect prediction.

Ma et al. [8] proposed new approach of transfer Naïve Bayes (TNB) algorithm. He worked on homogeneous set of cross project metric and applied TNB. The results showed comparable results to WPDP. Nam et al. [9] further in the direction of transfer learning also proposed a novel approach named as transfer component analysis

(TCA) for dimensionality reduction and data distribution. The results outperformed compared to other baselines. Herbold [10] commented that training data set selection plays an important role in CPDP. He used strategy based on distance and NN filtering for selection of the training data. Jing et al. [11] gave new dimensions to CPDP. Nam et al. [12] proposed that when two projects have different set of metrics, then it is termed as heterogeneous projects. The prediction on heterogeneous projects is called heterogeneous defect prediction. He proposed a novel metric matching and selection approach for selecting the set of features. Lin et al. [13] in 2015 stated that negative samples reduce the performance of the model. He proposed a double transfer learning approach to eliminate the negative samples, thereby enhancing the performance of the model.

The difference of defective and non-defective classes leads to the problem of class imbalance learning. Ryu et al. [14] in 2016 proposed a novel combination of cognitive boosting and SVM to reduce the class imbalance learning problem. Zhang et al. [15] in 2016 focused on heterogeneous metrics, thereby performed unsupervised learning. Xia et al. [16] proposed novel-HYDRA model (hybrid model reconstruction approach) for defect prediction in cross projects. Herbold et al. [17] investigated the cross-project defect prediction approaches.

### 3 Imperative Knowledge

#### 3.1 *Description of Metrics and Dataset*

The datasets used in this paper are <http://openscience.us> and [www.Promisedata.org](http://www.Promisedata.org). We have selected the projects with homogeneous (same) set of metrics, therefore, performing homogeneous CPDP.

#### 3.2 *Dimensionality Reduction*

The technique of reducing the dimension of the feature space is known as dimensionality reduction technique. The principal component analysis is a technique of feature extraction. In feature extraction, we create new independent variables by combining the old independent variables. These new independent variables can predict the dependent variable. PCA also combines the input variables by eliminating the least important variables and retains the most vital and important parts of all the variables. In PCA, each of the new variables will be independent of each other.

#### **Data Sampling Technique**

In this paper, we have used SMOTE for sampling of the data for handling the class imbalance problem. SMOTE [18]—SMOTE is oversampling method of sampling.



The new occurrences of the class with the minority instances are created. Instead of repetition of the data, new synthetic instances are created. This leads to the balanced data. It handles the issue of over-fitting.

### ***3.3 Ensemble Learning Modeling Approach***

Among the various ensemble learning models, we have used random forest and XGBoost for classification. Random forest is a bagging algorithm [19], whereas XGBoost is a boosting approach. Random forest is a bagging technique that is applied to decision trees. It is an ensemble of weak decision tree models. To classify a class each tree votes, the forest chooses the classification which has the highest number of votes XGBoost is an implementation of gradient boost decision trees. It is known as extreme gradient boosting. The performance of the model and the execution speed are improved using XGBoost classifier.

### ***3.4 Performance Measures***

To determine the effectiveness of the model performance measures are used.

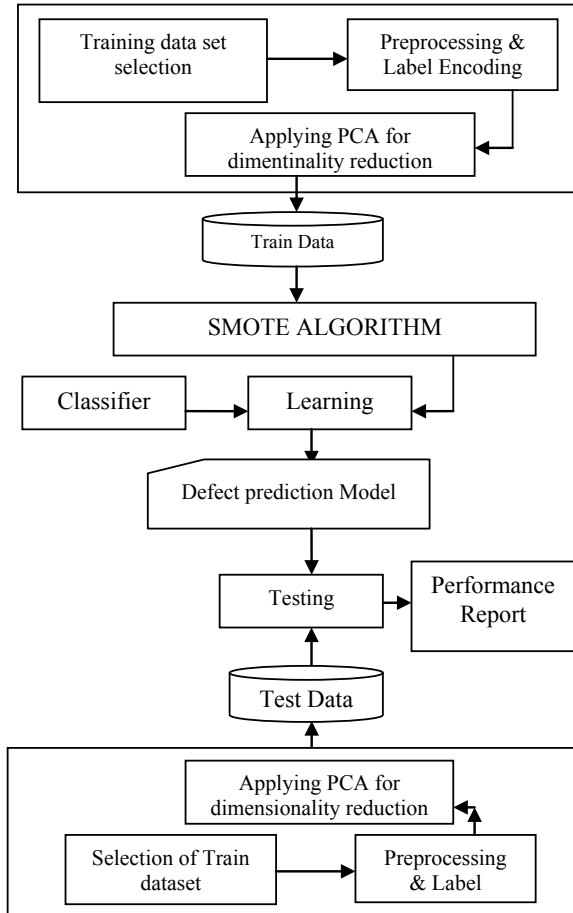
The following are the parameters of the confusion matrix [20]:

- TP: true specimen forecasted as true.
- FP: false specimen forecasted as true.
- TN: false specimen forecasted as false.
- FN: true specimen forecasted as false.

## **4 Proposed Methodology**

The experiment is conducted in twofold: Whether the PCA technique of feature extraction is effective in CPDP? Secondly, whether CPDP perform after suitable proposed framework as compared traditional WPDP? The proposed methodology of the framework is in Fig. 1. Two experiments are conducted to answer the above-stated questions.

**Fig. 1** Proposed model



**4.1 Experiment 1**

For varying combinations of the dataset, the model is trained. The model is trained with imbalanced dataset. Random forest and XGBoost are used as classifiers for classification. Various performance measures were evaluated.

**4.2 Experiment 2**

For all the different combinations of the dataset, the model is trained. After preprocessing of the datasets, PCA was employed for dimensionality reduction into two components. Then, the SMOTE is implemented to handle class imbalance nature of

the dataset. Random forest and XGBoost are used for classification. The performance of each classifier is computed.

## 5 Results

In the result section, the results acquired from the experiments are tabulated. It is tabulated in Tables 1, 2 and 3. Table 1 summarizes the results with no dimensionality reduction method and no sampling method being implemented to train data. Random forest and XGBoost are used as a classifier to the cross projects. Table 2 tabulates the results using the proposed framework. Table 3 tabulates the results with no dimensionality reduction technique and no data sampling technique in WPDP.

**Table 1** Results of CPDP without PCA and SMOTE-experiment 1

Classifier	Project	Accuracy	Precision	Recall	F-score	ROC value
Random forest	ivy-2.0	0.81	0.76	0.82	0.77	0.523
	ant-1.7	0.85	0.82	0.86	0.84	0.541
	camel 1.4	0.78	0.75	0.79	0.73	0.554
	jedit-4.1	0.86	0.78	0.76	0.77	0.52
	camel 1.6	0.81	0.86	0.63	0.7	0.651
XGBoost	ivy-2.0	0.74	0.72	0.8	0.75	0.507
	ant-1.7	0.88	0.83	0.88	0.84	0.529
	camel 1.4	0.78	0.77	0.79	0.72	0.544
	jedit-4.1	0.88	0.78	0.79	0.78	0.504
	camel 1.6	0.78	0.86	0.67	0.73	0.673

**Table 2** Results of CPDP using the proposed framework

Classifier	Project	Accuracy	Precision	Recall	F-score	ROC value
Random forest	ivy-2.0	0.82	0.78	0.82	0.79	0.574
	ant-1.7	0.86	0.83	0.87	0.84	0.544
	camel 1.4	0.91	0.81	0.81	0.8	0.624
	jedit-4.1	0.87	0.8	0.88	0.85	0.614
	camel 1.6	0.84	0.86	0.84	0.85	0.671
XGBoost	ivy-2.0	0.81	0.77	0.75	0.76	0.598
	ant-1.7	0.88	0.86	0.89	0.86	0.587
	camel 1.4	0.82	0.85	0.85	0.83	0.695
	jedit-4.1	0.87	0.79	0.78	0.78	0.558
	camel 1.6	0.81	0.87	0.83	0.84	0.695

**Table 3** Results of WPDP without PCA and SMOTE

Classifier	Project	Accuracy	Precision	Recall	F-score	ROC value
Random forest	ivy-2.0	0.83	0.79	0.83	0.81	0.678
	ant-1.7	0.87	0.84	0.88	0.86	0.556
	camel 1.4	0.93	0.83	0.85	0.84	0.621
	jedit-4.1	0.86	0.82	0.89	0.85	0.612
	camel 1.6	0.86	0.85	0.88	0.86	0.667
XGBoost	ivy-2.0	0.83	0.78	0.76	0.77	0.589
	ant-1.7	0.87	0.85	0.91	0.88	0.599
	camel 1.4	0.81	0.86	0.88	0.87	0.693
	jedit-4.1	0.87	0.78	0.79	0.78	0.598
	camel 1.6	0.83	0.88	0.85	0.87	0.699

The recall, accuracy, F-measure, and precision are used for evaluation of the model. The explanation to above-stated research questionnaire is as below:

*RQ1. Whether the PCA technique of feature extraction is effective in CPDP?*

Ans: From Tables 1 and 2, the average accuracy for cross-project defect prediction without using PCA and SMOTE is lower by 4.62% than the accuracy value using the proposed framework. The precision and recall also showed similar results. The precision is higher by 4.54% and recall by 4.32% using the proposed framework. The results improved for F-measure. The F-score is improved by approx. 4.5% when using the proposed framework. It can be concluded that PCA along with SMOTE as proposed in the framework enhances the forecasting ability of the model.

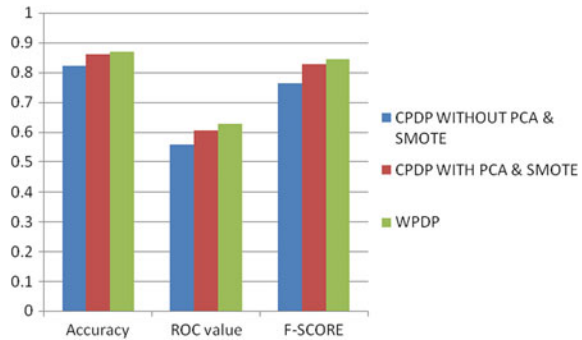
*RQ2. How does CPDP perform after suitable proposed framework as compared traditional WPDP?*

Ans: From Tables 2 and 3, the following inferences can be made:

- 0.91 is the highest accuracy achieved for CPDP using the proposed framework. The value is proportionate to the highest accuracy of 0.93 obtained for within-project defect prediction.
- The highest precision is 0.86 for CPDP which is almost equivalent to the value of 0.88 for within-project defect prediction.
- The Recall was as high as 0.91 for within-project defect prediction. The highest recall in the case of CPDP is 0.89, which is proportionate to within-project defect prediction.
- The F1-score has the highest value of 0.86 for cross-project defect prediction. This is approximately close to 0.88 achieved for within-project defect prediction.

The above considerations infer that cross-project defect prediction using the proposed framework is comparable to within-project defect prediction (Fig. 2).

**Fig. 2** Comparative performance measures of average accuracy, ROC and F-score



## 6 Conclusion

CPDP is a popular research domain in software reliability. The limit in the availability of historical data of the projects leads to the motivation of the cross projects. Many dimensionality reduction techniques have been proposed. This paper focuses on PCA for feature extraction. The class imbalance problem degrades the performance of the CPDP predictive model. In the paper, SMOTE technique for over sampling is implemented. An analysis is done to infer whether dimensionality reduction technique can improve the predictive performance in CPDP. Besides this, the results of traditional WPDP were compared to the proposed framework. Sixteen open source object-oriented projects were used to conduct the experiments. The results concluded, the principal component analysis technique of dimensionality reduction improved the performance of the classifier. Besides, these results also inferred that the proposed framework which used PCA and SMOTE technique gave comparable results to WPDP.

In the future, validation of the conclusion can be done on different real-time data using various other projects. Furthermore, validation of the results on projects with different metrics, i.e. heterogeneous cross projects can also be focused. A novel framework with improved dimensionality reduction approaches can be implemented for more application domains.

## References

1. Caglayan B, Misirli AT, Bener AB, Miransky A (2015) Predicting defective modules in different test phases. *Softw Quality J* 23(2):205–227
2. Monden A et al (2013) Assessing the cost effectiveness of fault prediction in acceptance testing. *IEEE Trans Softw Eng* 39(10):1345–1357
3. Hall T, Beecham S, Bowes D, Gray D, Counsell S (2012) A systematic literature review on fault prediction performance in software engineering. *IEEE Trans Softw Eng* 38(6):1276–1304
4. Kitchenham BA, Mendes E, Travassos GH (2007) Cross versus within company cost estimation studies: systematic review. *IEEE Trans Softw Eng* 33(5):316329

5. Anushree A, Malhotra R (2019) Cross project defect prediction for open source software. *Int J Inf Technol* 1–15. <https://doi.org/10.1007/s41870-019-00299-6>
6. Turhan B, Menzies T, Bener AB, Di Stefano J (2009) On the relative value of cross-company and within-company data for defect prediction. *Empirical Software Eng* 14(5):540–578
7. Zimmermann T, Nagappan N, Gall H, Giger E, Murphy B (2009) Cross-project defect prediction. In: Proceedings of the 7th joint meeting of the european software engineering conference and the ACM SIGSOFT symposium on the foundations of software engineering (ESEC/FSE), pp 91–100
8. Ma Y, Luo G, Zeng X, Chen A (2012) Transfer learning for cross-company software defect prediction. *Inf Softw Technol* 54(3):248–256
9. Nam J, Pan SJ, Kim S (2013) Transfer defect learning. In: Proceedings of the 35th international conference on software engineering (ICSE), pp 382–391
10. Herbold S (2013) Training data selection for cross-project defect prediction. In: Proceedings of the 9th international conference on predictive models in software engineering (PROMISE), pp 6–16
11. Jing XY, Wu F, Dong X, Qi F, Xu B (2015) Heterogeneous cross company defect prediction by unified metric representation and CCA-based transfer learning. In: Proceedings of the 10th joint meeting on foundations of software engineering, pp 496–507
12. Nam J, Kim S (2015) Heterogeneous defect prediction. In: Proceedings ESEC/FSE 2015 proceedings of the 10th joint meeting on foundations of software engineering, ACM. New York, pp 508–519. <https://doi.org/10.1145/2786805.2786814>
13. Lin C, Fang B, Shang Z, Tang Y (2015) Negative samples reduction in cross-company software defects prediction. *Inf Softw Technol* 62:67–77
14. Ryu D, Choi O, Baik J (2016) Value-cognitive boosting with a support vector machine for crossproject defect prediction. *Empirical Softw Eng*. <https://doi.org/10.1007/s10664-014-9346-4>
15. Zhang, F et al. (2016) Cross-project defect prediction using a connectivity-based unsupervised classifier. In: Proceedings of the 38th international conference on software engineering. ACM
16. Xia X et al (2016) Hydra: massively compositional model for cross-project defect prediction. *IEEE Trans Softw Eng* 42(10):977–998
17. Herbold S et al (2017) A comparative study to benchmark cross-project defect prediction approaches. *IEEE Trans Softw Eng*
18. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 16:321–357
19. He P, He Y, Yu L, Li B (2018) An improved method for cross-project defect prediction by simplifying training data. *Math Probl Eng* 18, Article ID 2650415
20. Limsettho N, Bennin, K, Keung J, Hata H, Matsumoto K (2018) Cross project defect prediction using class distribution estimation and oversampling. *Inf Softw Technol* 100. <https://doi.org/10.1016/j.infsof.2018.04.001>

# Performance Analysis of Square and Triangular CNT Bundle Interconnects Driven by CNTFET-Based Inverters



P. Uma Sathyakam, Ananyo Banerjee and P. S. Mallick

**Abstract** This paper proposes the use of triangular cross-sectioned CNT (T-CNT) bundle interconnects for VLSI circuits. The geometry of T-CNT bundles has the advantage of offering least possible crosstalk between adjacent interconnects. The performance factors like propagation delay, power dissipated, crosstalk delay, crosstalk power, On/OFF time of the output waveforms and the output waveform swing are analyzed and compared with traditionally used square CNT (S-CNT) bundle interconnects. Results show that T-CNT bundles offer lesser power dissipation at longer lengths  $>1000 \mu\text{m}$ . Also the crosstalk delay is lesser for T-CNT bundles compared to S-CNT bundles. Both the types of interconnects are driven by GAA CNTFET based inverter circuits.

**Keywords** Interconnects · Carbon nanotubes · CNTFET · Delay · Crosstalk · Power dissipation

## 1 Introduction

Miniaturization of devices in integrated circuits is underway to improve the performance of futuristic electronic devices and applications [1]. Beyond the CMOS technology, large-scale integration of circuits is possible due to innovations in design of new transistors using new materials that have the potential to drastically change the process technologies for IC fabrication in a big way [2]. Performance factors like delay, power dissipation and crosstalk can be optimized by using nanoscale devices like CNTFETs [3]. Further, the performance of interconnects is of paramount importance that decides the overall performance of integrated circuits. In an attempt to

---

P. Uma Sathyakam (✉) · A. Banerjee · P. S. Mallick  
Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India  
e-mail: [umasathyakam.p@vit.ac.in](mailto:umasathyakam.p@vit.ac.in)

A. Banerjee  
e-mail: [ananyo.emon@gmail.com](mailto:ananyo.emon@gmail.com)

P. S. Mallick  
e-mail: [psmallick@vit.ac.in](mailto:psmallick@vit.ac.in)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_32](https://doi.org/10.1007/978-981-15-2329-8_32)

combine the benefits of advances in nano-fabrication processes and to minimize the component size in ICs, we propose to use triangular CNT bundles as VLSI interconnects. In fact, we have proposed the use of T-CNT bundles as VLSI interconnects due to their inherent advantage of lesser propagation, crosstalk delay and power dissipation [4–6]. Fabrication of ‘V’-shaped grooves by Smith et al. [7] has encouraged us to propose this new geometry of CNT bundle interconnects. The basic performance analysis of triangular CNT bundle interconnects shows that they can outperform traditionally proposed square CNT bundles [4–6].

Many types and geometries of CNT bundles are proposed earlier. We earlier proposed the use of mixed CNT bundles that contain both single-walled and multi-walled CNTs in it [8]. This type of bundles proved to be advantageous than SWCNT and MWCNT bundle interconnects [9]. Later, we also proposed CNT interconnects with air-gaps where the dielectric medium is considered as air. This is advantageous as the dielectric constant is unity between adjacent wires and the electrostatic coupling can be reduced [10]. Crosstalk-induced delay and noise effects are studied when semiconducting CNTs are used as shields around CNT bundle interconnects [11]. It was also proposed earlier the use of semiconducting CNTs around mixed CNT bundles for crosstalk reduction [12]. Lastly, emerging ultra low-k (ULK) materials as dielectrics for CNT interconnect technology is proposed and reviewed [13]. The impact of performance of SWCNT bundle interconnects when shielded by semiconducting materials is also analyzed and simulated [14]. Results are encouraging.

In this paper, we compare the performance of square CNT (S-CNT) bundles and triangular CNT (T-CNT) bundles when they are driven by CNTFET-based inverters. Factors like propagation delay, power dissipation, crosstalk delay, crosstalk power and power delay product are compared in this paper.

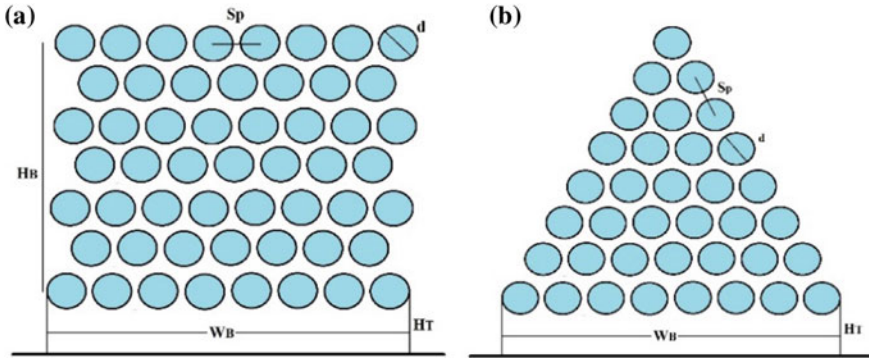
## 2 Proposed Interconnect Models

Carbon nanotube interconnects are modeled as equivalent single conductor transmission lines (ESC-TL). Since a single CNT is having very high resistance of around 6.45 k $\Omega$ , bundles of CNTs are proposed as interconnects [8, 9]. Figure 1 shows the schematic diagrams of S-CNT and T-CNT bundle interconnects placed over a grounded substrate.  $H_B$  is the height of the bundle and  $W_B$  is the width of the bundle. Each CNT is having a diameter  $d$  and  $S_p$  is the inter-CNT distance in a bundle.  $H_t$  is the height at which the CNT bundle is placed from the ground plane. Table 1 shows the dimensions and the geometrical parameters of the CNT bundles.

The width of the bundles is considered as same for the ease of comparison. So, the number of CNTs in T-CNT bundles will be lesser than that of S-CNT bundles. This makes the intrinsic resistance of T-CNT bundles to be higher than the S-CNT bundles as evident in Table 2. The ESC-TL parameters of CNT bundle interconnects are listed in Table 2.

The ESC resistance is the CNT bundle resistance which is the sum of quantum resistance and scattering-induced resistance. The ESC kinetic inductance is due to the





**Fig. 1** Geometry of **a** square CNT (S-CNT) bundle and **b** triangular CNT (T-CNT) bundle interconnects

**Table 1** Dimensions and geometry of the CNT bundles

Parameter	S-CNT	T-CNT
Number of CNTs in the bundle	564	300
Number of CNTs facing ground	24	24
Width of the bundle (nm)	34	34
Height of the bundle (nm)	34	29.4
H/W aspect ratio (AR)	1	0.864

kinetic energy of the electrons and their Fermi velocity. The intrinsic ESC quantum capacitance is due to the density of states available at Fermi level. It must be noticed that the electrostatic capacitance is same for both S-CNT and T-CNT bundles as the width of both the bundles is same. The expressions of these ESC parameters per unit length are discussed earlier in many papers [3–6].

### 3 Transient Analysis

Now, we perform HSPICE simulations of the proposed ESC-TL circuits of the interconnects. We use a standard driver-interconnect-load model for simulation as shown in Fig. 2 [9].

The driver circuit is an inverter made of CNTFETs from the VS-CNTFET library by Stanford University [15]. We give an input pulse of 0.71 V with an ON time and OFF time of 50 ns, rise and fall time of 0.1 ns and a period of 100 ns.

Table 3 enlists the ON and OFF times of the output waveforms of the DIL model for different interconnects lengths. The ON time is higher for S-CNT bundle interconnects at higher lengths compared to T-CNT bundle interconnects. On the other hand, the OFF time is slightly higher for T-CNT bundles for all lengths compared to S-CNT bundles. The ON and OFF times are significant for circuit operation as it

**Table 2** ESC-TL parameters of S-CNT and T-CNT bundles

Length ( $\mu\text{m}$ )	$R_{\text{ESC}}$ (k $\Omega$ )		$L_{k,\text{ESC}}$ (nH)		$C_{Q,\text{ESC}}$ (pF)		$C_{E,\text{ESC}}$ (fF)		$C_C$ (nF)	
	S-CNT	T-CNT	S-CNT	T-CNT	S-CNT	T-CNT	S-CNT	T-CNT	S-CNT	T-CNT
500	7.27	10.75	3.55	6.665	112.8	60	360	158.16	146.1	
1000	14.54	21.5	7.1	13.33	225.6	120	720	316.32	292.2	
1500	21.81	32.25	10.65	19.99	338.4	180	1080	474.48	438.3	
2000	29.08	43	14.2	26.66	451.2	240	1440	632.64	584.4	

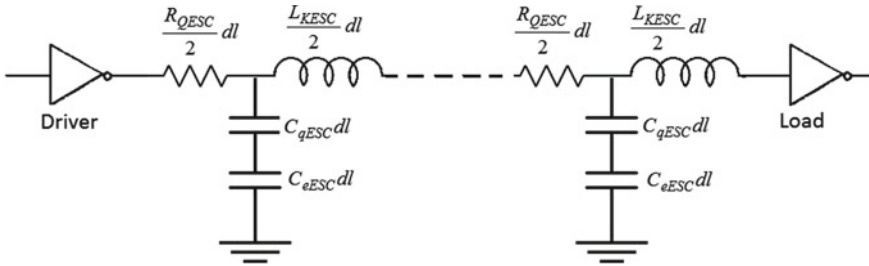


Fig. 2 Circuit model of driver-interconnect-load model used for simulations

Table 3 Timing and output waveform swing of proposed interconnects

Length ( $\mu\text{m}$ )	ON time (ns)		OFF time (ns)		Swing	
	S-CNT	T-CNT	S-CNT	T-CNT	S-CNT	T-CNT
500	50.083	50.105	49.92	49.89	Excellent swing	Good swing
1000	50.117	50.072	49.89	49.91	Good swing	Good swing
1500	49.857	48.893	49.91	49.95	Delayed waveform	Delayed waveform
2000	47.695	42.631	50.12	51.54	Delayed waveform	Better waveform (delayed)

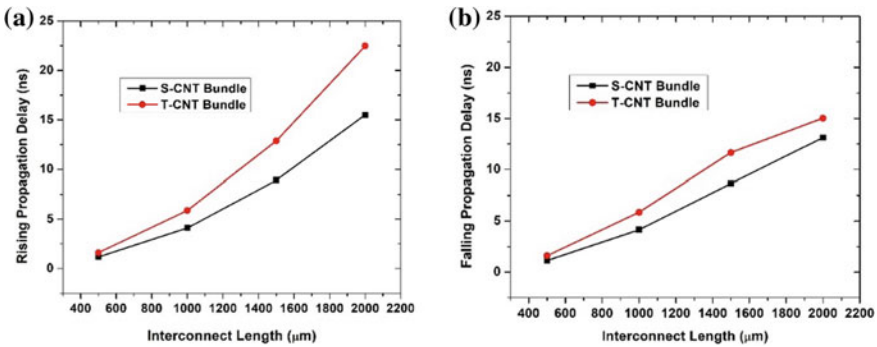


Fig. 3 Propagation delay a rising delay b falling delay

can yield information on the power dissipated by the device and interconnects. Further, the power dissipated can be reduced by identifying longer ON/OFF times and sizing the CNTFET device geometry accordingly to optimize the ON/OFF timing. Secondly, the swing of the output waveform is very important. Generally, the swing is good for smaller lengths of smaller than 1000  $\mu\text{m}$ , while it degrades for higher lengths. Particularly, when the circuit is operating at very low voltages ( $<V_{Th}$ ), the swing may not be from  $V_{dd}$  to  $0V$  or vice versa [3].

**Fig. 4** Power dissipated by the interconnects

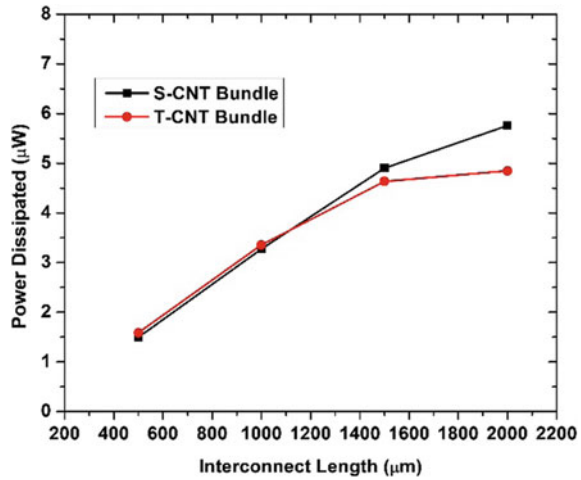


Figure 3 shows the rising and falling propagation delays of S-CNT and T-CNT bundles interconnects at different lengths. In both the cases, the delay increases as the length increases. The delay is higher for T-CNT bundle interconnects compared to S-CNT bundle interconnects. This is due to the fact that the resistance of the T-CNT bundles is higher than S-CNT bundles as evident from Table 1.

Figure 4 shows the power dissipated by the interconnects at different lengths. Again, the power dissipated increases as the length increases. However, the T-CNT bundles dissipate lesser power at lengths greater than 1000 µm. This can be due to the fact that, long T-CNT bundles have lesser quantum capacitance ( $C_{Q,ESC}$ ) compared to S-CNT bundles. So, this is one of the advantages of using T-CNT bundles for low-power interconnects applications, where speed is not the primary constraint. Furthermore, the ON time of T-CNT bundle interconnects is lesser than S-CNT bundles as evident from Table 3, which is also accountable for lower power dissipated.

## 4 Crosstalk Analysis

Adjacent interconnects are capacitively coupled to each other. When signals pass through them, electrostatic crosstalk happens and the propagation delay induced is called as the crosstalk delay. Here, we employ an aggressor-victim model where the interconnect that is energized is the aggressor and the one that is connected to the ground is the victim as shown in Fig. 5.

From Fig. 6a, it can be seen that the crosstalk delay is higher for T-CNT bundle interconnects than the S-CNT bundle interconnects. This is because of the higher resistance of the wires. Moreover, the crosstalk power dissipated by the coupled interconnects decreases when the wire length increases, which is due to the lesser coupling capacitance between adjacent interconnects for T-CNT bundles than

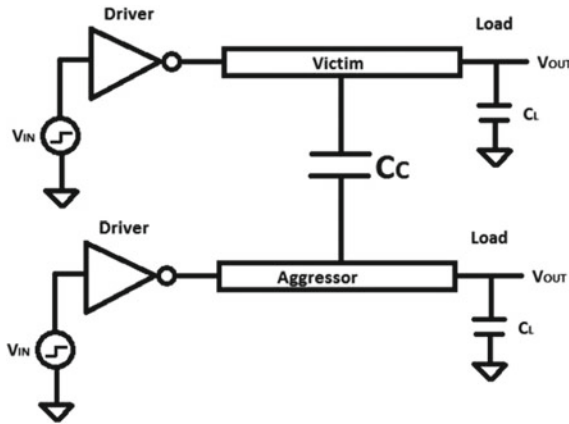


Fig. 5 Aggressor-Victim model of capacitively coupled interconnects

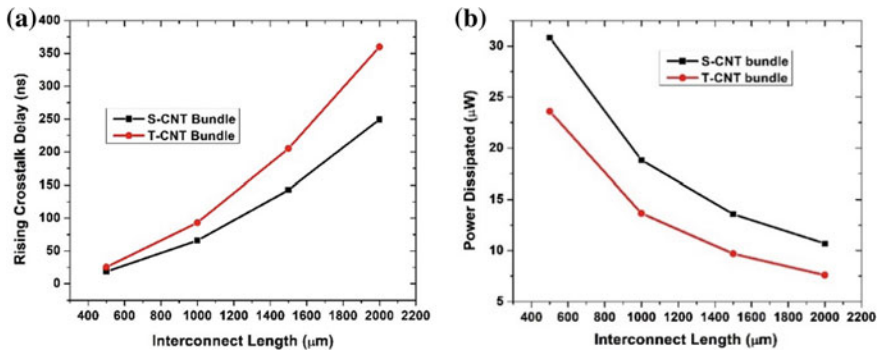


Fig. 6 a Crosstalk delay of coupled interconnects b power dissipated by coupled interconnects

S-CNT bundles. Also, the smaller ON time of the output pulse indicates that the power dissipated is lesser for T-CNT bundles as the length increases.

### 5 Conclusions

This paper discussed the implications of using T-CNT bundle interconnects in VLSI circuits and the performance is compared to traditionally proposed S-CNT bundles. Overall, it is found that T-CNT bundles perform better at longer interconnect lengths (>1000 μm) in terms of crosstalk delay and power dissipated. On the other hand, S-CNT bundles are better performers in terms of propagation delay. The crosstalk delay is lesser for T-CNT bundles as they are least coupled because of their geometry. Further, the ON/OFF timings and the swing of the waveforms are analyzed to find

out their implications on the performance of the interconnects. It was found that the timing and swing influences the power dissipated by the interconnects and it follows a reverse trend w.r.t length compared to traditionally achieved power dissipation trends in literature.

## References

1. Kyeremateng NA, Brousse T, Pech D (2017) Microsupercapacitors as miniaturized energy-storage components for on-chip electronics. *Nat Nanotechnol* 12:7–15. <https://doi.org/10.1038/nnano.2016.196>
2. Rajshekar K, Hsu HH, Kumar KUM, Sathyanarayanan P, Velmurugan V, Cheng CH, Kannadassan D (2019) Effect of plasma fluorination in p-type SnO TFTs: experiments, modeling, and simulation. *IEEE Tran Electron Dev* 66:1314–1321. <https://doi.org/10.1109/TED.2019.2895042>
3. Sathyakam PU, Mallick PS, Saxena AA (2019) High speed subthreshold operation of carbon nanotube interconnects. *IET Circuits Devices Syst* 13:526–533. <https://doi.org/10.1049/iet-cds.2018.5118>
4. Sathyakam PU, Bisht A, Tandon Y, Mallick PS (2017) Triangular CNT bundles as VLSI interconnects. In: 2016 3rd international conference on emerging electronics, ICEE 2016. <https://doi.org/10.1109/icemelec.2016.8074592>
5. Sathyakam PU, Mallick PS (2019) Triangular carbon nanotube bundle interconnects for subthreshold VLSI circuits. *J Electronic Mater*. <https://doi.org/10.1007/s11664-019-07431-z>
6. Sathyakam PU, Mallick PS, Singh P (2019) Geometry based crosstalk reduction in carbon nanotube interconnects. *J Circuit Syst Comp*. <https://doi.org/10.1142/S0218126620500942>
7. Smith CLC et al (2014) Efficient excitation of channel plasmons in tailored UV-lithography-defined V-grooves. *Nano Lett* 14:1659–1664. <https://doi.org/10.1021/nl5002058>
8. Sathyakam PU, Mallick PS (2012) Inter-CNT capacitance in mixed CNT bundle interconnects for VLSI circuits. *Int J Electron* 99:1439–1447. <https://doi.org/10.1080/00207217.2012.669721>
9. Sathyakam PU, Mallick PS (2012) Towards realisation of mixed carbon nanotube bundles as VLSI interconnects: A review. *Nano Commun Netw* 3:175–182. <https://doi.org/10.1016/j.nancom.2012.09.004>
10. Sathyakam PU, Mallick PS (2017) Carbon nanotube interconnects with air-gaps: effect on thermal stability, delay and area. *J Nano Res* 48:29–37. <https://doi.org/10.4028/www.scientific.net/JNanoR.48.29>
11. Sathyakam PU, Mallick PS (2018) Reducing crosstalk induced delay and peak noise in carbon nanotube interconnects. *Current Nanosci* 14:76–80. <https://doi.org/10.2174/1573413713666170821123024>
12. Sathyakam PU, Karthikeyan A, Rajesh JK, Mallick PS (2014) Reduction of crosstalk in mixed CNT bundle interconnects for high frequency 3D ICs and SoCs. In: 2014 International conference on advances in electrical engineering (ICAEE). <https://doi.org/10.1109/icaee.2014.6838461>
13. Sathyakam PU, Mallick PS (2018) Future dielectric materials for CNT interconnects—possibilities and challenges. *J Nano Res* 52:21–42. <https://doi.org/10.4028/www.scientific.net/JNanoR.52.21>
14. Sulochana V, Agrawal S, Singh B (2019) Impact on performance of Bundled SWCNT interconnects surrounded with semiconductor shielding materials. In: Mishra S et al (eds) *Applications of computing, automation and wireless systems in electrical engineering*, LNEE. vol 553. Springer, Heidelberg, pp 837–849. [https://doi.org/10.1007/978-981-13-6772-4\\_72](https://doi.org/10.1007/978-981-13-6772-4_72)
15. VS-CNFET Model. Available at <https://nano.stanford.edu/stanford-cnfet2-model>

# A Current Tunable Third-Order Oscillator Using CCDDCC



Sunil Kumar Yadav, Manoj Joshi and Ashish Ranjan

**Abstract** This paper reports a tunable sinusoidal oscillator using current-controlled differential difference current conveyor (CCDDCC) with grounded capacitor and resistor. The proposed oscillator circuit has the following advantages, viz. minimum number of active components, suitable for high-frequency operation and electronically tunable frequency of oscillation (FO) and condition of oscillation (CO) using bias current of CCDDCC. The workability test has been examined using PSpice with 0.18  $\mu\text{m}$  CMOS technology.

**Keywords** Current-controlled differential difference current conveyor (CCDDCC) · Third order · High-frequency oscillator · Electronically tunable

## 1 Introduction

An ever-increasing demand of a portable system with lightweight and low-cost electronic devices in consumer electronics increases the interest of researchers to design a low-supply voltage-based high-accuracy electronic device. To fulfill these advance demands of consumers, a current mode approach has become a line of interest in analog integrated circuits due to its high-performance parameter in comparison with traditional voltage mode op-amp [1–3]. Therefore, the existence of modern applications for analog circuits for lower voltage swings and low-supply voltage by using the current mode approach has contributed a new era of research in analog circuit design. Simultaneously, the interface between different active blocks that works for different modes is also analyzed and implemented in several literatures. Some of

---

S. K. Yadav · M. Joshi · A. Ranjan (✉)

Department of Electronics and Communication Engineering, National Institute of Technology Manipur, Imphal, India

e-mail: [ashish.ism@rediffmail.com](mailto:ashish.ism@rediffmail.com)

S. K. Yadav

e-mail: [sunilgwal0543ei@gmail.com](mailto:sunilgwal0543ei@gmail.com)

M. Joshi

e-mail: [manojjoshi1506@gmail.com](mailto:manojjoshi1506@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication*

*Engineering*, Lecture Notes in Networks and Systems 106,

[https://doi.org/10.1007/978-981-15-2329-8\\_33](https://doi.org/10.1007/978-981-15-2329-8_33)

active current mode blocks are used for the realization of filter [4, 5], Schmitt trigger [6], Colpitts oscillator [7], chaotic oscillator [8–10] and instrumentation amplifier [11], active inductor [12] and many more circuit designs.

Oscillators are broadly used in various advance signals processing as an individual unit in communication system, control system, instrumentation and measurement systems [13, 14]. Also, quadrature oscillator allows us for critical and accurate applications in telecommunication instrument, viz. modulators, transmission of radio signal, quadrature mixer, etc. In the last two decades, third-order oscillators can be implemented in various signal processing circuits by using active devices, viz. CCII [15], DVCC [16], OTA [17], CDTA [18–21], CCCII [22], CCCDTA [23], VDTA [24], OTRA [25], DVCCTA [26], MCCFTA [27], MO-DVCCTA [28], CDBA [29], etc. In the literature, many current tunable oscillators generate continuous periodic oscillating waveforms whose frequency is controlled by the biasing current ( $I_B$ ) as well as the phase shift between the two output waveforms. Most of the literature comes with more than one active block to realized third-order oscillators, lack of electronically tunable and low-frequency operation. However, in most of the realization the CO and FO are not controlled electronically and independently which are not suitable for oscillator with variable frequency. The literature reveals that there are many circuits designed [15–29] to have some of the mentioned features above, but there is none of them having all the features.

A new active building block is an advance modification of CCII termed as Differential Difference Current Conveyor (DDCC) [30]. The properties of DDCC and CCII can be utilized in the realization of current mode circuit termed as current-controlled differential difference current conveyor (CCDDCC). The main advantage of CCDDCC block is that the internal resistance of X-terminal is depending on bias current ( $I_B$ ) because of this there is no need to add any external resistance at this terminal.

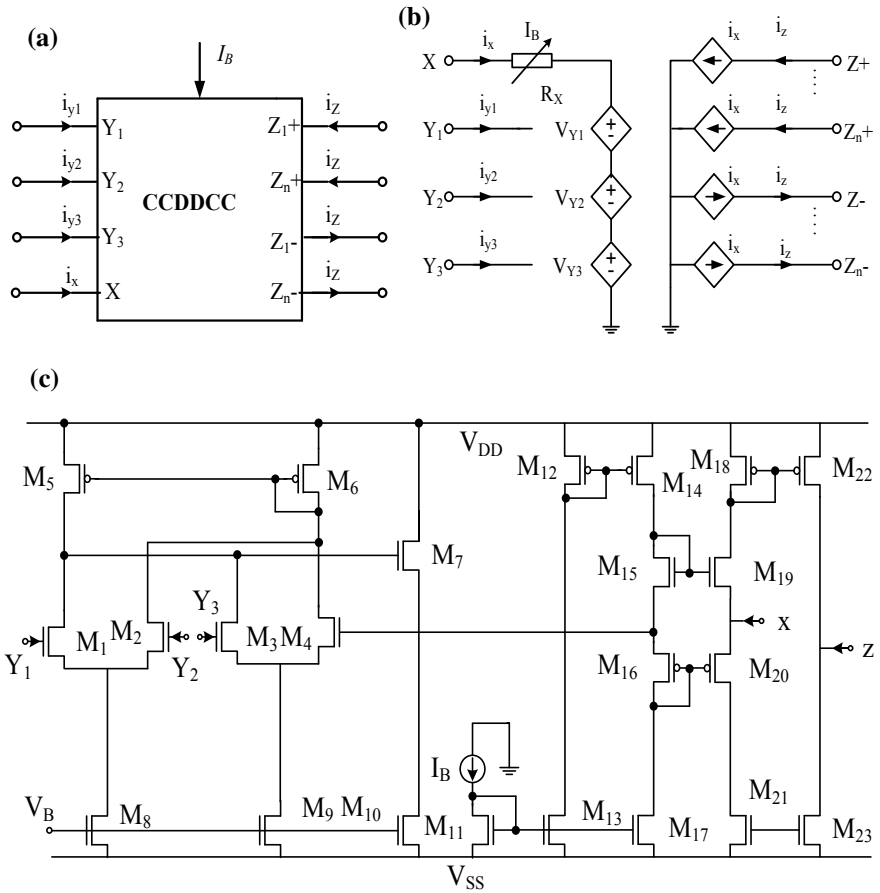
In this paper, a current tunable oscillator using CCDDCC as the active building block is reported. The proposed sinusoidal oscillator is realized with single CCD-DCC with grounded passive components. The simulation results are well demonstrated through PSpice circuit simulator using 0.18  $\mu\text{m}$  CMOS technology process parameters.

## 2 Circuit Description

### 2.1 Circuit Topology of CCDDCC

A complete circuit topology of CCDDCC is shown in Fig. 1. It includes circuit symbol, equivalent electrical circuit and CMOS implementation of CCDDCC. Moreover, the port characteristics can be examined by the following set of voltage and current equations as:





**Fig. 1** Current-controlled differential difference current conveyor (CCDDCC). **a** Circuit symbol, **b** equivalent electrical circuit, **c** CMOS-based internal structure

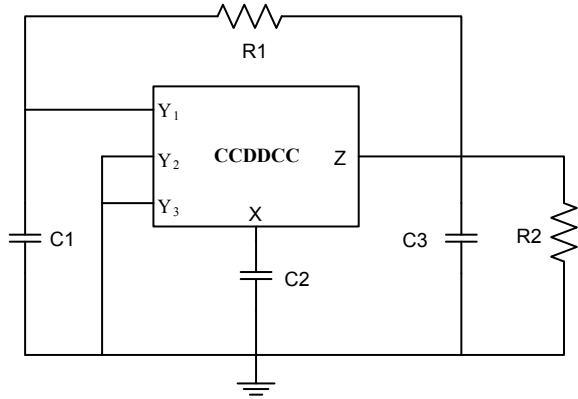
$$I_X = \pm I_Z, V_X = I_X R_X - V_{X1} - V_{Y2}, I_Y = 0 \tag{1}$$

where  $R_X$  is the intrinsic resistance CCDDCC at  $X$ -terminal port and mathematically expressed as:

$$R_X = \frac{1}{g_{m_{19}} + g_{m_{20}}} \tag{2}$$

where  $g_{m_{19}}$  and  $g_{m_{20}}$  are transconductance terms for transistor  $M_{19}$  and  $M_{20}$  and the matched transistor ( $M_{19}$  and  $M_{20}$ ) gives the  $R_X$  values as:

**Fig. 2** Proposed third oscillator using CCDDCC



$$R_X = \frac{1}{\sqrt{8\mu C_{ox} \left(\frac{W}{L}\right) I_B}} \tag{3}$$

where  $\mu$ ,  $C_{ox}$ ,  $W$  and  $L$  are well-known electrical parameters of a MOS transistor. The intrinsic resistance  $R_X$  can be adjusted by the bias current ( $I_B$ ) of CCDDCC.

### 2.2 Proposed Current Tunable Third-Order Oscillator

An active design for a third-order oscillator using CCDDCC with passive components is given in Fig. 2. In general, a third-order oscillator can be characterized by the polynomial equation ( $q(s) = 0$ ) as:

$$q(s) = As^3 + Bs^2 + Cs + D = 0 \tag{4}$$

A routine analysis of Fig. 2 provides a third-order characteristic equation as:

$$\left. \begin{aligned} & s^3 C_1 C_2 C_3 R_1 R_2 R_X + s^2 [C_3 R_X (C_1 R_1 + C_2 R_2 + C_1 R_2) - C_1 C_2 R_1 R_2] \\ & + s [C_3 R_X + C_3 R_2 - C_1 R_1 - C_2 R_2 - C_1 R_2] + 1 = 0 \end{aligned} \right\} \tag{5}$$

By comparing Eq. (4) and (5), its gives the coefficients A, B, C and D as:

$$\left. \begin{aligned} A &= C_1 C_2 C_3 R_1 R_2 R_X \\ B &= C_3 R_X (C_1 R_1 + C_2 R_2 + C_1 R_2) - C_1 C_2 R_1 R_2 \\ C &= C_3 R_X + C_3 R_2 - C_1 R_1 - C_2 R_2 - C_1 R_2 \\ D &= 1 \end{aligned} \right\} \tag{6}$$

By putting the value  $s = j\omega$  yields:

$$-Aj\omega^3 - Bw^2 + Cj\omega + D = 0 \quad (7)$$

To achieve the CO and FO, the real and imaginary part of Eq. (7) gives fundamental oscillator characteristics as:

$$\left. \begin{aligned} -A\omega^3 + C\omega &= 0 \\ -B\omega^2 + D &= 0 \end{aligned} \right\} \quad (8)$$

From Eq. (8), CO and FO are written as:

$$\left. \begin{aligned} AD &= BC \\ \omega &= \sqrt{\frac{C}{A}} = \sqrt{\frac{D}{B}} \end{aligned} \right\} \quad (9)$$

By assuming,  $C_1 = C_2 = C$ ,  $C_3 = 6C$ ,  $R_1 = R$  and  $R_2 = 2R$  provide the modified characteristic equation as:

$$2s^3C^3R^2R_X + s^2[5C^2RR_X - 2C^2R^2] + sC[R_X - 3R] + 1 = 0 \quad (10)$$

Hence, the modified CO and FO may be obtained:

$$\text{FO: } f = \frac{1}{2\pi} \sqrt{\frac{R_X - 3R}{2C^2R^2R_X}} \quad (11)$$

and

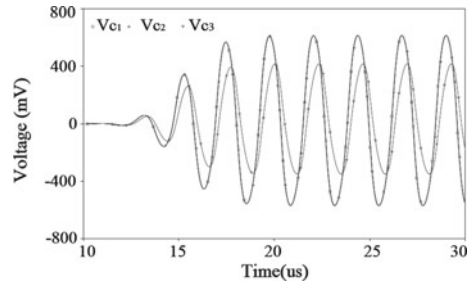
$$\text{CO: } \frac{R_X - 3R}{2C^2R^2R_X} = \frac{1}{5C^2RR_X - 2C^2R^2} \quad (12)$$

As observed by Eq. (11) and (12), the CO and FO can be adjusted by  $R_X$  which is a function of  $I_B$ .

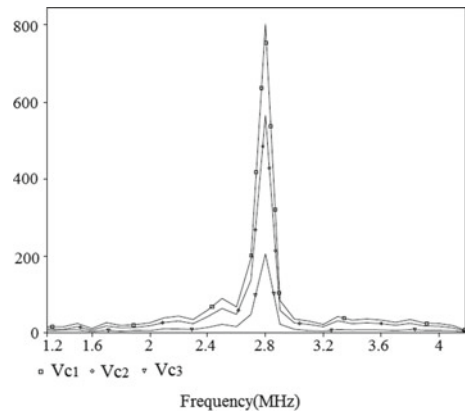
### 3 Simulation Results

The workability test of the proposed design is verified by using PSpice. Figure 2 shows proposed current tunable third-order grounded capacitor oscillator using single CCDDCC as the active device with  $0.18 \mu\text{m}$  CMOS technology parameter, and given supply voltage to the active block is  $\pm 1.25 \text{ V}$ . In the realization of proposed circuit, the range of bias current  $I_B$  can be taken from  $100 \text{ nA}$  to  $100 \mu\text{A}$ . For this range of current, the  $X$ -terminal resistance will be from  $12$  to  $1 \text{ k}\Omega$ . On the basis of frequency requirement, the value of resistance on the terminal  $X$  can be adjusted with  $I_B$  ( $60 \mu\text{A}$ ) of CCDDCC. The passive component values being adjusted to  $C_1 = C_2 = 20 \text{ pF}$ ,  $R_1 = 5 \text{ k}\Omega$ ,  $R_2 = 10 \text{ k}\Omega$  and  $C_3 = 120 \text{ pF}$  to obtain the output of the sinusoidal oscillator

**Fig. 3** Output waveforms of the proposed oscillator



**Fig. 4** Frequency spectrum of the proposed oscillator



with FO range in MHz. The transient response of the proposed oscillator is shown in Fig. 3. The obtained FO from the PSpice simulation is  $f \approx 2.8$  MHz. That agrees well with the theoretical value (3 MHz). Also, Fig. 4 shows the frequency spectrum of the proposed oscillator output. Finally, a brief comparative study of the oscillator is investigated in Table 1 in terms of the active and passive component, technology used, electronically tunable and maximum range of frequency. It reveals that the proposed circuit uses only single active block to implement the third-order oscillator [27]. However, it uses minimum number of passive components in comparison with [15, 16, 25]. In comparison with circuits [15–17, 25, 29], proposed circuit provides electronically tunability. Also, proposed oscillator works in high-frequency range (in MHz).

## 4 Conclusion

A third-order high-frequency tunable sinusoidal oscillator is designed by using CCD-DCC. The proposed oscillator has the following features realized by using single active components, and all the capacitors are grounded and provide electronically tunable CO and FO. Finally, the theoretical observations of the reported oscillator are well verified through PSpice simulator.

**Table 1** Comparative study of the different chaotic circuits

References	Active elements	Technology parameter ( $\mu\text{m}$ )	No. of active elements	No. of $R + C$	Electronically tunable	Maximum range of frequency (Hz)	Supply voltage (V)
[15]	CCII	–	3	$3C + 3R$	No	$\approx k$	5
[16]	DVCC	0.5	3	$3C + 3R$	No	7.96 M	2.5
[17]	OTA	5	3	$>3C$	No	2 M	3
[18]	CDTA	0.18	3	$3C$	Yes	–	1.25
[20]	CDTA	0.18	3	$3C$	Yes	$\approx M$	2.5
[22]	CCCII	0.5	3	$3C$	Yes	8 M	2.5
[24]	VDTA	0.35	2	$3C$	Yes	1.10 M	2
[25]	OTRA	0.25	3	$3C + 4R$	No	315 k	2.5
[26]	DVCCTA	0.25	2	$3C + 2R$	Yes	$\approx M$	1.25
[27]	MCCFTA	0.25	1	$3C$	Yes	90 k	1
[28]	MO-DVCCTA	0.18	2	$3C + 2R$	Yes	2.1 M	0.9
[29]	CDBA	0.5	2	$3R + 2C$	No	500 k	2.5
This work	CCDDCC	0.18	1	$3C + 2R$	Yes	2.8 M	1.25

$C$  = capacitor,  $R$  = resistor

## References

1. Ferri G, Guerrini NC (2003) Low-voltage low-power CMOS current conveyors. Cluwer Academic Publishers
2. Toumazou C, Lidgey FJ, Haigh DG (1990) Analogue IC design: the current mode approach. IEE Circuits and Systems Series 2 Peter Peregrinus Ltd
3. Sedra AS, Smith KC (1970) A second generation current conveyor and its applications. IEEE Trans CT-17, 132–134
4. Tarunkumar H, Ranjan A, Perumulla S, Pheiroijam NM (2017) Four input single output based third order universal filter using four terminal floating mellor. Analog Integr Circ Sig Process 93:89–98
5. Ranjan A, Ghosh M, Paul SK (2014) Third-order voltage-mode active-C band pass filter. Int J Electron 102:781–791
6. Ranjan A, Pamu H, Tarunkumar H (2018) A novel Schmitt trigger and its application using a single FTFN. Analog Integr Circ Signal Process 1–13 (2018)

7. Joshi M, Bhatt V, Ranjan A, Benjwal P (2017) Realization of Colpitts oscillator using second generation current controlled current conveyor. In: Proceedings of the second international conference on research in intelligent and computing in engineering, pp 49–52
8. Joshi M, Ranjan A (2018) Realization of multi scroll 2D chaotic oscillator using DVCC. In: Mishra S, Sood Y, Tomar A (eds) Applications of computing, automation and wireless systems in electrical engineering. Lecture notes in electrical engineering, vol 553. Springer, Singapore
9. Joshi M, Ranjan A (2019) An autonomous chaotic and hyperchaotic oscillator using OTRA. *Analog Integr Circuits Sig Process*. <https://doi.org/10.1007/s10470-019-01395-0>
10. Joshi M, Ranjan A (2019) New simple chaotic and hyperchaotic system with an unstable node. *Int J Electron Commun (AEU)* 108:1–9
11. Abuelmaatti MT, Al-Qahtani MA (1998) Low component second-generation current conveyor-based multiphase sinusoidal oscillator. *Int J Electron* 84:45–52
12. Skoti GD, Psychalinos P (2010) Multiphase sinusoidal oscillators using second generation current conveyors. *AEU—Int J Electron Commun* 64:1178–1181. <https://doi.org/10.1016/j.aeu.2009.11.013>
13. Maheshwari S (2009) Analogue signal processing applications using a new circuit topology. *IET Circuits Devices Syst* 3(3):106–115
14. Keskin AU, Bielek D (2009) Current-mode quadrature oscillator using CDTAs. *IEE Proc Circuits Devices Syst*. 153:214–218
15. Soliman AM (1998) Current mode CCII oscillators using grounded capacitors and resistors. *Int J Circuit Theor Appl* 26:431–438
16. Elwan HO, Soliman AM (1997) Novel CMOS differential voltage current conveyor and its applications. *Proc IEE Circuits Devices Syst* 144:195–200
17. Prommee P, Dejhan K (2002) An integrable electronic-controlled quadrature sinusoidal oscillator using CMOS operational trans-conductance amplifier. *Int J Electron* 89(5):365–379
18. Horng JW (2009) Current-mode third-order quadrature oscillator using CDTAs. *Active Passive Electron Componen*. <https://doi.org/10.1155/2009/789171>
19. Horng JW, Lee H, Wu JY (2010) Electronically tunable third order quadrature oscillator using CDTAs. *Radio Eng* 19:326–330
20. Jin J, Wang C, Sun J (2015) Novel third-order quadrature oscillators with grounded capacitors. *Automatika* 56(2):207–216
21. Prasad D, Bhaskar DR, Singh AK (2008) Realisation of single-resistance-controlled sinusoidal oscillator: a new application of the CDTA. *WSEAS Trans Electron* 5(6):257–259
22. Maheshwari S (2010) Current-mode third-order quadrature oscillator. *IET Circuits Devices Syst* 4:188–195. <https://doi.org/10.1049/iet-cds.2009.0259>
23. Sakul C, Jaikla W, Dejhan K (2011) New resistorless current mode quadrature oscillators using 2 CCCDTAs and grounded capacitors. *Radio Eng* 20(4):890–897
24. Phatsornsiri P, Lamun P, Torteanchai U (2014) Current-mode third-order quadrature oscillator using VDTAs and grounded capacitors. In: Proceeding of the 4th joint international conference on information and communication technology, electronic and electrical engineering (JICTEE). Thailand, pp 1–4
25. Kungern M, Kansiri I (2014) Single-element control third order quadrature oscillator using OTRAs. In: Proceedings of CT and KE. Thailand, pp 24–27
26. Pandey N, Pandey R (2015) Approach for third order quadrature oscillator realization. *IET Circuits Devices Syst* 9:161–171. <https://doi.org/10.1049/iet-cds.2014.0170>
27. Khaw-Ngam K, Kungern M, Khateb F (2017) Mixed-mode third-order quadrature oscillator based on single MCCFTA. *Radio Eng* 26(2):522–535
28. Chen HP, Hwang YS, Ku YT (2017) A systematic realization of third-order quadrature oscillator with controllable amplitude. *Int J Electron Commun* 79:64–73. <https://doi.org/10.1016/j.aeu.2017.05.039>
29. Prommee P, Somdunyanok M (2009) CMOS-based current-controlled DDCC and its applications to capacitance multiplier and universal filter. *Int J Electron Commun (AEU)* 65:1–8
30. Chiu W, Liu SI, Tsao HW, Chen JJ (1996) CMOS differential difference current conveyor and their applications. *IEE Proc Circuits Devices Syst* 143:91–96

# Parametric Classification of Dynamic Community Detection Techniques



Neelu Chaudhary and Hardeo Kumar Thakur

**Abstract** The community detection in a given network is the idea to find a cluster in the structure. A community is the most densely populated part of the graph. The observed network is mostly sparse having multiple dense partitions in it, for example, a protein–protein interaction network where different proteins interact with each other. Here, communities can be detected by finding the cluster of proteins in the network to find different functional modules. Another example is of Facebook friendship network. Several authors try to find the structure and communities in this type of network. Multiple clusters in one network can also be detected which can overlap with each other. This paper covers the classification of different community detection techniques in dynamic networks and then compares them on the basis of different features, e.g., parallelization, network models, community instability, temporal smoothness, etc.

**Keywords** Community · Dynamic community detection · Community instability · Temporal smoothness

## 1 Introduction

### 1.1 Dynamic Graphs

The scientific literature dedicated to community detection is modeled on static network that is static or unchanged with time [1]. These approaches do not fit the evolving or changing nature of the real-world phenomena. For example, the social network which is dynamic in nature, i.e., evolves with time, contains a temporal dimension which contains valuable information to analyze. These are the dynamic

---

N. Chaudhary (✉) · H. K. Thakur  
Manav Rachna University, Faridabad, India  
e-mail: [neelu@mru.edu.in](mailto:neelu@mru.edu.in)

H. K. Thakur  
e-mail: [hkthakur@mru.edu.in](mailto:hkthakur@mru.edu.in)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_34](https://doi.org/10.1007/978-981-15-2329-8_34)

networks which contain dynamic communities that also change with time. To find communities in dynamic network is the challenging task as both nodes and edges can join and leave a dynamic network at any instant of time. These approaches track the local topologies of the network. In this paper, we discuss the dynamic community detection techniques on dynamic networks.

## 1.2 Community

Community is a set of entities in a complex network that shows the similar set of actions with the other entities of that set (community). To detect and analyze community in complex network is a typical problem as this type of networks is not well posed. These are the cluster of nodes which can overlap with each other; i.e., one node can be a part of different clusters or communities as shown in the figure. Dynamic community thus can be defined as a dynamic cluster of nodes which can change or evolve over time.

## 1.3 Community Life Cycle

A community or a cluster of node goes through stages in its life cycle by the time of its appearance to the stage of vanishing. These transformations can be categorized in different stages, i.e., begin, vanished, extension, shrink, merge and split as shown in figure. In addition to these operations, two more operations were added that is continue when community remains unchanged and re-emergence when a community vanished for some time and then reoccurs as if it has never stopped. Among these operations, some (merge, split and re-emergence) are dealt with different strategies in community detection. For example, for merge operation several strategies can be followed, i.e., absorption when one community vanished and other one remains or replacement when both communities vanished and a new one creates and leads to immediate disruptions of communities.

## 2 Dynamic Community Detection

For a evolving graph that is dynamic in nature, a community is also dynamic and is defined as a set of different pairs of nodes and period  $(N, P)$  such as  $N = (v_1, v_2, v_3 \dots v_n)$  and  $P = (P_1, P_2, P_3 \dots P_n)$ ; then,  $DC = \{(v_1, p_1), (v_2, p_2), (v_3, p_3) \dots (v_n, p_n)\}$ . Here, time period  $p_n$  consists of the starting and ending time of node  $v_n$  that remains connected in a community.  $P_n = ((ts_0, te_0). (ts_1, te_1) \dots (ts_N, te_N))$  with  $ts < te$  for all nodes.



The dynamic community detection in complex network also depends on the several features that affect the complexity of analysis. Based on these features, various approaches of community detection can be classified. These features are as follows.

### 2.1 Network Models

Network modeling process affects the crucial role in shaping network topologies. We can differentiate the time-evolving networks with the static networks as shown in Fig. 1. The static networks depict the complete temporal networks where a snapshot shows the static view of dynamic process. The dynamics can slowly increase by considering weights on edges and nodes. This leads to the aggregate networks but covers only limited dynamics. Thus, a series of time-ordered network snapshots [2] are considered for modeling which also covers the perturbations in topologies. The temporal networks however avoid aggregation but give complete and exact view of dynamics of network. As the expressivity increased, the complexity to analyze that type of network is also increased as shown in Fig. 2.

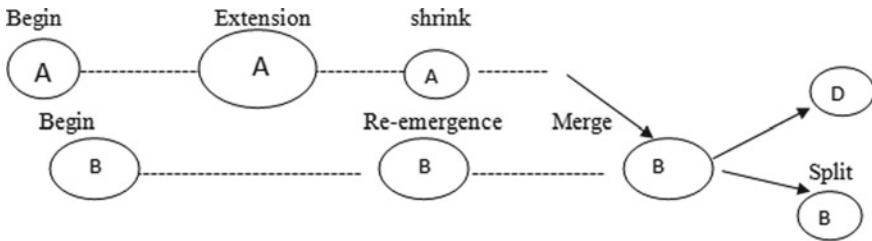


Fig. 1 Community life cycle

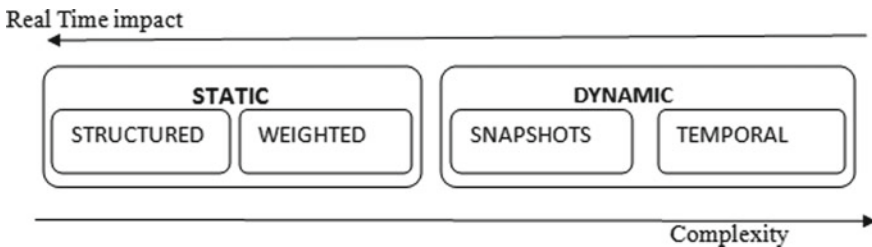


Fig. 2 Network models

## 2.2 *Network Memory*

The analysis of dynamic networks depends highly on the type of system memory. Dynamic data like emails or calls can be transformed into relation network depicting each edge with time duration to consider. This transformation can be taken as synchronized snapshots or changing temporal networks [1]. The storage can be perfect memory network type nodes and edges can only join but the existing edges and nodes will also remain intact, i.e., perfect memory. It can be a limited network type where nodes and edges can disappear with time.

## 2.3 *Community Instability*

Community instability is the major concern in the area of dynamic community detection. There can be multiple decompositions in complex network communities. This generates the ambiguity in the partitions of communities [3]. But mostly algorithms generate the partitions assigning nodes to one community only for simplification. But nodes often belong to multiple communities as discussed early in overlapping of communities. Also, the partition selection process is arbitrary in nature. Different types of selection lead to different results. All these lead to the complexity in identifying the decomposition in stable networks due to community instability.

## 2.4 *Temporal Smoothing*

Smoothing is the process to solve the instability of communities to a great extent. This process can be done explicitly requiring certain level of similarity in partitions  $t$  with partition  $t - 1$ . The smoothing can be implicit [4] which does not explicitly integrate the partition similarity but considers this similarity by optimal construction. The smoothing can also be done by bootstrap by searching stable partition in communities by running same algorithm on the same network.

## 3 **Classification of DCD Techniques**

The dynamic community detection can be classified in different perspectives. We discuss the following of them:

- 3.1 DCD can be classified in classes [5] such as two-stage approach where each cluster when detected at one time stamp is then matched across multiple time stamps. Another class is evolutionary clustering where community detection at time  $t$  is based on the previously found communities in that topology. It can be

a coupling graph in which a typical community detection technique runs on a graph constructed by adding edges between different instances of nodes.

- 3.2 Another classification is based on high level that is online and offline methods. Most of the algorithms are online which can be identified as temporal smoothness methods as discussed previously or dynamic updated that instead of running a new process update the process according to methods found in previous time stamps.
- 3.3 One approach to classify DCD techniques is based on what are the dynamic communities and not restricting on the techniques used to find them [1]. These classes are:

**Instant optimal:** This class considers those communities that depend on the current state of the network. Communities found at  $t$  are considered optimal at that instant with respect to the structure or topology of the network. These classes are not temporally smoothed, thus limiting by the community instability. Mostly, network snapshots are used for these techniques where complete network is not considered but only the changes in between are effective. The advantage of this approach is that it can be directly used on the existing works as communities are instant optimal. Also, this type of detection technique offers parallelization where communities at each step are found parallel which leads to fast process otherwise to a time-consuming process on large networks. The authors used this approach [2, 6, 7].

**Temporal Trade-off:** The main difference between temporal trade-off and instant optimal approaches is that formal approach considers the communities found in previous time stamp also to detect the communities in current time stamp. It is an iterative process which first initializes and then just updates the existing one. It however cannot be parallelized as in optimal as previous communities also play a significant role in detecting current communities (e.g. [8–11]). Both types of network, i.e., snapshot and temporal, can be used for this approach. The main advantage of this approach is the ability to smooth the instability of community as compared to instant optimal.

**Cross-time:** This approach does not consider independently the various steps of dynamic network; instead, it is done in one single process taking all states of the network in account at that time. Examples of these approaches are network transformation and community detection on transversal network (e.g. [12–14]). It is a novel technique based on assumptions on the basic nature of dynamic communities; thus, operations like merge and split can be applicable to these techniques. Also, these are not real-time community detection technique due to requirement of complete knowledge of previous stages of network. Both snapshot network and temporal network can be used for this approach. The main advantage of cross-time approach is that it does not suffer from community drift and instability problem as the previous one. Also, they are more capable to deal with slow evolution and local anomalies.

In Table 1, we have classified different dynamic community detection approaches and compare these approaches on the basis of several features like network model used, community instability, coherency, temporal smoothness, etc.

**Table 1** Comparison of different DCD approaches

Class	Network model	Community instability	Parallelization	Coherent	Network memory	Online/offline	Temporal smoothness
Two-stage approach	Snapshot	×	×	×	Limited	Online/offline	✓
Evolutionary clustering	Snapshot/temporal	×	×	×	Perfect	Online	✓
Coupling graph	Snapshot	✓	×	×	Limited	Offline	×
Instant optimal	Snapshot	✓	✓	×	Limited	Online	×
Temporal trade-off	Snapshot/temporal	×	×	✓	Perfect	Online	✓
Cross-time trade-off	Temporal	×	×	✓	Limited	Online	✓

## 4 Conclusion and Future Work

In this paper, we discussed the dynamic communities and what are the different ways to detect them in dynamic network. We also explain the different stages of the dynamic community life cycle and the types of network models these communities used. Comparison between these models is also discussed based on the features like real-time impact and complexity. Then, we classify the different dynamic community detection techniques or approaches based on certain features like network model, type of memory used, community instability, temporal smoothing, etc. We also discussed the advantages and limitation of each approach as compared to each other. Further, we propose the approach for the dynamic community techniques that overcome the problem of community drift and community instability and should also have the perfect memory and coherent in nature. All these features are lacking in advanced temporal smoothness techniques also. This will lead to more optimized and fast community detection approach that can be used in multiple applications like social networks and mobile applications.

## References

1. Rossetti G, Cazabet R (2018) Community discovery in dynamic networks: a survey. *ACM Comput Surv* 51(2), Article 35
2. Morini M, Flandrin P, Fleury E, Venturini T, Jensen P (2017) Revealing evolutions in dynamical networks. [arXiv:1707.02114](https://arxiv.org/abs/1707.02114)
3. Aynaud T, Guillaume J-L (2011) Multi-step community detection and hierarchical time segmentation in evolving networks. In: Proceedings of the 5th social network mining and analysis workshop, (SNA-KDD Workshop'11)
4. Shang J, Liu L, Xie F, Chen Z, Miao J, Fang X, Wu C (2014) A real-time detecting algorithm for tracking community structure of dynamic networks. [arXiv:1407.2683](https://arxiv.org/abs/1407.2683)
5. Aynaud T, Fleury E, Guillaume J-L, Wang Q (2013) Communities in evolving networks: definitions, detection, and analysis techniques. In: Dynamics on and of complex networks, vol 2. Springer, pp 159–200
6. Bota A, Kresz M, Pluhar A (2011) Dynamic communities and their detection. *Acta Cybernetica* 20(1):35–52
7. Dhouioui Z, Akaichi J (2014) Tracking dynamic community evolution in social networks. In: Proceedings of the IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM'14). IEEE, Los Alamitos, CA, pp 764–770
8. Agarwal MK, Ramamritham K, Bhide M (2012) Real time discovery of dense clusters in highly dynamic graphs: identifying real world events in highly dynamic environments. *Proc VLDB Endowment* 5(10):980–991
9. Ma HS, Huang JW (2013) Cut: community update and tracking in dynamic social networks. In: Proceedings of the 7th workshop on social network mining and analysis. ACM, New York, NY, 6. Macon KT, Mucha PJ, Porter MA (2012) Community structure in the United Nations general assembly. *Phys A: Stat Mech Its Appl* 391(1):343–361
10. Rossetti G, Pappalardo L, Pedreschi D, Giannotti F (2017) Tiles: an online algorithm for community discovery in dynamic social networks. *Mach Learn* 106(8):1213–1241
11. Zakrzewska A, Bader DA (2015) A dynamic algorithm for local community detection in graphs. In: Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM'15). ACM, New York, pp 559–564

12. Ghasemian A, Zhang P, Clauset A, Moore C, Peel L (2016) Detectability thresholds and optimal algorithms for community structure in dynamic networks. *Phys Rev X* 6(3):031005
13. Xu H, Wang Z, Xiao W (2013a) Analyzing community core evolution in mobile social networks. In: Proceedings of the international conference on social computing (SocialCom'13). IEEE, Los Alamitos, CA, pp 154–161
14. Xu H, Xiao W, Tang D, Tang J, Wang Z (2013b) Community core evolution in mobile social networks. *Sci World J* 2013:781281
15. Boyack K, Borner K, Klavans R (2008) Mapping the structure and evolution of chemistry research. *Scientometrics* 79(1):45–60
16. Brodka P, Saganowski S, Kazienko P (2013) GED: the method for group evolution discovery in social networks. *Soc Network Anal Min* 3(1):1–14
17. Casteigts A, Flocchini P, Quattrociocchi W, Santoro N (2012) Time-varying graphs and dynamic networks. *Int J Parallel Emergent Distrib Syst* 27(5):387–408
18. Cazabet R, Amblard F (2011) Simulate to detect: a multi-agent system for community detection. In: Proceedings of the IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology (WI-IAT'11), vol. 2. IEEE, Los Alamitos, CA, pp 402–408
19. Cazabet R, Chawuthai R, Takeda H (2015) Using multiple-criteria methods to evaluate community partitions. [arXiv:1502.05149](https://arxiv.org/abs/1502.05149)
20. Michele Coscia, Fosca Giannotti, and Dino Pedreschi. 2011. A classification for community discovery methods in complex networks. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 4, 5, 512–546
21. Harry Crane and Walter Dempsey. 2015. Community detection for interaction networks. [arXiv:1509.09254](https://arxiv.org/abs/1509.09254)
22. Fortunato S, Hric D (2016) Community detection in networks: a user guide. *Phys Rep* 659:1–44
23. Li XL, Tan A, Philip SY, Ng SK (2011) ECODE: event-based community detection from social networks. In: Proceedings of the international conference on database systems for advanced applications (DASFAA'11), pp 22–37
24. Matias C, Miele V (2016) Statistical clustering of temporal networks through a dynamic stochastic blockmodel. *J Royal Statistical Soc: Series B (Statistical Methodology)* 79(4):1119–1141
25. Rossetti G (2017) RDYN: graph benchmark handling community dynamics. *J Complex Netw* 5(6):893–912. <http://dx.doi.org/10.1093/comnet/cnx016>
26. Rossetti G, Pappalardo L, Rinzivillo S (2016) A novel approach to evaluate community detection algorithms on ground truth. In: *Complex networks VII*. Springer, pp 133–144
27. Rosvall M, Bergstrom CT (2010) Mapping change in large networks. *PLoS ONE* 5(1):e8694
28. Xie J, Kelley S, Szymanski BK (2013) Overlapping community detection in networks: the state-of-the-art and comparative study. *ACM Comput Surv* 45(4):43
29. Xie J, Szymanski BK (2013) LabelRank: a stabilized label propagation algorithm for community detection in networks. In: Proceedings of the 2nd network science workshop (NSW'13). IEEE, Los Alamitos, CA, pp 138–143
30. Xu KS, Hero AO (2014) Dynamic stochastic blockmodels for time-evolving social networks. *J Selected Topics Signal Process* 8(4):552–562
31. Yang J, Leskovec J (2014) Structure and overlaps of ground-truth communities in networks. *Trans Intell Syst Technol* 5(2):26
32. Yang J, Leskovec J (2015) Defining and evaluating network communities based on ground-truth. *Knowl Inf Syst* 42(1):181–213

# A Low-Profile Compact Ultra-Wideband Antenna for Wireless Applications



Preeti Pannu and Devendra Kumar Sharma

**Abstract** The proposed design is composed of simple ultra-wideband (UWB) monopole antenna. The structure consists of semi-rectangular ground plane and rectangular patch on opposite side of the FR4 substrate. In order to achieve ultra-wideband characteristics, half ring-shaped tapered microstrip feed is utilized at the patch. By introducing a square slit at lower edge of the rectangular patch, wide bandwidth can be easily realized. The bandwidth can be further enhanced by creating defect in ground structure. The most impressive feature of the proposed design is its compactness ( $17 \times 25 \text{ mm}^2$ ). Also, this antenna structure has reflection coefficient  $S_{11} < -10 \text{ dB}$  and voltage standing wave ratio  $VSWR \leq 2$  in the whole realized band from 2.1 to 15.8 GHz. The overall antenna structure is simulated on substrate having dielectric constant 4.4. The bandwidth of proposed antenna is wide enough to be used in wireless communication applications such as Wi-MAX (3.3–3.8 GHz), WLAN (5.1–5.8 GHz), and X-band communication systems (7.25–7.75 GHz).

**Keywords** UWB · MIMO · VSWR · DGS

## 1 Introduction

UWB antennas are extremely in demand for their low transmission energy, high data rate, high image resolution, high target characterization, etc. These characteristics make them most attractive for UWB communication. UWB systems make use of very small pulses for transmission and reception. UWB based systems which got license by FCC (Federal communications commission) in 2002 provide very high bandwidth, low gain, and low cost and utilize for multiple wireless applications in the range of 3.1–10.6 GHz. UWB systems include various wireless applications such as 3.6 GHz IEEE 802.11y Wireless Local Area Networks (3.65–3.69 GHz), 5 GHz IEEE 802.11a/h/j/n WLAN (5.15–5.825 GHz), and 7.5 GHz X-band (7.1–7.9 GHz) [1]. Also, UWB antennas are gaining high attention in biomedical field

---

P. Pannu (✉) · D. K. Sharma

Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Delhi-NCR Campus, Ghaziabad, India

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication*

*Engineering*, Lecture Notes in Networks and Systems 106,

[https://doi.org/10.1007/978-981-15-2329-8\\_35](https://doi.org/10.1007/978-981-15-2329-8_35)

[2, 3]. In study, various patch shapes (square, rectangular, circular, octagonal, and elliptical) with various feeding techniques (such as CPW, microstrip line, proximity feed) are proposed for obtaining ultra-wideband [4–8]. In literature, various methods to obtain ultra-wideband are discussed. The most common method is the use of defected ground structure (DGS). This includes cutting of symmetric half-circular slits from rectangular ground plane [9], T-shaped inverted slot in the ground structure [10], hexagonal slot in the ground plane [11], triangular notch loaded defected ground plane [12], and open T-shaped slot [13]. Another approach is modification in radiating element by insertion of slots and slits such as O-slot in rectangular monopole [14], heart-shaped patch [15], dumb-bell-shaped defected metallic structure [16], and Koch fractal geometry in patch [17]. In order to enhance antenna bandwidth, the third method focuses on modifications in feeding structure which includes semicircular radiator fed by tapered feed line [18], and modified feed wrench in shape [19]. Moreover, fractal-shaped geometry is also surveyed to obtain the UWB response with low iterations [20, 21]. In [22], low-profile heart-shaped antenna consists of triangular-shaped patch is proposed but it does not able to achieve the lower cutoff frequency of UWB. Referenced in [23], proposed monopole antenna has wide operating range from 2.7 to 25 GHz but it was large in size. Therefore, it is quite difficult to design compact UWB antennas and to achieve its lower cutoff frequency. Here, a compact rectangular-shaped radiator with square slit at its lower edge for UWB communication is communicated. The proposed design is composed of defected ground structure in order to achieve wide bandwidth as well as lowest cutoff frequency of UWB antennas. Also, a half ring-shaped feed line is used to achieve wideband characteristics. Various parameters such as S11 parameter, voltage standing wave ratio, and impedance bandwidth are analyzed. Table 1 shows comparison in terms of size and bandwidth with some previous reported designs.

**Table 1** Comparison of proposed layout of UWB with other reported

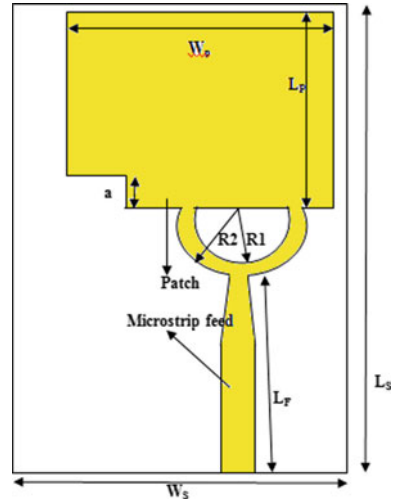
References	Size ( $W \times L$ geometry) (in mm <sup>2</sup> )	Operating band (in GHz)
[23]	36 × 43	2.7–25
[22]	33 × 35	3.5–20
[17]	22.5 × 22.5	2–10.6
[16]	32 × 25	2.85–15.25
[15]	39 × 36.6	3.5–17.8
[14]	29 × 17	2.9–10.70
Proposed	17 × 25	2.1–15.8



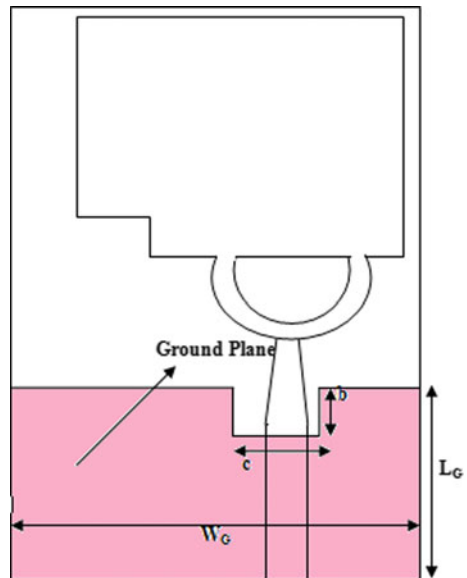
## 2 Antenna Layout

The planned layout is demonstrated in Fig. 1. It consists of rectangular-shaped patch (highlighted with yellow color) with common ground structure (pink color) as represented in Fig. 2. The radiator is fed by  $50 \Omega$  tapered microstrip line of size  $(L_F \times W_F)$ . The proposed antenna design has total area of  $17 \times 25 \text{ mm}^2$  (i.e., denoted as  $L_S \times W_S$ ) and has printed on low loss FR4 substrate with 4.4 relative permittivity, 0.02 loss

**Fig. 1** Proposed geometry highlighting patch



**Fig. 2** Geometry of ground plane



**Table 2** Parameters values used for proposed antenna

Parameter	Values (in mm)	Parameter	Values (in mm)
$L_s$	25	$L_P$	10.4
$W_s$	17	$W_P$	14.5
$a$	2	$R1$	3.2
$R2$	4.4	$L_f$	10.6
$c$	3.5	$b$	2
$L_G$	9.1	$W_F$	2.4

tangent, and 1.6 mm thickness. The overall wideband is achieved through modified feed line. Major design issue of UWB antennas is to get resonance at lower/upper cutoff frequency and keeping compactness of the layout. This can be obtained by increasing current path in radiator, so we take slot in radiator as well as in ground.

This design provides wideband of 2.1–15.8 GHz and also maintains compactness of the antenna. Designing and optimization are done by using high-frequency structure simulator (HFSS). The optimized parameters are given in Table 2.

### 3 Results and Discussion

#### 3.1 Results

Figure 3 represents the return loss characteristics S11 in plot 1 and VSWR in plot 2 of the UWB proposed design. The reflection coefficient S11 is less than  $-10$  dB for entire ultra-wideband (2.1–15.8 GHz) as well as it shows dual-band characteristics.

In plot 1, maximum return loss of  $-15$  and  $-22$  dB is obtained at 2.9 GHz and 13.4 GHz, respectively. In plot 2, Voltage Standing Wave Ratio is less than equal to 2, considered most suitable value for most antenna applications for entire operating band. The utilization of half ring-shaped microstrip line provides improvement in bandwidth.

#### 3.2 Various Design Stages

The evolutionary stages involved in proposed structure are shown in Fig. 4, whereas Fig. 5 shows corresponding return loss graphs. In step 1, simple rectangular radiator, half ring-shaped microstrip line, and simple ground plane are used (shown as Antenna I of Fig. 4).

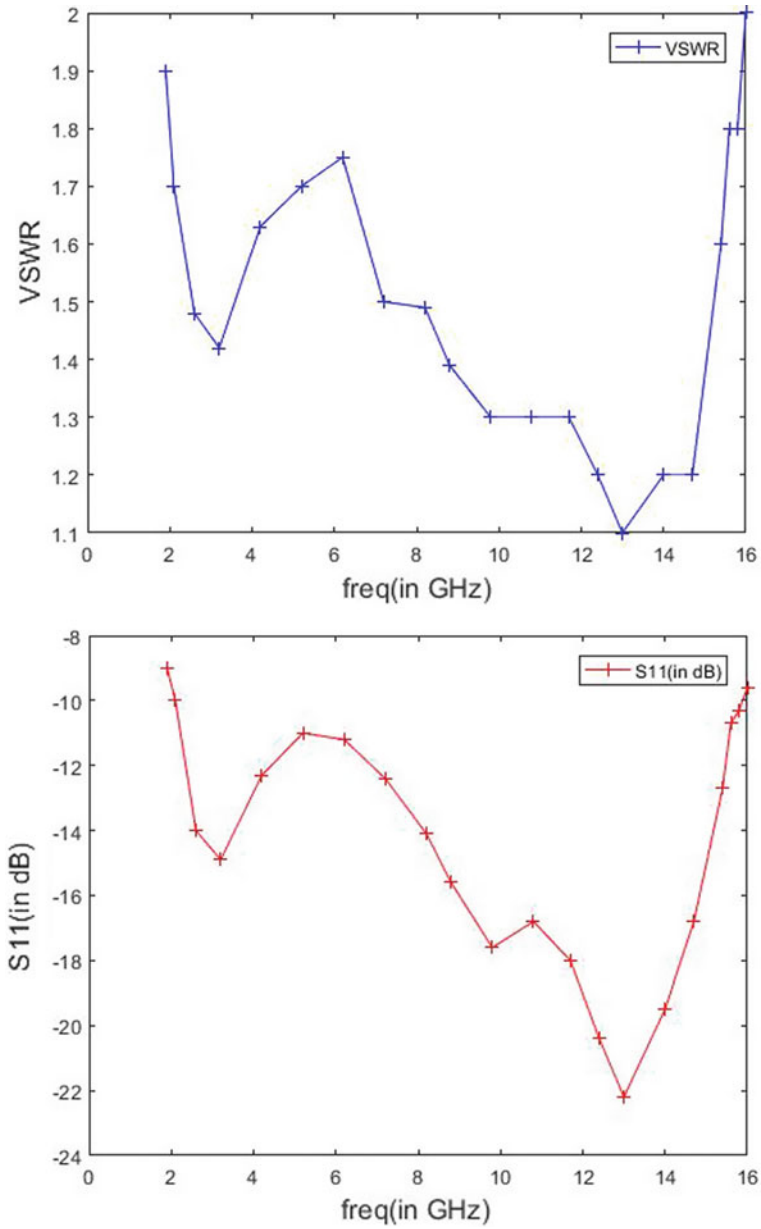
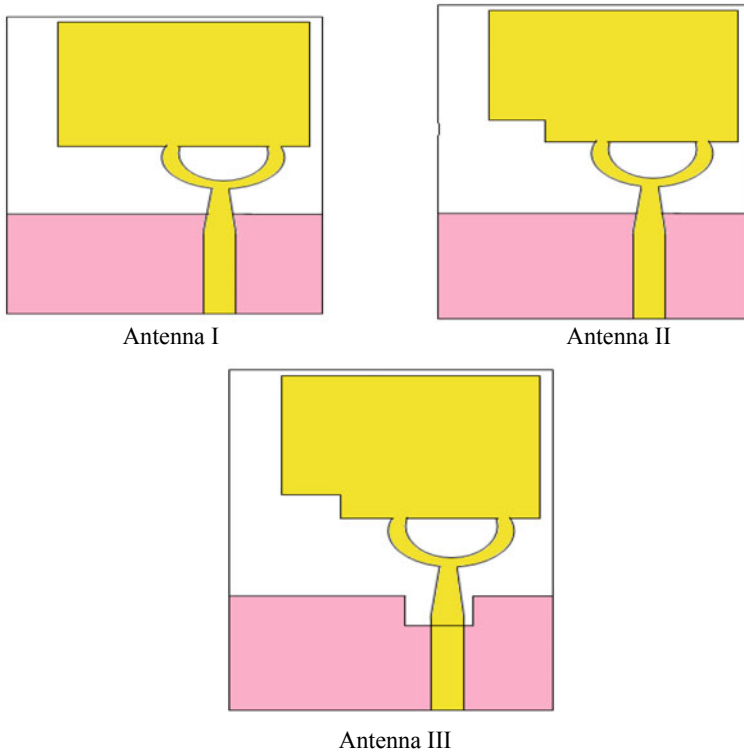


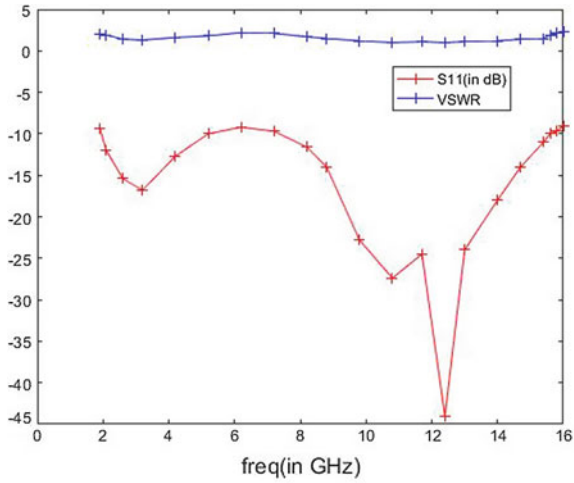
Fig. 3 Simulated results for VSWR and return loss



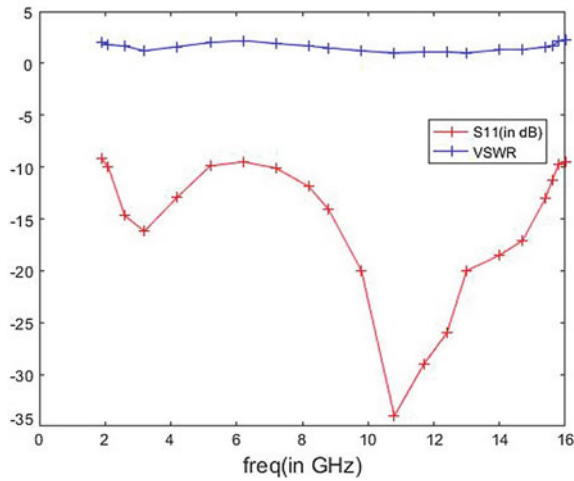
**Fig. 4** Evolution of proposed design

This Antenna I gives a UWB response but not in 5–7.5 GHz band represented in Fig. 5a. In step 2, a square slit from the lower side of the radiator is cut down. This design (Antenna II) gives entire UWB range but not covering band ranging from 6 to 7.5 GHz (Fig. 5b). Finally, by using asymmetric rectangular slot of ( $b \times c$ ) in the ground plane, Antenna III is obtained which covers entire wideband. This spectrum gives an impedance bandwidth varying from 2.1 to 15.8 GHz. In Fig. 5a, Antenna I shows triple-band characteristics with return loss of  $-16.8$  dB,  $-24$  dB, and  $-44$  dB at 3 GHz, 10.2 GHz, and 12 GHz, respectively, whereas Antenna II shows dual-band characteristics with return loss of  $-16.2$  dB and  $-25$  dB at 3 GHz and 12 GHz, respectively.

The best suitable geometry for wideband is Antenna III with lower cutoff frequency 2.1 GHz and covering entire UWB band, whereas in case of Antenna I and Antenna II, it is difficult to maintain wideband at 5–7.5 GHz frequency band. The reason behind this is distribution of different surface current due to different geometry of ground plane and patch. The current distribution of Antenna III is represented



(a) Comparison result for S11 and VSWR of Antenna



(b) Comparison result for S11 and VSWR of Antenna

Fig. 5 Return loss and VSWR for different antenna design techniques

(Fig. 6). The current distribution is shown at 6.5 GHz. The surface current distribution of the proposed design changes due to the incorporation of slit in radiator and slot in ground plane.

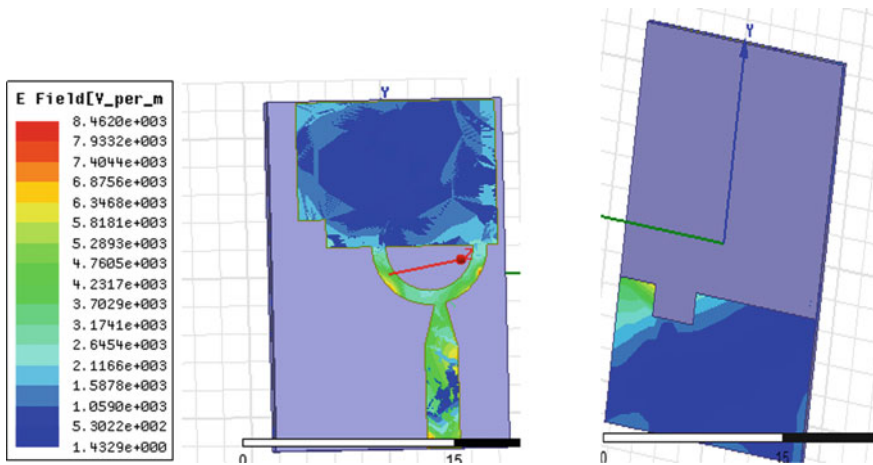


Fig. 6 Current distribution of Antenna III

## 4 Conclusion

A single port miniaturized  $17 \times 25 \text{ mm}^2$  UWB antenna layout is proposed. It has simple design composed of rectangular patch and defected ground structure. The defects in ground plane have great impact on the  $-10 \text{ dB}$  bandwidth for entire UWB. The use of slit in patch provides ultra-wideband characteristics. The insertion of modified tapered feed line further improves the bandwidth. The proposed antenna has  $S_{11} < -10 \text{ dB}$  from 2.1 to 15.8 GHz and  $VSWR < 2$ . The simulated results are quite attractive for various wireless applications.

## References

1. Abbas SM, Ranga Y, Verma AK, Esselle KP (2014) A simple ultra wideband printed monopole antenna with high band rejection and wide radiation patterns. *IEEE Trans Antennas Propag* 62(9):4816–4820
2. Wang F, Arslan T (2017) Body-coupled monopole UWB antenna for wearable medical microwave imaging applications. In: *IEEE-APS topical conference on antennas and propagation in wireless communications (APWC)*, pp 146–149
3. Mahmud M, Islam MT, Samsuzzaman M (2016) A high performance UWB antenna design for microwave imaging system. *Microwave Opt Technol Lett* 58(8):1824–1831
4. Pannu P (2018) A small square monopole ultra-wide band antenna with band stop behavior. *Int J Eng Technol* 7(3.12):697–700
5. Azim R, Islam MT, Misran N (2011) Compact tapered shape slot antenna for UWB applications. *IEEE Antennas Wirel Propag Lett* 10:1190–1193
6. Chandel R, Kumar A (2018) Design and packing of an eye-shaped multiple-input multiple-output antenna with high isolation for wireless UWB applications. *IEEE Trans Compon Packag Manuf Technol* 8(4)

7. Gautam AK, Chandel R, Knaujia BK (2013) A CPW-fed hexagonal shape monopole like UWB antenna. *Microwave Opt Technol Lett* 55(11):2582–2587
8. Raheja DK, Kanaujia BK, Kumar S (2019) Compact four-port MIMO antenna on slotted-edge substrate with dual-band rejection characteristics. *Int J RF Microwave Comput Aided Eng* 1–10
9. Tiwari RN, Singh P, Kanaujia BK (2018) Small-size scarecrow-shaped CPW and microstrip-line-fed UWB antennas. *J Comput Electr*
10. Ojaroudi M, Ghobadi C, Nourinia J (2009) Small square monopole antenna with inverted T-shaped notch in the ground plane for UWB application. *IEEE Antennas Wirel Propag Lett* 8:728–731
11. Palaniswamy SK, Selvam YP, Gulam Nabi Alsath M, Kanagasabai M, Kingsly S, Subbaraj S (2017) 3-D Eight-port ultrawideband antenna array for diversity applications. *IEEE Antennas Wirel Propag Lett* 16:569–572
12. Sahoo S, Mishra LP, Mohanty MN, Mishra RK (2018) Design of compact UWB monopole planar antenna with modified partial ground plane. *Microwave Opt Technol Lett* 60:578–583
13. Sim CYD, Chung WT, Lee CH (2010) Compact slot antenna for UWB applications. *IEEE Antennas Wirel Propag Lett* 9:63–66
14. Chen S, Pan T, Yan Z (2018) Flexible ultra-wideband rectangular monopole antenna with O-slot insertion design. *Sci China Inf Sci* 61
15. Rahman N, Islam MT (2018) The broken heart printed antenna for ultrawide-band applications. *IEEE Antennas Propag Mag*
16. Salamin MA, Ali W, Zugari A (2018) Design and analysis of a miniaturized band-notched planar antenna incorporating a joint DMS and DGS band rejection technique for UWB applications. *Microsyst Technol*
17. Tripathi S, Mohan A, Yadav S (2015) A compact Koch fractal UWB MIMO antenna with WLAN band rejection. *IEEE Antennas Wirel Propag Lett* 14:1565–1568
18. Nikolaou S, Abbasi MAB (2017) Design and development of a compact UWB monopole antenna with easily-controllable return loss. *IEEE Trans Antennas Propag* 65(4):2063–2067
19. Ellis MS, Zhao Z, Wu J, Nie Z, Liu QH (2014) Unidirectional planar monopole ultrawideband antenna using wrench-shaped feeding structure. *Electron Lett* 50(9):654–655
20. Gupta M, Mathur V (2017) Wheel shaped modified fractal antenna realization for wireless communications. *AEU-Int J Electron Commun* 79:257–266
21. Qrazi H, Soleiman H (2017) Miniaturization of UWB triangular slot antenna by the use of DRAF. *IET Microwaves Antennas Propag* 11(4):450–456
22. Zhang Y, Li C (2015) Design of small dual band-notched UWB slot antenna. *Electron Lett* 51(22):1727
23. Kaziel S (2016) Structure and design optimization of compact UWB slot antenna. *Electron Lett* 52(9):681–682

# Weather Monitoring System Using Smart Sensors Based on IoT



Suresh Kumar, M. A. Ansari, Shivam Pandey, Pragati Tripathi and Mukul Singh

**Abstract** The system obtained in this paper is an advanced result for monitoring weather conditions in certain places and anywhere in the world. Prepared with physical equipments, cars, structures and sensors, electronics, software, and other network-related matters make these items in statistical collection and conversation. Sensor control through environmental monitoring information and sensor temperature, humidity, light intensity, rainfall, MQ-135 pollution sensor, and BMP180 air pressure sensor send information to web pages and then sensor data as graph data. The latest data design of the implemented system can enter the Internet anywhere in the world.

**Keywords** GSM SIM 900A module · ATmega328P · DHT11 sensor · LDR sensor · Raindrops module · MQ-135 gas sensor · BMP180 sensor

## 1 Introduction

In an IoT enabled weather surveillance system, the microcontroller applies six weather parameters using each sensor. These sensors are temperature and moisture sensor (DHT 11 sensor), light sensor, rain surface sensor, pollution sensor (MQ-135 gas sensor), and air pressure sensor (BMP180). Then send these parameters to the Internet via IoT techniques. The process of sending data to the Internet using the GSM SIM900A module is repeatedly repeated over a time interval. Then users need to look at some of the Web sites to view this data [1]. The microcontroller is the MCU, which receives and processes the humidity and temperature data from the DHT11

---

S. Kumar (✉) · M. A. Ansari · S. Pandey · P. Tripathi · M. Singh  
Department of Electrical Engineering, Gautam Buddha University, Greater Noida, UP, India  
e-mail: [suresh.gbu@gmail.com](mailto:suresh.gbu@gmail.com)

M. A. Ansari  
e-mail: [ma.ansari@ieee.org](mailto:ma.ansari@ieee.org)



sensor and gives it to the GSM SIM900A module. The GSM SIM 900A module is a Wi-Fi module, which is one of the major Internet platforms. It can send data to IoT cloud [2, 3]. With the reduction in the cost of devices supporting Wi-Fi, this trend will only accelerate [4]. In cloud services, people can use this data to communicate in a variety of ways, such as using the bell or sending them an email or sending an SMS [5]. As mentioned earlier, IoT only allows human-to-human interaction but does not interact with the human-to-device and device-to-device interaction [6]. Local levels and major changes in energy consumption will become an integral part of any country's energy policy [7].

## 2 Internet of Things (IoT)

Internet of things (IoT) is a related quantity environment that is connected to the Internet. In IoT, "objects" can be cars with a core monitor or built-in sensor, e.g., things are given to IP addresses and they have the capacity to accrue data on the network [8, 9]. Technology embedded in objects helps work together with the core state or external environment, which impacts choices [10]. Entrenched (such as sensors) by electronics, Internet connectivity and other types of hardware can interconnect with others on the Internet and IoT can control tenuously. There are technical changes in the information commerce in rappings of the Internet (IoT) [11, 12] (Fig. 1).

## 3 Implementation

IoT-based weather monitoring system with smart sensors completes implementation setup, and components can be described in Table 1. The complete implementation setup of IoT-based weather monitoring system with smart sensor is shown in Fig. 2.



**Fig. 1** Internet of things (IoT)

**Table 1** Rating of components

S. No.	Components	Description
1	GSM SIM900A module	Operating voltage +5 V, used for establishing communication
2	Microcontroller	Supply voltage +5 V, CPU type 8-bit AVR, pin count 28, max. operating frequency 20 MHz
3	Voltage regulator	Power supply +5 V, use a simple feed-forward
4	DHT11 sensor	Operating voltage 3.5–5 V, used for temperature and humidity
5	Raindrops sensor	Operating voltage 5 V, used for detection of rain
6	LDR sensor	Operating voltage 5 V, used for light
7	MQ-135 gas sensor	Operating voltage 5 V, used for air pollution
8	BMP180 sensor	Operating voltage 5 V, used for air pressure



**Fig. 2** Complete setup of the device

## 4 Methodology

In this IoT-based weather monitoring system with smart sensors, it uses different types of parameters such as GSM SIM900A, voltage regulator, and air pressure sensor (BMP180 sensor) which are given below [12].

### 4.1 GSM SIM 900A

A GSM or GPS module is a chip or circuit that will be used to find communication among a mobile device and computing machine. This is the international standard for mobile that is used for communication across a wide range. In a GSM/GPRS modem, you have an internal TCP/IP stack to allow the Internet to connect to the GPRS [13]. It is appropriate for SMS, voice, and database transfer requests in the M2M interface’s SIM900A module as shown in Fig. 3.

### 4.2 Temperature and Humidity Sensor (DHT11)

It feels the temperature around it. This is a four-pin device. This sensor includes a component for measuring resistive type humidity and an NTC temperature measurement component [14]. The DHT11 sensor is shown in Fig. 4.

### 4.3 Raindrop Module

This is used to feel the rain. This can also be used to measure rain intensity. This module treats humidity above the analog output pin and provides a digital output when humidity is high. More water or less resistance means less output voltage. We can stop using this code [15]. The raindrops module is shown in Fig. 5.

### 4.4 Light-Dependent Resistor (LDR) Sensor

The LDR is an element which has variable resistance which changes the severity of light. A light-dependent resistor (LDR) is a lightweight device resistor [16]. The LDR light can be applied to sensitive detectors circuits and switching circuits that turn on by light as shown in Fig. 6.



Fig. 3 GSM SIM 900A



Fig. 4 DHT 11 sensor

**Fig. 5** Rain level sensor



**Fig. 6** Light-dependent sensor (LDR)

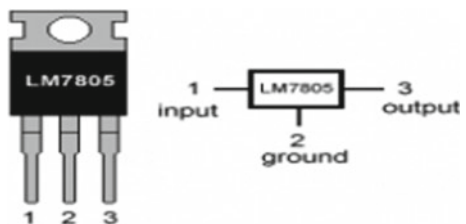


### 4.5 Voltage Regulator

Power supply is an electrical expedient that electronically transfers electromagnetic power. The basic function of the control source is to transform a type of electric energy into other forms, so it is sometimes known as an electrical converter.

Consumer electronics use small and separate power supply AC adapter or different types of electric bricks, wall tablets, or chargers. AC adapter is used with electrical appliances which require electricity, but the power does not contain internal ingredients to achieve voltage and power. Internal power supply internal circuit design is such as used for internal or internal supply [6]. In this paper, a 12VDC adapter to run the setup is shown in Fig. 7.

**Fig. 7** Voltage regulator



**Fig. 8** MQ-135 gas sensor

#### **4.6 Pollution Sensor (MQ-135 Gas Sensor)**

An air pollution sensor is a device that senses and filters air pollution in the immediate area. It can still be used for the inner and outer environment. These sensors can still be finished at home or received from some finished goods. Sensors are very exclusive in the past, but with technical development, the population is more affordable and grows more in all of these sensors. These sensors can help serve many purposes and help focus on environmental problems outside the human eye circle. MQ135 content is  $\text{SnO}_2$ , and this is a special ingredient. MQ-135 gas sensor sensors such as ammonia nitrogen, oxygen, alcohol, vertical compounds, sulfur, and smoke [17]. The MQ-135 gas sensor is shown in Fig. 8.

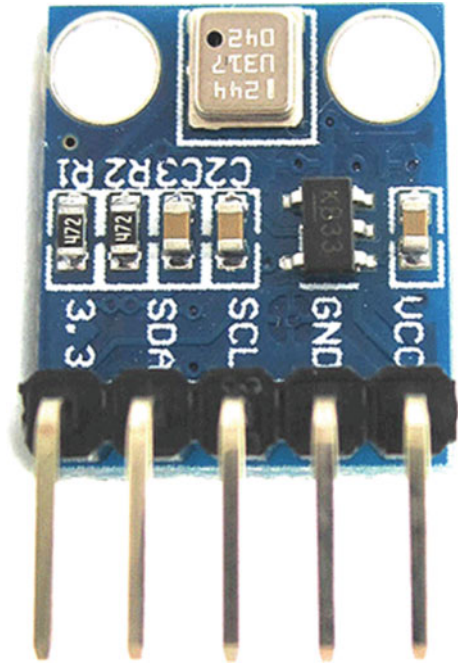
#### **4.7 BMP180 Sensor**

The pressure sensor is expedient for the measurement of gas or liquid pressure. Increasing pressure is a look of power needed to prevent a liquid, and generally, according to the unit per unit area, BMP180 barometer combines pressure, temperature, and height [6, 17–19]. A BMP180 air pressure sensor is shown in Fig. 9.

### **5 Implementation and Working**

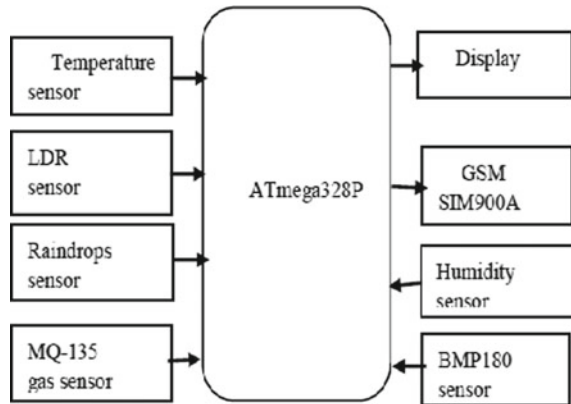
The system implementation involves the ATmega328P microcontroller as the core processing unit for the entire system and all sensors and plans can be connected to the microcontroller. The core connection types are shown in Fig. 2. The second pin is DHT11 which is a detail pin, can mention to temperature and humidity heights up to pin 5 of the microcontroller. Pins 1 and 4 DHT11 are  $V_{cc}$  and  $G_{nd}$ , and Pin 3 is not connected.

Fig. 9 BMP180 sensor



The sda and scl pins are the air pressure sensor pins that are connected to the 28 and 27 pin microcontrollers. The fourth pin and air pollution sensor are  $V_{cc}$  and  $G_{nd}$ , the third pin has no connection, the second pin of the air pollution sensor is the data pin, and it can send the pollution value to the 26 pin microcontroller. Pins 1 and 4 of the rain level module are  $V_{cc}$  and GND, the third pin has no connection, and the second pin of the rain module is a data pin that can send rain values or not to the 24 and 23 pin microcontrollers. The microcontroller produces the temperature and humidity equal and sends it to the Wi-Fi component. The Tx and Rx pins of the GSM SIM900A are connected to the second (Rx) and third (Tx) microcontrollers. Brand certain the GSM SIM 900A input voltage must be 3.3 V, not 5 V (otherwise it will damage the device). The block diagram of the system is given in Fig. 10, and the working flow chart is given in Fig. 11.

**Fig. 10** Block diagram of the system



## 6 Result and Discussion

The flowchart which we have used in the proposed system. First of all, connect to sensors needed for microcontrollers and then process sensor data using microcontrollers and embedded C. Send monitoring parameters (temperature and humidity sensors, LDR sensors, rain sensors, air pollution sensors, air pressure sensors, etc.) to the cloud using modules Wi-Fi. Develop an application to read data from Wi-Fi and save that data to the cloud. Plot the values on the graph and display the monitoring parameters on the web page. In the data source collected by the weather monitoring system, we have developed four graphs showing variations between weather parameters (Fig. 12). The various measurements and sensing may be discussed as follows:

*Case 1—Temperature:* The first graph is plotted of temperature which depends on the conditions of the surroundings. In this graph, temperature is represented with respect to time as shown in Fig. 13. This graph is drawn according to the data available from Fig. 12.

*Case 2—Temperature and Humidity:* The second graph is plotted between temperature and humidity whose values depend on the conditions of the surroundings. When humidity is low, the temperature is high and vice versa as shown in Fig. 14.

*Case 3—Light Intensity:* The day and night conditions are detected by the LDR sensor, and the output graph from the LDR sensor tells what kind of lighting conditions are present in the surroundings as shown in Fig. 15.

*Case 4—Raindrops:* The rain level sensor at the weather monitoring station tells about rain conditions in the environment and is the channel on web page as shown in Fig. 12. Successively, it displays results based on the number of water drops on the rain sensor module.

- a. If the rain device panel really rains, it will start and the “RAINING” determination will still be sent to the consecutive display.

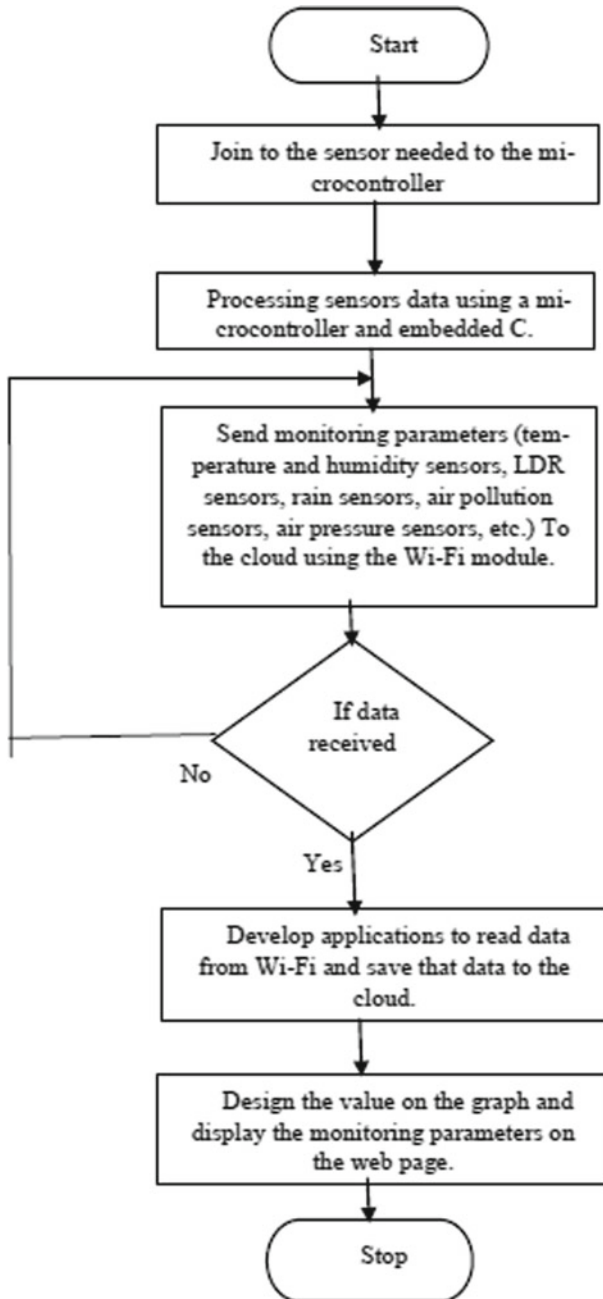


Fig. 11 Flowchart of the working process



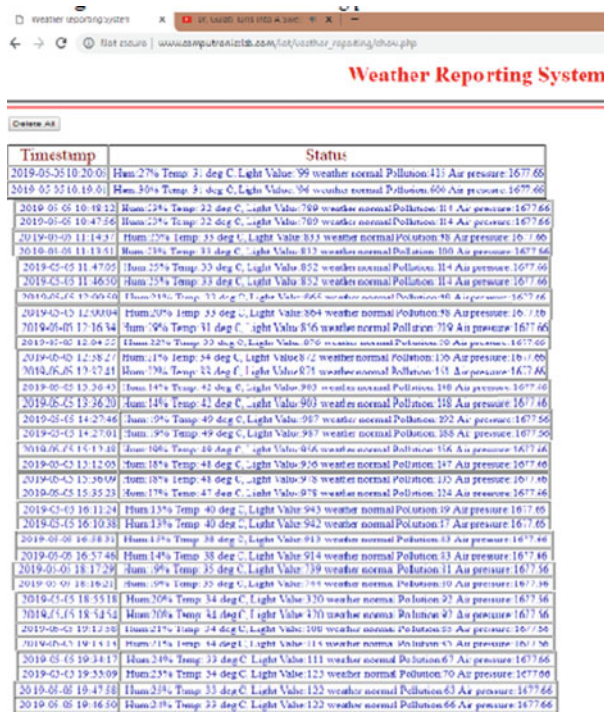


Fig. 12 Channel on web page

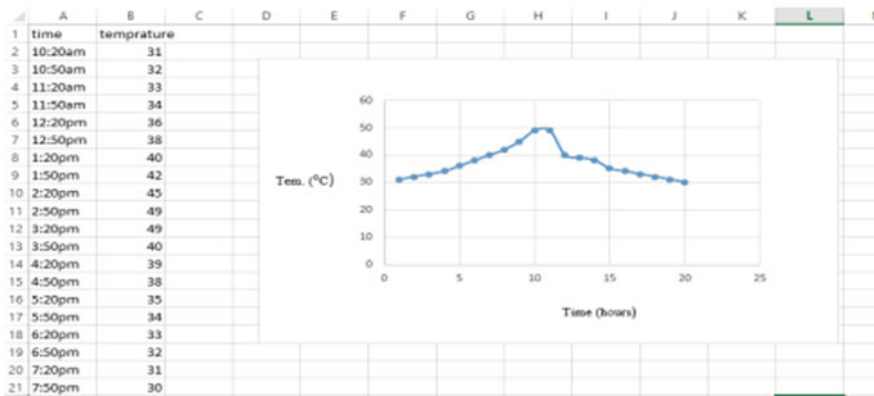


Fig. 13 Representation of temperature

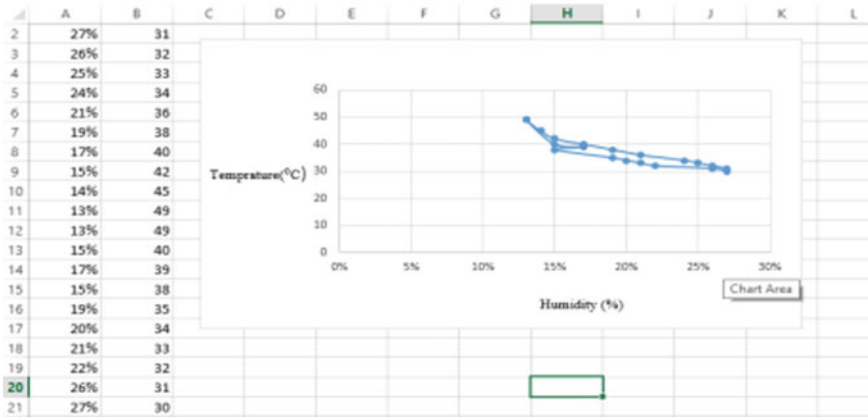


Fig. 14 Representation of temp. and humidity

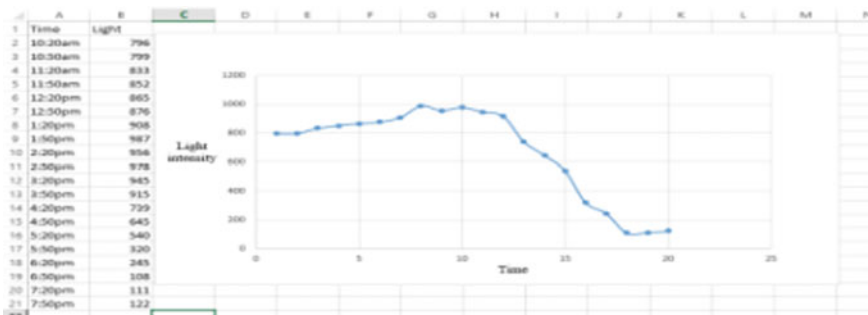


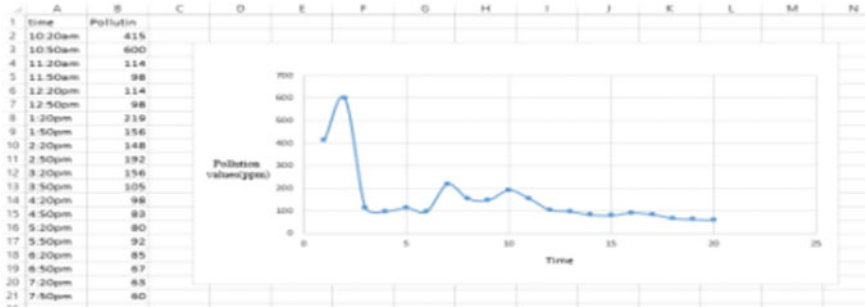
Fig. 15 Representation of lighting data

- b. If the sensor panel has water droplets on it, the solution still starts and “RAIN WARNING” will still be sent to the sequential display.
- c. If the sensor panel dries up, the “NO RAIN” resolution continues to be sent to the sequential display.

Case 5—Air Pollution: The air pollution sensor in the weather monitoring station tells us about the value less than 135 ppm when the pollution is normal and value is greater than 135 ppm when the pollution is high as shown in Fig. 16.

## 7 Conclusion

Based on the IoT-based weather reporting system, which gives us ATmrga328P values with different parameters such as temperature, humidity, light, pollution, air pressure, and rainfall. It can be got wherever from a laptop, PC, or smartphone as we have industrialized an application that classifies evidence. By using sensor devices



**Fig. 16** Representation of air pollution

in the environment, we can convey the environment into real life, interrelating with other objects to complete the network. Then assembly data and study results will be obtainable to end users via Wi-Fi. This perfect can be used to screen effluence in emerging cities and manufacturing sectors.

## References

1. Rao BS, Rao KS (2016) Internet of things (IOT) based weather monitoring system. *Int J Adv Res Comput Commun Eng* 5(9). ISO 3297:2007 Certified
2. Lehmann G, Reefer A, Blumendorf M (2010) A 3-layer architecture for smart environment models. A model-based approach. Labor Technische University Berlin, Germany
3. Asghar MH, Negi A, Mohammadzadeh N (2015) Principle application and vision in internet of things (IOT). In: *International conference on computing, communication automation*, pp 427–431, May 2015
4. Thaker T (2016) Esp8266 based implementation of wireless sensor network with Linux based web-server. *Int J Adv Res Comput Commun Eng*
5. Setiyono B, Sumardi, Harisuryo R (2015) Measurement system of temperature, humidity and air pressure over 433 MHz radio frequency—an application on quadrotor, Oct 2015
6. Ram KSS, Gupta ANPS (2016) IoT based data logger system for weather monitoring using wireless sensor networks. *Int J Eng Trends Technol* 32(2)
7. Xiaojun C, Xianpeng L, Peng X (2015) IOT-based air pollution monitoring and forecasting system. In: *International conference on computer and computational sciences (ICCCS)*
8. Patil PS, Sawant SR, Mudholkar RR (2016) AVR microcontroller based embedded weather monitoring system. *Int J Res Appl Sci Eng Technol* 4(7)
9. Zhou Y, Zhou Q, Kong Q, Cai W (2012) Wireless temperature amp; humidity monitor and control system. In: *International conference on consumer electronics, communications and networks (CECNet)*, pp 2246–2250, Apr 2012
10. Bosisio AV, Cadeddu MP (2015) Rain detection from ground-based radiometric measurements: validation against rain sensor observations. In: *IEEE international geoscience and remote sensing symposium (IGARSS)*, pp 2323–2326, July 2015
11. Salim GM, Ismail H, Debnath N, Nadya A (2015) Optimal light power consumption using LDR sensor. In: *IEEE International symposium on robotics and intelligent sensors (IRIS)*, pp 144–148, Oct 2015
12. Lazarescu MT (2013) Design of a WSN platform for long-term environmental monitoring for IoT applications. *IEEE J Emerg Sel Topics Circuits Syst* 3(1):45–54

13. Xie H, Ma F, Bai Q (2009) Prediction of indoor air quality using artificial neural networks. In: International conference on natural computation (ICNC 2009), pp 414–418
14. Hasler A, Talzi I, Tschudin C, Gruber S (2008) Wireless sensor networks in permafrost research—concept, requirements, implementation and challenges. In: Proceedings of 9th international conference on permafrost, vol 1, pp 669–674, June 2008
15. Romer K, Matter F (2004) The design space of wireless sensor networks. *IEEE Wireless Commun* 11(6):54–61
16. Tozlu S, Senel M, Mao W, Keshavarzian A (2012) Wi-Fi enabled sensors for internet of things: a practical approach. *IEEE Commun Mag* 50(6):134–143
17. Munandar A, Fakhrurroja H et al (2018) Design of real-time weather monitoring system based on mobile application using automatic weather station. In: 2nd International conference on automation, cognitive science, optics, micro electro-mechanical system, and information technology (ICACOMIT), Jakarta, pp 44–47, Jan 2018
18. Rohini, Ansari MA, Pal NS (2019) Remote monitoring of vital parameters with IoT based sensing system. In: 3rd International conference on micro-electronics and telecommunication (ICMETE-2019)
19. Kumar D, Ansari MA (2018) Condition monitoring of electrical assets using digital IRT and AI technique. *J Electr Syst Inf Technol* 5(3):623–634

# Brain Tumor Detection Using Image Processing Based on Anisotropic Filtration Techniques



Aditya Garg, Aditya Bajaj and Roshan Lal

**Abstract** Whether it is common cold or something as big as a brain tumor, a timely detection of a disease goes a long way in curing of a disease or even the survival of the patient. There are just so much more options available at the initial stage than at the last. There are various ways to detect brain tumor such as neurological exams, computerized tomography (CT), positron emission tomography (PET), and magnetic resonance imaging which have various steps that if not applied carefully may or may not yield helpful results, and these steps are pre-processing which includes noise removal, image enhancement, filtering, edge detection, etc., segmentation, feature extraction such as thresholding and image subtraction, and finally, area estimation. For filtering, we are going to be using anisotropic diffusion filtering techniques which reduces the contrast with nearby neighboring pixels.

**Keywords** CT · PET · MRI · Clustering techniques · Anisotropic filter

## 1 Introduction

### 1.1 Magnetic Resonance Imaging

It is generally known as MRI which is a diagnostic technique that utilizes the properties of magnetism and radio waves to develop detailed images of the various organs of human body. The magnetic field aligns the protons present in hydrogen atoms of our body, which on interaction with the radio waves results in a spinning of protons

---

A. Garg (✉) · A. Bajaj · R. Lal  
Amity University, Noida, UP, India  
e-mail: [adityagarg011@gmail.com](mailto:adityagarg011@gmail.com)

A. Bajaj  
e-mail: [adityabajaj63@gmail.com](mailto:adityabajaj63@gmail.com)

R. Lal  
e-mail: [rhhokar@amity.edu](mailto:rhhokar@amity.edu)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_37](https://doi.org/10.1007/978-981-15-2329-8_37)

creating a faint signal, and these signals are then used by the computer to develop an image. The images are quite detailed and can help in detection of even the slightest of changes in the structure.

## ***1.2 Tumor of the Brain***

Consider a large accumulation of waste, now think of this in terms of a human body. Accumulation of abnormal cells in a part of a body is tumor. Normally, cells in our body have a fixed life cycle which ends with their death, but some cells have their life cycles altered due to radiations or some other unknown reasons. As a result, instead of dying they just keep on existing without any use, and as time goes by, they keep on adding resulting in a mass called tumor.

## ***1.3 Problem Definition***

MRI is an imaging technique used to detect disorders/abnormalities of our body. MRI utilizes strong magnetizing field and bursts of radio waves to cause excitations within the protons/hydrogen atoms present in our body. Protons when hit with the radio waves of selected frequencies get excited and start resonating, thus disturbing the magnetic field. On coming back to this ground state, these protons release energy which gets caught by the receiver present within the scanner.

As different parts of the body resonate at their own specific frequency, we can have slices of images for the body to make up a whole picture. Similarly, different tissues have a different rate of relaxation (i.e., time taken by protons to get back to the ground state). Two times are measured for every scan: First is the T1 which is the time duration required for the magnetic alignment to fall back to the relaxing state, while the other is T2 which is the time required by the spin to get back to the resting phase.

More than one type of radio waves can be used in conjunction to highlight or hide different types of tissues. Consider, for example, fat tissue layers usually which do not have any kind of abnormalities, so together with the area-specific frequency, we use a fat suppressing pulse to remove signal emitted by that fat layer.

## ***1.4 Project Definition***

**The first step** in image processing is pre-processing, which involves basic operations such as contrast enhancement, noise removal, deblurring, smoothing, and so on.

**The second step** consists of two steps: thinning and thresholding. Thresholding is the process of converting any form of an image to a binary image, and this is done by

considering a threshold value and comparing the value at each pixel to that threshold. Thinning is used to reduce the thickness of the boundaries to just a pixel thick this helps in clearly marking the boundaries.

**The third step** is filtering; it is used to remove small uncertainties or deformities at pixel level by giving the whole image a certain texture.

**The fourth step** is morphological operations; it is in this step that we scan the density of the image to determine whether the brain has a tumor or not.

**The fifth step** is marking the tumor, outlining, and overlapping the tumor region on the original input image.

**The sixth and final step** is bounding the tumor in a bounding box for easier detection.

## 1.5 Objective

The objective of this report was to scan an input image for tumor detection and marking of that tumor over the original input image. These image processing techniques are used in MRI for tumor detection in brain. The techniques used in these image processing are thinning, thresholding, anisotropic diffusion filter,  $k$ -means segmentation techniques, and morphological operations [1].

## 2 Literature Review

In this section, we are going to be discussing the various brain tumor detection and segmentation techniques used for MRI scans for brain tumor detection. Several studies have been performed and are reported in literature for detection of brain tumor. Some of the works are as follows:

In year 2010, Ehab F. Badran, Esraa Galal Mahmoud, and Nadder Hamdy published a work, “An Algorithm for Detecting Brain Tumors in MRI Images.” In this paper, they proposed a computer-based method that can be applied on MRI images of a brain for defining the tumor region and classification of tissues. This computer-based method uses various operations like pre-processing, image segmentation, feature extraction, and classification of different types of tumor with the help of neural network and artificial intelligence.

In year 2010, T. Logeswari and M. KARNAN described their work in, “An Enhanced Implementation of Brain Tumor Detection Using Segmentation Based on Soft Computing.” They have described in detail the process of segmentation by dividing the process into two phases. In the first phase, MRI image of brain is collected. Using the pre-processing technique, image is then converted into standard form. In the second phase for image segmentation, hierarchical self-organizing map (HSOM) method is applied on image [2].

## 3 Research and Methodology

### 3.1 *Limitations to Existing Systems*

#### 1. **Only large tumors can be diagnosed**

We get very clear and vivid images of the tissues, but with the present methods, it is very difficult to detect tumors or cancer cells at an early stage, when the tissue is small.

#### 2. **Image is not 100% accurate**

We use different filtering techniques to normalize the uncertain pixels. Thus, there is some loss of data which leads to a not so accurate image.

### 3.2 *Solutions in Proposed System*

#### 1. **Detection at an early stage for tumors**

As we are using differential filtering and  $k$ -means clustering techniques, we are going to be taking in small clusters of cells at a time. Thus, even the small tumors or cancer cells can be detected before they develop into a full-blown tumor.

#### 2. **Better and accurate image after pre-processing**

With the use of anisotropic diffusion filter, we are reducing the blurring effect caused by the other filters such as the gaussian blur filter generally used in image processing. This improves the quality of image significantly and helps us in achieving accurate images after pre-processing.

#### 3. **Region-based growing**

It is very important to know whether the tumor is growing or not, and in order to get some information about this, we are using  $k$ -means clustering technique that selects various seed locations and then observes the neighboring cells and study them. Then it decides whether to include them in the cluster or not.

## 4 Proposed System

### 4.1 *Getting an MRI Image*

First and foremost, important thing we need for image processing is an image, and for doing so, we have various commands in MATLAB. The one we are going to use is **uigetfile**, applying these commands opens up a dialog box which asks you to select a file from one of the pre-decided formats. Once the file is selected, its address gets stored in a local variable that refers to the image wherever it is used/called upon.



## 4.2 Anisotropic Diffusion Filter

The anisotropic diffusion filter is an iterative method deployed to restore the intensities of MR images. This is achieved with the help of edge-stopping functions, and the anisotropic diffusion filter is far more capable than any of the previously used filtering techniques such as the gaussian filter, gobel filter, and so on.

At its definitive form, anisotropic diffusion filter or ADF is a type of a repetitive algorithm that mimics the diffusion process as follows:

$$I_s^{t+1} \approx I_s^t + (\lambda/|\eta_s|) \sum_{p \in \eta_s} \left[ g \left( \left| \nabla I_{s,p}^t \right|, \gamma \right) \nabla I_{s,pt} \right] \quad (1)$$

where  $I_s^t$  represents the intensity of a pixel (S) from an image ( $I$ ) at any instant  $t$ ,  $\lambda$  is a scalar that relates to the rate of diffusion,  $\gamma$  is a +ve constant selected in accordance with the need of smoothness,  $\eta_s$  represents the set of neighboring pixels of (S),  $g(\cdot)$  is an edge-stopping function, and  $\nabla I_{s,p}^t$  represents the magnitude of an image directional gradient from pixel (S) to (P) at any instant  $t$ . To simplify the process for understanding, we will replace  $\nabla I_{s,pt}$  with  $N$  whenever the pixel information and the number of repetitions are irrelevant to the context of the image. Noise pixels get filtered at a much more faster rate than the pixels that make up the edge for two main reasons. First is that the gradients of the opposite direction nullify their effects, in accordance with the equation and that usually happens to the pixels at the edge, but this is not the case with the noise pixels. Second is that anisotropic diffusion filter has an effect that can be considered big enough only when the number of neighbors with high gradient magnitude in the same direction is large. As noise pixels are not collected at a particular region and more evenly spread throughout the image, thus keeping their intensity distinct from the neighboring pixels. Edge pixels, on the other hand, have a much more similar pixel signature to most of their neighboring pixels [3].

## 4.3 Image Thresholding

Image thresholding is the simplest form of image segmentation, and using thresholding what we are doing is converting the image into a binary format. What happens in binary format is that each pixel can have either of the two states, either it can be in the foreground (i.e., defining the picture), or it can be in the background (i.e., helps in defining the boundaries of the image). This helps us to remove unwanted intrusions that will lead to difficulty in tumor detection later in morphological operations. There are various ways in which we can apply thresholding and MATLAB has some inbuilt thresholding functions, but the general rule for thresholding is

$$\text{If } f(x, y) > T \text{ then } f(x, y) = 0 \text{ else } f(x, y) = 255 \quad (2)$$

What this does is if the value at a particular pixel is less than the threshold value, then that pixel is considered in the background, or else if the pixel value is above the threshold value, then it is considered in the foreground [4].

#### **4.4 Thinning**

One of the most essential parts of image processing is to represent the structural shape of the brain, and for this, we need to have a good way to determine the boundary of our image. One way to accomplish this is by reducing it to a graph. This is done by a process called skeletonization or thinning. We can describe thinning as the extraction of the skeleton up to an extent that the image preserves the points necessary for image processing and reconstruction.

Thinning gives us some pretty good advantages in optimization of our image processing algorithm as it reduces processing time, and unimportant features and noise can be easily filtered out. As the whole amount of data gets reduced, it takes less time for data processing and easy handling of data.

#### **4.5 Segmentation Technique**

Segmentation technique is used to divide the image into various contours for analyses of different abnormalities. For doing so, we apply different clustering algorithms such as  $k$ -means, fuzzy  $c$ -means, hierarchical, mixture of gaussians, and now neural network. Here, we are using the  $k$ -means clustering algorithm, and clustering is a method to divide the total pixels into a specific number of groups. In  $k$ -means, we divide the pixels into exactly  $k$  number of groups, and further, the algorithm is divided into two phases. During the first phase of the algorithm, we calculate all the  $k$ -centroids corresponding to each cluster, and in the second phase, we select each point in the cluster which has nearest centroid to the respective data point. As the grouping is done, we try to find the centroid for each of the new clusters and the above process is repeated again. Thus,  $k$ -means is an iterative method, which minimizes the sum of distance from each object to the center of the cluster centroid.  $k$ -means is easier to implement, but at the time of designing a  $k$ -means algorithm, we should take care of the computation's complexity [5].

### **5 Comparative Study**

Now that we have learned about the various techniques that can be used to implement image processing such as region-based segmentation (i.e., we are applying filtration technique on individual pixel in the neighborhood), clustering-based technique (i.e.,

**Table 1** Comparative analysis

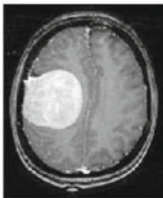
Parameters	Segmentation technique			
	Region-based technique	Clustering-based technique	Threshold-based technique	Anisotropic diffusion filter and threshold-based technique
Processing time	Moderate	Moderate	Low	Moderate to high
Accuracy	Moderate	Moderate	Low to moderate	High
Noise	Low	Low	Moderate	Very low

we are grouping pixels into cluster and then applying filtering techniques onto them), and finally threshold-based segmentation that considers each individual pixel and check it with the threshold limit to see whether to keep it in background or foreground. We are implementing region-based technique which compares the selected pixel with its neighboring pixels and applies filtering technique on them [6] (Table 1).

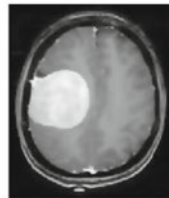
## 6 Result and Discussions

Now that we have read and understood the process of how image processing is done and how we detect tumor. Let us look at some results of how each and every step helps in the detection of tumor and image processing.

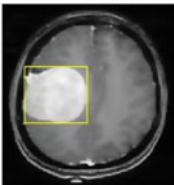
**Input Image**



**Filtered Image**



**Bounding Box**



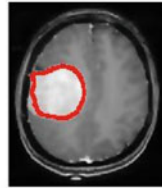
**Tumor Alone**



## Tumor Outline



## Detected Tumor



## 7 Conclusions

### 7.1 Limitations of the System

1. Very small tumors or cancer cells that are just starting to grow are undetectable by this method.
2. Tumors that have holes in them or have hollow body are not yet detectable through this method.

### 7.2 Advantages of the System

1. Gives very accurate nearly 97% accurate image after pre-processing and filtering of image is done.
2. Patient does not have to go through any ionizing radiations.
3. Gives an idea about the regional growth of tumor.
4. Gives a clear and specific enlarged image of the abnormal area.

### 7.3 Implications for the Future

In the future, we would like to further look at different and more advanced segmentation techniques such as mixture of gaussian techniques. We would also like to use artificial intelligence and neural network in helping us in better detection of tumor.

## References

1. Nerurkar SN (2017) Brain tumor detection using image segmentation. *Int J Eng Res Comput Sci Eng* 4(4):25–70
2. Poonam JP (2014) Review of image processing techniques for automatic detection of tumor in human brain. *Int J Comput Sci Mob Comput* 3(3):371–378
3. Palma CA, Cappabianco F, Ide J, Miranda P (2014) Anisotropic diffusion filtering operation and limitations—magnetic resonance imaging evaluation. In: *IFAC proceedings volumes (IFAC-PapersOnline)*, 19
4. Natarajan P, Krishnan N, Kenkre NS, Nancy S, Singh BP (2012) Tumour detection using threshold operation in MRI brain images. In: *2012 IEEE international conference on computational intelligence and computing research, Coimbatore*, pp 1–4. <https://doi.org/10.1109/iccic.2012.6510299>
5. Norouzi A, Rahim MSM, Altameem A, Saba T, Rad AE, Rehman A, Uddin M (2014) Medical image segmentation methods, algorithms, and applications. *IETE Tech Rev* 31:199–213. <https://doi.org/10.1080/02564602.2014.906861>
6. Shishodiya D, Shukla S, Soni A, Singh SA (2018) Comparative study of brain tumor detection in MRI images. *ISSN (Online): 2319–7064*

# Sales Terminal Interactive Device for Disabled People



Priyam Shah, Harsh Patel, Roshni Rao and Manan Shah

**Abstract** ‘Disability’ is a state of body where it lacks any vital senses or organ functioning which limits a person’s usual activities or movements. Due to a disability, any individual had to suffer a lot, as a disabled person had to be dependent on some other individual. It is generally observed that disability not only affects their physical condition, but it also exacerbates their mental condition as they are treated differently than normal humans in society. Being treated differently by society also affects their psychological health which is generally seen in disabled children. Several times, it is quite difficult to find teachers who are prepared for students with special needs in every geographic area which overall affects their education. The current work proposes how modern hardware along with the implementation of appropriate technology can be made useful to disabled people up to a great extent. The present invention relates to devices operable by the deaf, mute or people having low vision. The present invention more particularly relates to a multi-use simple, low cost and efficient point of sale terminals operable by the deaf, mute or people having low vision. The device can also be implemented at several public as well as private platforms where there is limited interaction. The device has an enormous scope of implementation at a departmental store, shopping malls and movie theatres, where range to employ such PWD person at the job position like cashier, store inventory manager and at several other places using present invention. The device is customizable, easy to use, efficient, economical and robust.

**Keywords** Disability · Deaf · Mute · Low vision person · IOT · Sales terminal · Employment

---

P. Shah (✉) · H. Patel

Department of Computer Engineering, LJ Institute of Engineering and Technology,  
Ahmedabad, India  
e-mail: [priyamshah14697@gmail.com](mailto:priyamshah14697@gmail.com)

R. Rao

School of Liberal Studies, Pandit Deendayal Petroleum University, Gandhinagar, India

M. Shah

Department of Chemical Engineering, School of Technology, Pandit Deendayal Petroleum University, Gandhinagar, Gujarat, India

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_38](https://doi.org/10.1007/978-981-15-2329-8_38)

375

## 1 Introduction

Approximately, 15% of the world's population experience some or other forms of disability and generally developing countries experience a higher rate of disability between 110 million and 190 million people, i.e. approximately one-fifth population experience significant disabilities. Government of India is a signatory to the promulgation adopted in the meeting to launch the Asian and Pacific decade of disabled persons year 1993–2002 at Beijing from 1st to 5th year December which had to implement a law for the benefit of the disabled citizen. Hence, the persons with disabilities (Equal Opportunities, Protection of Rights and Full Participation) act 1995 were passed in the parliament [1]. Disability includes vision impairment, deaf or hard of hearing, physical disability, mental health condition and acquired brain injury. Present work focuses on three types of disability, i.e. deaf, mute and low vision. Technological inventions are rapidly advancing in every field. However, surprisingly there are very few researches or inventions which are conducted for disabled people.

Young people transition from a state of the dependent minority to an independent adult identity has been measured in terms of a developmental stage model. The key markers of adulthood have commonly been regarded as leaving full-time education and entering the labour market, moving out of the parental home to establish an independent household [2]. However, in the case of deaf and mute people, they lack quality knowledge and education which have direct effects on their upcoming future. Generally, it is seen in disabled children who have limited long-term memory, and, in normal education sessions, they experience the difficulty to retain a learned concept. It is necessary to design certain activities for disabled children which on the one end introduce novelty and create new challenges, and on the other end include moments in which children repeat and consolidate what they have previously understood [3].

Disable person finds quite difficult to get proper employment which affects their social life as well and if they can find a job then sometimes it is seen that these jobs are below dignity jobs which overall sabotage their confidence towards work. However, it is understood that disable people are more focused as well as dedicated to their jobs. Due to the expected deterioration in the working-age group, especially in European countries, people with disabilities are now more often acknowledged as a valuable resource in the workforce, and research work in the field of disability and their employment facility is more important than ever [4].

POS is an efficient method to automate the checkout method, providing faster and better customer experience. It can be found almost in every retail store, restaurants, clubs and bars and supermarkets [5]. Point of sale (POS) terminals have been in the market for many years now but there are hardly any devices or inventions which enable a low vision, or a deaf or mute operator to interact with customers without the need of an additional helper. If a person has low vision, is deaf or mute, it is very difficult for him to get employment in any industry as communication with them is crucial, and thus, the person has to be dependent on some alternative sources of

income which may be not sufficient to earn his livelihood which leads them to happy life.

The present paper discusses regarding the device which is specially designed for mute, deaf person people, and the device has been developed which enables a person with relatively low vision (not blind) to view the screen placed at the sales terminal (counter) of shopping malls, theatres, departmental stores, offices, etc. The device has 'braille' engraved on it so such low vision people can also operate the system efficiently. Systems have been developed which enable a person with low vision (not blind) to view the screen placed at the point of sale counter of departmental stores, shopping malls, offices, etc. Such people can be trained to use braille script also. The device is designed keeping in mind the needs of disabled people so that they can easily learn and operate the device without any technical expertise.

## 2 Technology Aspects and Methodology

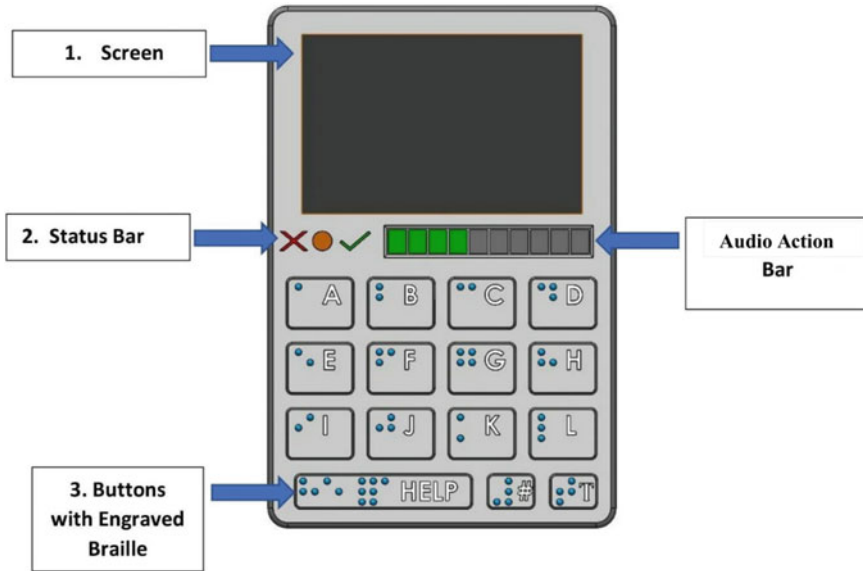
Improvements for varied human and organizational problem-solving attempts, through the design, development and implementation of system based on certain technological aspects along with enhanced efficiency and effectiveness of information for strategic, tactical and operational situations are termed as 'Technology'.

'Smart billing system' has a great implementation in the shopping malls involving some automation in the customarily practiced process. Malls and supermarkets are self-service shops offering a wide variety of food and household products. Moreover, the system developed by using Raspberry processor has excellent efficiency as well as the productivity, and also this system reduces the overall time and increases the productivity of the store [6]. Python is a programming language that helps you to integrate systems more effective and lets you work swiftly; also Python provides excellent support of 'Tkinter Library' which will be helpful for the system [7].

### 2.1 Screen

The screen will display the output from the processor. Screen size will be 5" inch LCD screen. The screen will display the same message which is played in the speaker on pressing the corresponding button. The screen will also project those inputs which are given from screen-2. The screen will also include GUI and animation for better understanding and representation to PWD person (Fig. 1).





**Fig. 1** Design of the device which is operated by the PWD person

## 2.2 Status Bar

The functionality of status bar is specially designed for the deaf person. The status bar contains three lights red, orange and green. The status bar will display the light according to the input given on screen-2. Suppose the customer standing at the point of sale terminal is not willing to take specific goods, then he will press the 'No' button and immediately on the status bar red light will be displayed. So, the deaf person can get the idea that the customer is not willing to take that product/goods. On the other end, if the customer standing at the sales terminal presses the 'Yes' button, the green light will be displayed on the status bar. Also, the orange light indicates that the customer standing at the sales terminal has not given any input (Fig. 1).

## 2.3 Buttons with Engraved Braille

Visual impairment may result in a restricted field of vision or in a diminished ability to see the sharpness of detail, read standard-size print, determine colour or depth perception, see contrasts, adjust to changes in light glare or locate objects. Visual loss can affect a person's daily activities, leisure pursuits, education, vocation and social interaction [8]. According to standards of Indian Government, the person who has 40% or more blindness comes in the category of 'low vision'. Thus, braille enables the low vision person to use the device with great ease. There are several

buttons on the device; on pressing each button an audio recording will be played. Audio recording is to help the mute person to interact with the customer standing at the sales terminal. Every button is engraved with the braille [A–Z] or [1–9] so that the person who has low vision can recognise each button and use the system with ease (Fig. 1).

### 2.4 Audio Action Bar

Functionality of the audio action bar is placed for convenience of the deaf person as the deaf person is unable to listen any voice in the surrounding. However, he will not be able to listen when the audio clip is played whenever he presses any button of the device. Thus, the audio action bar gets incremented automatically along with the audio so whenever the audio is completed the audio status gets ‘full green’ so the deaf person gets idea that audio clip has been completed (Fig. 2).

This customer-end screen, i.e. (screen-2) at the billing terminal from which person who is standing at the counter for billing purpose can interact with the PWD person side device. This device contains a 7" touch screen which will display all the messages and instructions, and even the audio clip which is played is printed on the screen (Fig. 2).

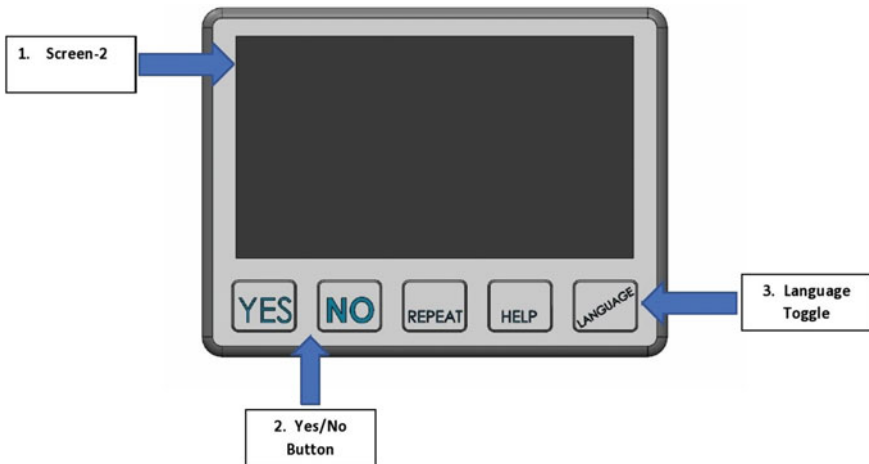


Fig. 2 Screen which will guide the customer of the shop at sales terminal

## **2.5 Yes/No Button**

Yes/No button has significance to help customer standing at the sales terminal to provide his input to the PWD person standing as a cashier or store manager. Suppose some audio clip is played 'Asking if customer needs a carry-bag or not?' then question arises who a person will who is deaf will get an idea if customer needs the bag or not! In that case if customer presses Yes/No button, the result will be displayed on the status bar of device used by PWD person.

## **2.6 Language Toggle**

Language toggle button is placed for the ease of the customer standing at billing terminal. Using language toggle button, customer standing at billing terminal can select the preferred language in which they want to interact. The audio clip will also be played in the same language which customer has selected prior, and the screen will also display the message in same language which customer has selected. Thus, it overall increases the customer satisfaction while interacting with device, and also the market can easily adopt the system.

## **3 Comparative Analysis**

Traditionally, there are certain stores across the nation who have employed PWD person at the billing counter or at the sales terminal of their respective stores. Traditionally, a play card or selection card is placed on the counter, and the customer needs to point the goods/items they are willing to purchase, according to that the PWD person will generate a bill and place the order. But, there are certain back drops in this system such as the PWD will not be able to properly understand the customer's needs. In traditional methodology, low vision person who are 40% blind will not be able to work at such platforms as the low vision person will not be able to see the things clearly! (Fig. 3).

Contrary to traditional methodology, the present invention has higher benefits as the system completely digitalized the sales terminal. System also provides language selection option which is helpful for customer so that they can communicate in their preferred language, and the system also incorporates certain features which are helpful for deaf/mute and low vision person to work on it as discussed earlier. System can also be implemented at store inventories, billing terminal and with cash counter attendant. Present invention also has certain functionality which reduces time of both



**Fig. 3** Deaf/mute lady working as a counter attendant/cashier at leading food chain KFC. *Source* <https://www.thestar.com.my/news/community/2014/07/22/disabled-shine-at-work-impairments-do-not-deter-them-from-excelling-in-their-duties>; <http://azonlifestyle.blogspot.com/2016/01/kentucky-fried-chicken-operated-by-deaf.html>



**Fig. 4** Author's invention

customer and the PWD person attending them. Audio action bar, language toggle, status bar and button with engraved braille are not possible in traditional cardboard or play card methodology.

Thus, using the present invention can help PWD person can get a better job opportunity to work at places where customer interaction plays a crucial role. And the device is economical as well as feasible for corporation which implements at their desk (Fig. 4).

## **4 Economic Analysis**

The present paper has provided a detailed economic analysis of the devices and points presented which has application for deaf, mute and low vision people describing the various aspects of the potential business benefits and opportunities that will lead to the economic benefit of the product. Cost-benefit analysis (CBA) is widely used administrative evaluation tool, though academics remain sceptical. This volume collects prominent contributors from law, economics and philosophy for discussion of cost-benefit analysis, specifically its moral foundations, applications and limitations [9]. Moreover, the cost-benefit analysis (CBA) which is also stated as benefit-cost analysis (BCA) is a systematic approach for strengths and weakness estimation to determine options and to provide the best feasible approach for achieving benefits along with preserving savings.

### ***4.1 Raw Materials Cost***

Basic raw material costs provide a better indication for the resource use within an economy than the production and material costs registered by producers. The basic costs do not include the added value from component-supplier production chains. In addition, basic raw material costs also are a better indicator for the potential behavioural effects of pricing such resources [10].

The raw material cost consists of the cost of the processor, cost of the LCDs, cost of power supply unit (PSU), cost of 3D body printing, cost of designing and manufacturing custom PCB board and circuitry for the device. Cost also consists of designing and manufacturing custom button which has braille embedded on it so it is helpful for low vision person.

### ***4.2 Utilities and Operating Labour Cost***

The cost of utilities (CUT) includes the cost of maintenance as well as the cost of labour. Proper maintenance for any device is a necessity for efficient and fault-free functioning. The requirement of an on-board support team can resolve any issue in the device or any sort of hardware failure (Fig. 5).

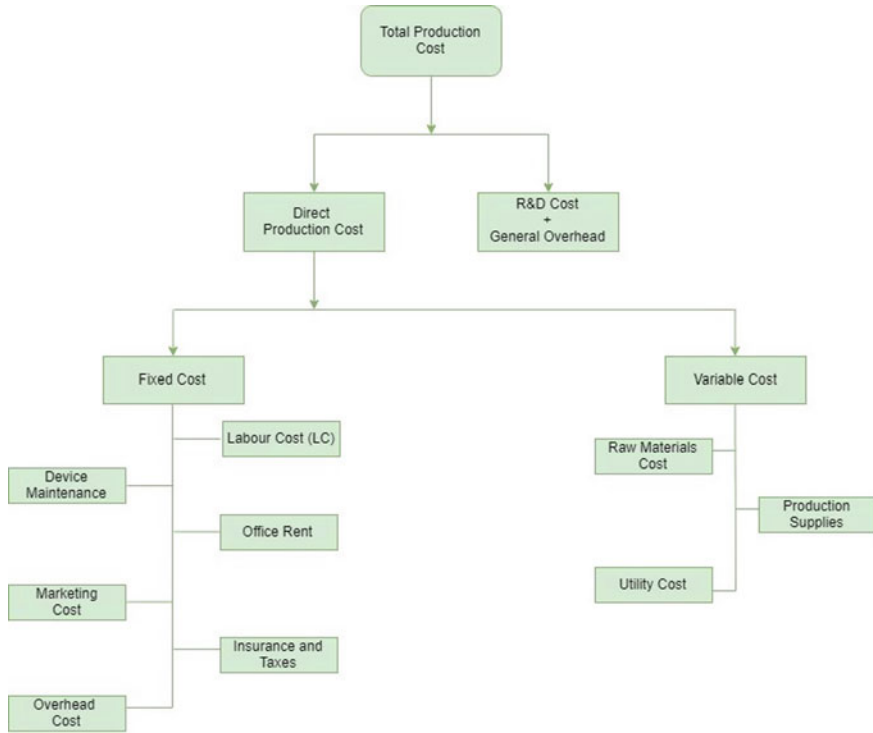


Fig. 5 Costing hierarchy

## 5 Challenges

Challenges faced by disabled workers are lack of understanding of the needs of employees with disabilities in the workplace. It is generally seen that most of the employers were unaware with and insensitive to the needs of their workers with disabilities. Work environment for disability friendly hindered the movement of employees with disabilities [11].

The main challenge which author will encounter is to train such PWD person to work with the device. Team needs to conduct separate training programmes for mute, deaf and low vision person so they can be trained to work with our device. Now, the device will act as a mediator between the PWD person standing as a cashier or any such platform where customer interaction is essential and the consumer who is there to bill his purchased goods or trying to interact with PWD person standing in front of him.

### Future Trends

Emerging future trends in the interaction of identity and disability are a homogeneous whole. There are certain issues whose identity relates to the type of impairment experienced with disable people and the range which impairment covered by ‘disability’

is enormous. Though categorising disabilities is problematical, this following paragraph will focus on three types of disability in order to demonstrate how different impairments can relate to issues of identity: intellectual disabilities and deafness.

The device provides an interconnection between all the systems working within a specific area to provide the performance boost to system functioning so the author will try to create a private network for the system which can enable the interconnectivity between them. Also, the devices for store inventory authors are planning to apply several 'data analytics algorithm', to keep a track record about goods sold and to provide a statistical report for the sales of a particular entity in the company. Thus, the PWD personal can act as a helping hand for the store's inventory management which is overall beneficial for the companies [12, 13].

## 6 Conclusions

The paper has presented the work on the device which proves to be useful for the disabled people; the paper also includes several technical aspects methodology along with the proposed design of the device and comparison with traditional techniques. The device lets the disabled people work at the sales terminal of the store where customer interaction is a crucial factor. The device economic feasibility is presented along with the estimated costing, as the costing factor is crucial for market survival. Hence, the device proves to be beneficial as well as sustainable to the global market.

**Acknowledgements** The authors are grateful to LJ Institute of Engineering and Technology and Pandit Deendayal Petroleum University for the permission to publish this research.

**Authors Contribution** All the authors make substantial contribution in this manuscript. PS, HP, RO and MS participated in drafting the manuscript. PS and HP wrote the main manuscript; all the authors discussed the results and implications on the manuscript at all stages.

**Availability of Data and Material** All relevant data and material are presented in the main paper.

**Competing Interests** The authors declare that they have no competing interests.

## References

1. Chandrashekar H, Naveen Kumar C, Prashanth NR, Kasthuri P (2010) Disabilities research in India. *Indian J Psychiatry* 52(Suppl 1): S281–S285
2. Valentine G, Skelton T (2007) Re-defining 'norms': D/deaf young people's transitions to independence. *Sociol Rev* 55(1):104–123
3. Garzotto F, Bordogna M (2010) Paper-based multimedia interaction as learning tool for disabled children. In: *Proceedings of the 9th international conference on interaction design and children*, pp 79–88
4. Vornholt K, Villotti P, Muschalla B, Bauer J, Colella A, Zijlstra F, Ruitenbeek GV, Uitdewilligen S, Corbière M (2017) Disability and employment—overview and highlights. *Eur J Work Organ Psychol* 27(1):40–55

5. Khaneja D (2017) SYSC 5708 model-driven development of real-time and distributed software. Point of Sales Terminal (Post), pp 1–24
6. Saraf PP, Khanolkar AD, Pawar AS, Ghodake UB, Salokhe SJ, Patil GA (2019) D2D smart billing system. *Int J Adv Res Ideas Innov Technol* 5(2):1204–1207
7. Pastell M (2016) Teaching instrumentation and data analysis using python. In: Cigr-Ageng 2016: international conference on agricultural engineering
8. Beaver KA, Mann WC (1995) Overview of technology for low vision. *Am J Occup Ther* 49(9):913–921
9. Adler MD, Posner EA (2001) Cost-benefit analysis: legal, economic and philosophical perspectives, pp 1–22
10. Narayanan S (2018) A study on challenges faced by disabled people at workplace in Malaysia. In: *Proceeding—5th Putrajaya international conference on children, women, elderly and people with disabilities*, pp 185–197
11. Gwernan-Jones R (2008) Identity and disability: a review of the current state and developing trends. *Beyond Current Horizons*, pp 1–19
12. Kakkad V, Patel M, Shah M (2019) Biometric authentication and image encryption for image security in cloud framework. *Multiscale Multidiscip Model Exp Des* 1–16. <https://doi.org/10.1007/s41939-019-00049-y>
13. Jha K, Doshi A, Patel P, Shah M (2019) A comprehensive review on automation in agriculture using artificial intelligence. *Artif Intell Agric* 2:1–12

### ***Web Links***

14. <https://www.thestar.com.my/news/community/2014/07/22/disabled-shine-at-work-impairments-do-not-deter-them-from-excelling-in-their-duties>
15. <http://azonlifestyle.blogspot.com/2016/01/kentucky-fried-chicken-operated-by-deaf.html>



# Ranking of E-Commerce Sites in India Using Decision-Making Approach



Kajal Sharma and Sanjay Kumar Dubey

**Abstract** The advent of the technological advancements has set up a stage for multifarious online businesses in India, therefore, changing the course of the ordinary shopping trends in the direction of online shopping. India has a total of 560 million end users of the Internet and has been emerged as the second largest applicant in this domain. As per a study in 2011, the retail market of India was estimated at 470 billion dollars. The E-commerce sites cater the innumerable needs of the shoppers. From such a big list of these sites, to pick one, it is termed as a multi-criteria decision-making (MCDM) problem which makes it hard for the online shoppers to come to a conclusion that which site is the best to go with. With the objective of exercising judgment upon the ranking of the E-commerce sites so as to free the shoppers from the tedious job of selection of a particular Web site, this paper uses AHP method used to assign the criteria weights which in turn assist in generating the ranking of the Web sites. The end result is substantiated by a dint of entropy method calculations.

**Keywords** E-commerce · AHP · Entropy · MCDM

## 1 Introduction

E-commerce is an acronym for electronic commerce that deals with goods and services over the Internet with the support of the electronic mediums. It cannot be merely termed as buying and selling items online rather involves a chain of numerous processes such as developing, marketing, selling, delivering and paying [1]. India witnessed the inception of the E-commerce via IRCTC that enabled an online ticket booking system and since then, it has become one of the largest growing sectors in India [2].

---

K. Sharma (✉) · S. K. Dubey

Department of Computer Science and Engineering, Amity University Uttar Pradesh, Sec.-125, Noida, UP, India

e-mail: [seaokajal@gmail.com](mailto:seaokajal@gmail.com)

S. K. Dubey

e-mail: [sanjukundan@gmail.com](mailto:sanjukundan@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication*

*Engineering*, Lecture Notes in Networks and Systems 106,

[https://doi.org/10.1007/978-981-15-2329-8\\_39](https://doi.org/10.1007/978-981-15-2329-8_39)

The abrupt increase in the Internet users has significantly contributed to the growth of E-commerce as well, recording 6 million new entrants per month. Rising from a market worth of about \$3.9 billion way back in 2009, the growth hits the figure of \$12.6 billion in 2013. The online retail revenue of India is expected to cross the magical figure of \$100 billion by the year 2020. The motivation behind the retail chains to get into this online business is that it gives the customers the liberty to shop anytime from anywhere. According to an analysis, the 4% of India's GDP share is going to be from E-commerce alone by 2020 [3].

The upward growth trajectory of an E-commerce site not only relies upon a single factor, rather its success is a blend of numerous significant factors [4]. In this paper, four major factors that are inescapable for governing the success of the Web sites are selected. These factors were picked after a survey was conducted with a couple of people including the expert web developers, along with the regular online shoppers.

The uncertainty is usually dealt with MCDM methods. The variants of the MCDM methods available are analytic hierarchy process (AHP in short), fuzzy set theory, analytic hierarchy process (ANP in short), fuzzy TOPSIS, compromise programming (CP) and Elimination Et Choix Traduisant la Realite (ELECTRE), etc. In this paper, AHP is being used that was proposed by Saaty and is often termed as Saaty method [5]. Instead of designating the exact decision, AHP searches for the best alternative to meet the target requirements.

### ***1.1 Development of the Hierarchy Model***

The MCDM method involving AHP is adopted to rank the sites. The problem is fragmented into certain levels which inclusively fabricate a hierarchical structure. This hierarchical model is constructed by keeping the goal at the first level and further the criteria or sub-criteria and the alternatives at the second and third level, respectively [6] (Fig. 1 and Table 1).

### ***1.2 Factors for Determination of the Best E-Commerce Site***

- A. **Approachability or Comfort of Use (Fc1)**—The cognitive effort to use a Web site must be reliable. The relevant factors like checkout process, transaction options, browser compatibility and embedded interactive content must be taken into consideration. According to a recent study by MarketingSherpa, it has been concluded that 60% of the customers abandon their shopping carts before the checkout process due to the complications embedded in them [7]. The interactive web content in India has risen by 155% over the past two years.

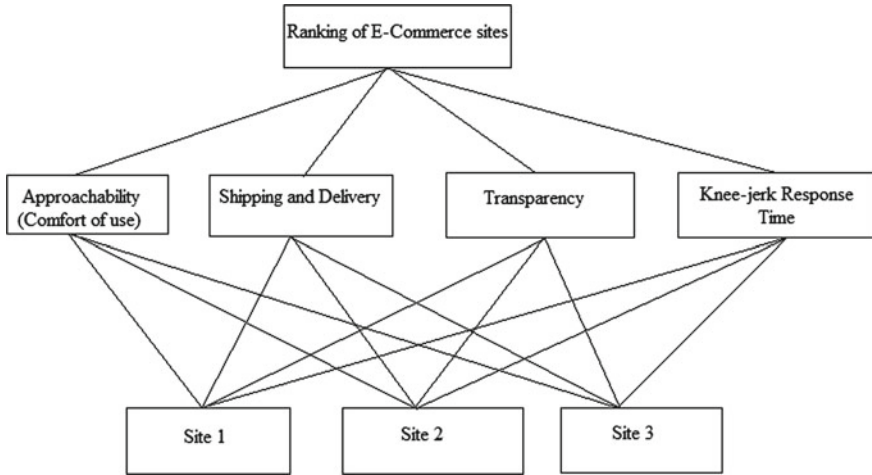


Fig. 1 Proposed hierarchical structure

- B. **Shipping and Delivery (Fc2)**—The safe and quick delivery is the vital factor on which the customers rely upon. During the shipment process, the products are more vulnerable to damages and ensuring a secure delivery to the customers becomes crucial. Low-cost shipment plays a vital role in ensuring satisfaction to the end users.
- C. **Transparency (Fc3)**—The online reviews by genuine customers only must be published on the Web sites. Generally, 50% of the customers like to see the reviews before jumping onto the decision of buying a certain product [7]. The radical essence of customer satisfaction is the transparency of a particular site [8].
- D. **Knee-jerk Response Time (Fc4)**—It is an arduous fact that the Web sites usually tend to lose 40% of the customers if the site takes more than 3 s to load [7]. Slow speed of a site discourages a buyer and affects the rankings too.

**Table 1** Selected criteria and sub-criteria

S. No.	Criteria	Sub-criteria	Checklist
1.	Approachability or comfort of use	i. Checkout process ii. Transaction options iii. Browser compatible iv. Embedded interactive content	i. How convenient the checkout process is? ii. Whether the site offers multiple transaction options? iii. Is the site compatible to the browser? iv. The content makes the user interaction trouble-free?
2.	Shipping and delivery	i. Safe delivery ii. Quick delivery iii. Low-cost shipment iv. Cashless proceedings	i. The delivered product is damage-free? ii. Site offers hasty delivery system for its customers? iii. Does the site offer fair shipment price? iv. Are there cashless proceedings available?
3.	Transparency	i. Online reviews ii. Product value iii. Reporting process iv. Product images	i. Are the reviews published on the site are by a genuine consumer? ii. Is the product value demanded at the time of checkout process is same as it was displayed initially? iii. Does the site provide updates on product tracking before it is shipped? iv. The product received by the consumer is same as shown in the image?
4.	Knee-jerk response time	i. Query interrogation ii. Load time of site iii. Product navigation iv. Home page downloading	i. How fast the site resolves the query: what the consumer is trying to find? ii. How much time does the site take to load? iii. How effortlessly the consumer is able to find the product demanded? iv. How quick the homepage is downloaded?

## 2 Adopted Methodology

AHP is one of the most efficient tools in the scenarios involving complex decision-making problems. Sooner after its evolution, the method was adopted in the military analysis [5]. The process initiates with the pair-wise comparison of the alternatives based on some shortlisted factors [9].

### 2.1 Implementation of AHP and Its Calculations

See Tables 2, 3, 4, 5, 6 and 7.

**Table 2** Pair-wise comparison matrix for approachability (*Fc1*)

	Site 1	Site 2	Site 3	<i>n</i> th root	E.V (W)
Site 1	1	3	5	2.466	0.637
Site 2	1/3	1	3	1	0.258
Site 3	1/5	1/3	1	0.405	0.105
Sum				3.871	1.000

$$(\lambda_{\max})_{\text{avg}} = 3.037, \text{C.I} = 0.0185, \text{C.R} = 0.0319$$

**Table 3** Pair-wise comparison matrix for shipping and delivery (*Fc2*)

	Site 1	Site 2	Site 3	<i>n</i> th root	E.V (W)
Site 1	1	9	5	3.557	0.751
Site 2	1/9	1	1/3	0.333	0.070
Site 3	1/5	3	1	0.843	0.178
Sum				4.733	0.999

$$(\lambda_{\max})_{\text{avg}} = 3.029, \text{C.I} = 0.0145, \text{C.R} = 0.025$$

**Table 4** Pair-wise comparison matrix for transparency (*Fc3*)

	Site 1	Site 2	Site 3	<i>n</i> th root	E.V (W)
Site 1	1	4	5	2.714	0.674
Site 2	1/4	1	3	0.908	0.225
Site 3	1/5	1/3	1	0.405	0.101
Sum				4.027	1.000

$$(\lambda_{\max})_{\text{avg}} = 3.086, \text{C.I} = 0.043, \text{C.R} = 0.074$$

**Table 5** Pair-wise comparison matrix for knee-jerk response time (*Fc4*)

	Site 1	Site 2	Site 3	<i>n</i> th root	E.V (W)
Site 1	1	1	7	1.913	0.515
Site 2	1	1	3	1.442	0.388
Site 3	1/7	1/3	1	0.362	0.097
Sum				3.717	1.000

$(\lambda_{\max})_{\text{avg}} = 3.086, C.I = 0.043, C.R = 0.074$

**Table 6** Pair-wise comparison matrix for criteria

	Fc1	Fc2	Fc3	Fc4	<i>n</i> th root	E.V
Fc1	1	2	1/3	1/4	0.639	0.116
Fc2	1/2	1	1/6	1/8	0.319	0.058
Fc3	3	6	1	1/2	1.732	0.314
Fc4	4	8	2	1	2.828	0.513
Sum					5.518	1.001

$(\lambda_{\max})_{\text{avg}} = 4.021, C.I = 0.006, C.R = 0.006$

**Table 7** List of indices assigned

	Fc1	Fc2	Fc3	Fc4	Index assigned
$W_p$	0.116	0.058	0.314	0.513	–
Site 1	0.637	0.751	0.674	0.515	0.593
Site 2	0.258	0.070	0.225	0.388	0.304
Site 3	0.105	0.178	0.101	0.097	0.104

### 3 Validation of Result

The result can be verified by employing Shannon’s entropy method (Table 8).

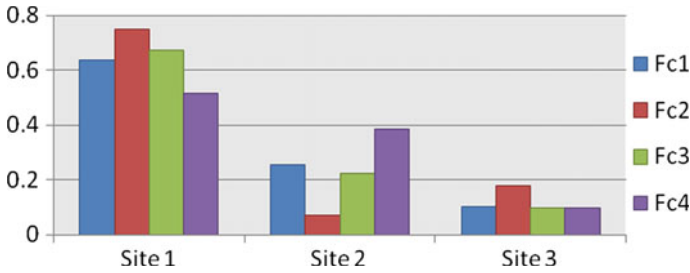
**First Step:**

As a first step, a matrix of criteria weights for each factor is made: [10, 11].

$\beta = 1/\ln(n)$ , which is a constant value. Therefore,  $\beta = 1/\ln(3) = 0.91023$ .

**Table 8** Determinant matrix

	Site 1	Site 2	Site 3
Fc1	0.637	0.258	0.105
Fc2	0.751	0.070	0.178
Fc3	0.674	0.225	0.101
Fc4	0.515	0.388	0.097



**Fig. 2** Dependency of each site on Fc1, Fc2, Fc3 and Fc4 (factors) results in providing the ranking of the Site 1, Site 2 and Site 3 (E-commerce sites) respectively

**Table 9** List of ranks

	Site 1	Site 2	Site 3
$R^3I_i$ values	0.66992	0.19966	0.13001
Ranking	I	II	III

**Second Step:**

$S_i = 0.91023 \sum_{k=1}^n (\alpha * \ln(\alpha))$ , where  $\alpha = W_{rc}$ :  $r$  = no. of rows and  $c$  = no. of columns [11, 12]. Hence,  $S_1 = 0.79505$ ,  $S_2 = 0.64482$ ,  $S_3 = 0.75830$ ,  $S_4 = 0.85142$ .

**Third Step:**

$C_i = [1 - S_i]$ . ( $C_1 = 1 - 0.79505 = 0.20495$ ;  $C_2 = 0.35518$ ,  $C_3 = 0.2417$ ,  $C_4 = 0.14858$ ).

**Fourth Step:**

In this step, the corresponding weights are computed:  $W_{t_i} = C_i / \sum C_i$  [11].

$W_{t_1} = 0.21564$ ,  $W_{t_2} = 0.37371$ ,  $W_{t_3} = 0.2543$ ,  $W_{t_4} = 0.15633$ .

**Fifth Step:**

It consists of calculation of ranks and henceforth validating outcome [11].

$R^3I_i = \sum A_{ij} * W_{t_{ij}}$  for all  $i, j = 1$  to  $k$ . ( $R^3I_1 = 0.66992$ ,  $R^3I_2 = 0.19966$ ,  $R^3I_3 = 0.13001$ ) (Fig. 2 and Table 9).

**4 Conclusion**

With the objective of evaluating the ranks of the E-commerce sites, this paper uses a technique employing AHP calculations. The entire study is based upon the determination of the factors that can ace the ranks of the sites. A contrast is obtained among the factors on which the various sites depend. Taking those factors into consideration, AHP is being implemented and the rankings are fetched. The latter result is

then verified via entropy calculations. Although, it does not cover the entire pool of such sites but based upon the most crucial factors it has been able to pick the sites that majorly take those factors into consideration. To conclude, it can be inferred that AHP extended with entropy acts as a preeminent approach for the ranking of the E-commerce sites and assists people in selecting the best site to shop online, thereby saving their valuable time.

**Acknowledgements** Authors acknowledged the persons who directly/indirectly helped to complete the research paper.

**Conflict of Interest** The authors declare that they have no conflict of interest (financial or non-financial).

**Ethical Approval** For this type of study, formal consent is not required. Present paper does not violate any ethical standards.

**Informed Consent** Data was collected by survey methods from the participants, and informed consent was obtained from all individual participants included in the study/survey.

## References

1. Kumar N (2018) E-commerce in India: an analysis of present status, challenges and opportunities. *Int J Manage Stud* V 2(3)
2. Mahipal D, Shankaraiah K (2018) E-commerce growth in India: a study of segments contribution. *Acad Mark Stud J* 22(2)
3. Rajasekar S, Agarwal S (2016) A study on impact of e-commerce on India's commerce. *Int J Dev Res* 6(3):7253–7256
4. Mitra A (2013) E-commerce in India—a review. *Int J Mark Fin Serv Manage Res* 2(2). ISSN: 2277-3622
5. Saaty TL (1980) *The analytic hierarchy process*. McGraw-Hill, New York
6. Taherdoost H (2017) Decision making using the analytic hierarchy process (AHP); a step by step approach. *Int J Econ Manage Syst* 2. ISSN: 2367-8925
7. <https://www.thewebbureau.com/blog/2016/april/top-8-ecommerce-website-success-factors>
8. Eid R, Sharief RYA, Hussein L (2011) Factors affecting the success of online branding: an empirical study. *Int J Online Mark* 1(4):20–32
9. Marfuah WS (2017) The implementation of analytical hierarchy process method for outstanding achievement scholarship reception selection at Universal University of Batam. In: *International conference on environment and technology (IC-Tech) 2017*. IOP conference series: earth and environmental science, vol 97. IOP Publishing, 012003
10. Al-Aomar R (2010) A combined AHP-entropy method for deriving subjective and objective criteria weights. *Int J Ind Eng* 17(1):12–24. ISSN: 1943-670X
11. Hung CC, Chen LH (2009) A fuzzy TOPSIS decision making model with entropy weight under intuitionistic fuzzy environment. In: *International multi conference of engineers and computer scientists 2009*, vol I, IMECS 2009, 18–20 Mar 2009. ISBN: 978-988-17012-2-0
12. Divyaa SK, Yadav N, Dubey SK (2016) Usability evaluation of mobile phones by using AHP-entropy approach. In: *3rd International conference on computing for sustainable global development*. IEEE, New Delhi, pp 4063–4067



# Computing Diet Composition for Patients with High Cholesterol Using Decision-Making Approach



Garima Rai and Sanjay Kumar Dubey

**Abstract** Cholesterol is a blended mixture that presents in tissues of the human body. Cholesterol and by-products of cholesterol are essential constituents of cell mucosa. A sudden increase in the level of cholesterol leads to increased risk of coronary heart diseases. Cholesterol is not required to be consumed as a part of the daily diet; human body can exaggerate cholesterol on its own. Cholesterol supports the proper functioning of human metabolism. Cholesterol is essential to build healthy body cells. Sudden increase in the level of cholesterol does not have any notable indication. Increased cholesterol levels in blood can only be suspected through a blood test. Factors such as poor diet, obesity, lack of exercise, smoking, age, and diabetes cause poor cholesterol level in the human body. Poor cholesterol level can cause complications like chest pain, heart attack, and stroke. Human liver secretes about 80% of the human body's cholesterol and the other comes from the dietary constituents such as fish, meat, eggs, and dairy products. The proposed work suggests the dietary patterns to recommend a diet for patients with poor cholesterol level. This diet will help in maintaining a good cholesterol level. Analytic hierarchy process is used as a technique to implement this proposed work. Estimation of ideal diet has been recommended to patients with poor cholesterol level. The final outcomes reveal the ideal diet plan for the patients with a bad cholesterol level. This recommends the diets to be consumed by the patients at breakfast, lunch, and dinner. Fuzzy methods are used to validate our work. The final results are in match with the results obtained by analytic hierarchy process.

**Keywords** Recommendation system · Cholesterol · AHP · Entropy · Model

---

G. Rai (✉) · S. K. Dubey  
Department of Computer Science Engineering, Amity University Uttar Pradesh, Sec-125,  
Noida, India  
e-mail: [garima.raii003@gmail.com](mailto:garima.raii003@gmail.com)

S. K. Dubey  
e-mail: [sanjukundan@gmail.com](mailto:sanjukundan@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_40](https://doi.org/10.1007/978-981-15-2329-8_40)

## 1 Introduction

This instruction file cholesterol is present in all the cells of the human body. Cholesterol is essential for brain, skin, and other human body organs to perform their functions efficiently, but poor cholesterol levels due to too much consumption of fatty dietary products can cause serious health issues. Some dietary products such as fruits, vegetables, and grains are not compromised of any cholesterol. An elevated cholesterol level is a notable mark for the risk of various heart diseases. As per recent statistical data, 71 million American adults have elevated cholesterol levels. Every one out of three adults has cholesterol conditions under control, and half the number of poor cholesterol patients gets the treatments. Maintaining a good cholesterol level helps in preventing the human body from numerous heart diseases like coronary diseases and bypass surgery.

We can maintain a healthy lifestyle by doing exercise on a daily basis or by including aerobics, jogging, and skipping habits such as smoking help in regulating the good cholesterol level in the human body. Elevated cholesterol levels are one of the major supporters in the burden of deaths in developing countries. In 2017, 16.9 million deaths recorded from coronary diseases due to poor cholesterol levels. Year 2018 estimates 17 million as of early death rates from coronary diseases. As per the latest statistical estimates, urban population is mostly to suffer from a high cholesterol level because of their poor lifestyle and eating habits. Annually, 10 million deaths are recorded due to coronary diseases caused due to poor cholesterol levels. In urban population, it is mostly the people above the age 20 mostly having elevated cholesterol levels. 2015–2016 reported more than 12% of adult age 20 and above that had total cholesterol above 200 mg/dl. In the USA, a little more than a half takes the benefits of medicine for cholesterol. As increased in the level of cholesterol level does not have any symptoms, most of the people do not know that the level of cholesterol has increased in their blood until an alarming stage is reached. One of the best ways to monitor your cholesterol level is to maintain a habit of regular checkup. A blood test named “lipid profile” measures the level of cholesterol in the human body. The blood test requires to fast means not to eat or drink for 8–12 h before the test. The test evaluates low-density lipoprotein (LDL), high-density lipoprotein (HDL), triglycerides, and total cholesterol. Low-density lipoprotein is also called bad cholesterol; high-density lipoprotein is known as good cholesterol; and high-density lipoprotein takes it back to the liver from the bloodstream. Low-density lipoprotein takes cholesterol from liver to the bloodstreams. In order to measure the cholesterol levels in the milligrams, the unit is deciliter. Adults are suggested to get their cholesterol levels examined in every 3–5 years of duration. People having records of elevated cholesterol level in history are suggested to maintain a habit to regularly examine their cholesterol level. Other components such as poor family background cannot be controlled but preventive measures are suggested to lower its impact so as to maintain a good cholesterol level. In order to suggest a proper diet plan for patients with elevated cholesterol level, a diet recommendation system is required. In the 1970s, Thomas L. Saaty developed a structural technique for categorizing and

examining complicated outcomes, based on mathematics and psychology termed as analytic hierarchy process [1]. A diet recommendation was proposed for diabetic patients using MCDM approach [2]. In earlier works, diet recommendation emphasizes more on exterior components compromising the need, methodology, latest trend, opportunities, etc., instead of internal components such as nutrients which should be present in the diet. In our work, nutrients play a vital role in determining the meals for patients with elevated cholesterol levels instead of the other exterior aspects. In the proposed work, we have implemented analytic hierarchy process for recommending the diet-containing nutrients with due importance. This technique has diverse improvement over previously proposed work as it focuses more on the vital nutrients essential for ideal meal. It involves AHP for determining an ideal diet plan comprising of three meals intake for the high cholesterol patients in order to preserve nutrients in body all through the day. As AHP formally implements the best substitute comprising of all essential elements, our paper aims to use this method in order to determine the ideal diet plan for high cholesterol patients [3].

## 2 Methodology

Multi-criteria decision-making (MCDM) strategy is implemented when the outcomes depend on certain criterion as a piece of circumstances and their influence of those criterions and circumstances on the final outcomes. Thomas L. Saaty arranged and proposed a MCDM approach called AHP. In this approach, both the aspects, numerical instruments and mental methodology, are incorporated. This approach requires an endeavor of making pair-wise connections of all the essential components as per requirement aspects. In relation to a particular factor, pair-wise relationships are formed. Components such as consistency rundown, and consistency extent help in examining the desired outcomes. Consistency extent in association under 10 percent presents an expected outcome. AHP makes the utilization of analyzing the tables (Table 1). The process of analysis was suggested by Saaty. Saaty introduced an initial size of internal and external numerals which are displayed in the grid (Table 2).

**Table 1** Factors and their importance in the diet

Components	Importance of the components in the diet [11]
Vitamin B5	It effectively reduces the LDL sustaining the CoQ10 levels. It is the derivative of pantethine. It lowers the LDL level in the human body safely
Omega-3 fatty acid	It may not support in lowering LDL level, but it assists in increasing HDL level
Calcium	It helps in lowering the cholesterol. Dairy products such as milk contain calcium
Vitamin D	It is vital for keeping human healthy and impacts blood sugar levels in the human body. Vitamin D influences the health of human bones

**Table 2** Alternatives of diet and their components

Dietary	Components of the diet [4, 12, 13]
Alternatives	
Diet 1	<b>Breakfast:</b> One-fourth tablespoon cinnamon, half tablespoon methi, three–four ounces of mint leaves, one cup fat-free milk, one banana
	<b>Lunch:</b> 1 seasonal fruit such as apple and mango. Oven-baked chicken with multi-grain bread, sliced avocado and tomato, black beans combined with vinegar and olive oil
	<b>Dinner:</b> One bowl soup of clear vegetables, leafy vegetables, half-bowl brown rice, few blackberries
Diet 2	<b>Breakfast:</b> Few slices of wheat buns, 6 or 8 oz of nuts, yogurt with berries and seeds
	<b>Lunch:</b> 2 chappattis, half-bowl rice and beans, half-bowl spinach or cauliflower with roasted fish or roasted chicken topping
	<b>Dinner:</b> Three ounces stir-fry broccoli with basil, one bowl rice with one tablespoon diced nuts, one bowl steamed mushroom, one bowl curd
Diet 3	<b>Breakfast:</b> Few ounces of nuts and one egg white with one glass skimmed milk
	<b>Lunch:</b> Beans or soybeans soup, tofu with toppings of wheat germ
	<b>Dinner:</b> Cucumber raita, spiced carrots, one bowl of oats meal along with barley
Diet 4	<b>Breakfast:</b> 1 bowl wheat flake, half-cup fat-free milk, half cup nuts
	<b>Lunch:</b> cucumber salad of half smashed tofu with diced nuts and apples
	<b>Dinner:</b> Two bowls of capsicum, one sautéed veg. broccoli, 1-cup fat-skimmed milk

This measure complements the numerals which are utilized to display an interconnection made and the options opted in respect of every scale units [4]. Utilization of AHP is effective as the regimen of eating eventually terminates up being very elementary to perform as this merges all the basic components in the various relatable manners. One other aspect of AHP is to examine the efficiency of a decision-maker’s conclusions in a similar manner carrying out some differences in the important primary procedures. In order to figure out the right outcomes and the limitations of AHP, we are utilizing this system in our work. Entropy technique is one of the assorted systems for discovering weights. This strategy is used for assurance of weight of surveying file in fuzzy mathematics, which joins the opinions of the expert with fuzzy analysis survey, realizing a specific level to sort. It is a blended strategy of subjective examination and quantitative examination [5–10]. The entropy weight is dictated by the framework developed, and its structure of the choice execution cross section is demonstrated in lattice beneath.

### 3 Experimental Work

The dietary summarization of high cholesterol patients depends on some important components. These components were examined with the assistance of an analysis carried out by the creators. Some popular dieticians and physician specialists marked their analysis. Important four components were concluded and are characterized in the grid (Table 1).

In accordance with the above dietary components, we are featuring the selected consuming summarization which was suggested to us by the reputed and admirable medicines and dieticians who have esteemed experience in the medication of highly elevated cholesterol patients and have tremendous knowledge in this area. These consumption system decisions are relatable to the above components. The nourishing analysis of the factors is utilized to estimate the regiment measures of components needed by adamant person suffering from the poor cholesterol levels in accordance with primarily evaluated supplements. Every consuming requirement includes evaluated three primary dinning of the daily human diets particularly breakfast, lunch, and dinner. The diets are shown in Tables 2, 3, 4, 5, 6, 7, 8, and Fig. 1.

**Table 3** Pairwise comparison matrix of criterions

	Vitamin B	Niacin	Phytosterols	Soluble fibers	Eigenvector
Vitamin B5	1.0000	3.0000	7.0000	7.0000	0.5654
Omega-3 fatty acids	0.3333	1.0000	5.0000	7.0000	0.3001
Calcium	0.1429	0.2000	1.0000	3.0000	0.0879
Vitamin D	0.1429	0.1429	0.3333	1.0000	0.0466

$\lambda_{avg(max)} = 4.2278$ , C. I = 0.0759, C. R = 0.0844

**Table 4** D1, D2, and D3 w.r.t vitamin B5

	DIET 1	DIET 2	DIET 3	Eigenvector
DIET 1	1.0000	0.3333	5.0000	0.2790
DIET 2	3.0000	1.0000	7.0000	0.6491
DIET 3	0.2000	0.1429	1.0000	0.0719

$\lambda_{avg(max)} = 3.0649$ , C. I = 0.0324, C. R = 0.0559

**Table 5** D1, D2, and D3 w.r.t omega-3 fatty acid

	DIET 1	DIET 2	DIET 3	Eigenvector
DIET 1	1.0000	3.0000	7.0000	0.6694
DIET 2	0.3333	1.0000	3.0000	0.2426
DIET 3	0.1429	0.3333	1.0000	0.0879

$\lambda_{avg(max)} = 3.0070$ , C. I = 0.0035, C. R = 0.0061

**Table 6** D1, D2, and D3 w.r.t calcium

	DIET 1	DIET 2	DIET 3	Eigenvector
DIET 1	1.0000	7.0000	5.0000	0.7306
DIET 2	0.1429	1.0000	0.3333	0.0810
DIET 3	0.2000	3.0000	1.0000	0.1884

$\lambda_{avg(max)} = 3.0649$ , C. I = 0.0324, C. R = 0.0559

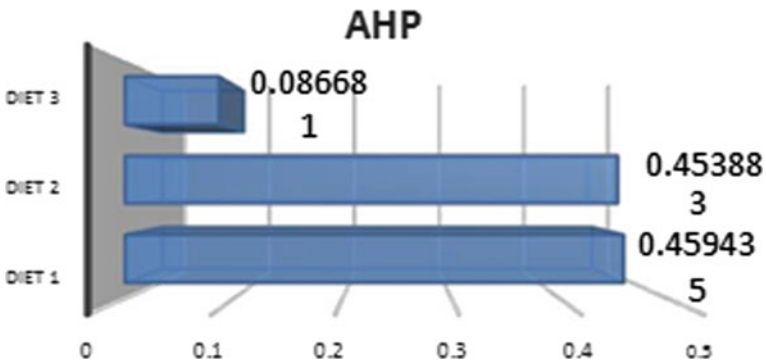
**Table 7** D1, D2, and D3 w.r.t vitamin D

	DIET 1	DIET 2	DIET 3	Eigenvector
DIET 1	1.0000	7.0000	9.0000	0.7854
DIET 2	0.1429	1.0000	3.0000	0.1488
DIET 3	0.1111	0.3333	1.0000	0.0658

$\lambda_{avg(max)} = 3.0803$ , C. I = 0.0401, C. R = 0.0692

**Table 8** Reliability index for diet alternatives

	Results	Rank
DIET 1	0.459435	1
DIET 2	0.453883	2
DIET 3	0.086681	3



**Fig. 1** Results through AHP

### 4 Validation of the Experimental Work

The methodology of entropy is among the variant procedures to evaluate the weights. The present regime (entropy) is used for validations of weight of reading data in fuzzy mathematics, which includes the evaluation of the studying of final fuzzy leading

**Table 9** Entropy calculations

Vitamin B5	Omega-3 fatty acids	Calcium	Vitamin D
0.7518	0.7519	0.6802	0.5937
0.2482	0.2481	0.3198	0.4063
0.203	0.2029	0.2616	0.3324

**Table 10** Ranking

	Diet 1	Diet 2	Diet 3
R3I	0.6447	0.2517	0.2517
Rank	1	2	3

to the sorting at some point. It is a combinational procedure of a particular analysis [5]. The entropy weight is managed (Tables 9, 10).

## 5 Conclusion and Future Aspect

Evaluating the hierarchy process is utilized to select the most suitable and effective eating diets among the three choices selected or chosen by esteemed dieticians and exceptionally trained specialists. The experimental study was done using AHP, and then the result was validated by using the entropy method, which provided the similar ranking for the diet recommendation. In the future, the study will be provided on the basis of more inputs and different parameters by using more different techniques and by using some real-life data.

**Acknowledgements** Authors acknowledged the person who directly/indirectly helped to complete the research paper.

**Conflict of interest** The authors declare that they have no conflict of interest (financial or non-financial).

**Ethical approval** For this type of study, formal consent is not required. The present paper does not violate any ethical standards.

**Informed consent** Data was collected by survey methods from the participant, and informed consent was obtained from all individual participants included in the study/survey.

## References

1. [https://en.wikipedia.org/wiki/Analytic\\_hierarchy\\_process](https://en.wikipedia.org/wiki/Analytic_hierarchy_process)
2. Nädäban S, Dzitac S, Dzitac I (2016) Fuzzy TOPSIS: a general view. *Proc Comput Sci* 91:823–831
3. Saaty T (1980) *The analytic hierarchy process: planning, priority setting, resource allocation*. McGraw-Hill, Texas
4. <https://www.bhf.org.uk/informationsupport/heart-mattersmagazine/nutrition/cooking-skills/10-heart-healthy-meals-in-less-than-10-minutes>
5. Singh MP, Dubey SK (2017) Recommendation of diet to anaemia patient on the basis of nutrients using ahp and fuzzy TOPSIS approach. *Int J Intell Eng Syst* 10(4)
6. Cheng, T, Zhang CX (2003) Application of fuzzy AHP based on entropy weight to site selection of solid sanitary landfill. *Environ Sanit Eng* 64–67
7. <https://www.lifeextension.com/Magazine/2015/5/Natural-Methods-To-Control-Cholesterol/Page-01>
8. <https://food.ndtv.com/food-drinks/cholesterol-diet-what-to-eat-and-avoid-to-lower-bad-cholesterol-1746881>
9. <https://www.pritikin.com/your-health/health-benefits/lower-cholesterol/1468-7-tips-for-improving-your-ldl-cholesterol.html>
10. <https://www.healthline.com/health/high-cholesterol/natural-cholesterol-reducers>
11. <https://www.betternutrition.com/features-dept/supplements-for-heart-health>
12. <https://www.mayoclinic.org/diseases-conditions/heart-disease/in-depth/hearthealthy-diet/art-20046702>
13. Bhatt A, Dubey SK, Bhatt AK (2018) Analytical study on cardiovascular health issues prediction using decision model-based predictive analytic techniques. In: Pant M, Ray K, Sharma T, Rawat S, Bandyopadhyay A (eds) *Soft computing: theories and applications*, vol 584. Springer, Singapore, pp 289–299 (*Adv Intell Syst Comput*)



# Automatic Vehicular Number Plate Recognition (VNPR) for Identification of Vehicle Using OCR and Tesseract



J. S. Nirmala, Rahul Banerjee and Rajath S. Bharadwaj

**Abstract** Vehicular number plate recognition (VNPR) is an important issue which can be solved using image processing and computer vision. VNPR process, when taken into consideration, can be used to solve a variety of issues in the fields of road safety and security. The problems can range from parking concerns to traffic control and might also include situations related to tollbooth or speed limit issues. Vehicle number plate recognition has a lot of side complexities arising due to a lot of factors including variable light dissipation, camera quality, and also speed at which the vehicle is moving. The proposed system will be able to identify number plates of cars and bikes and help trace the owner of that vehicle. We use various machine learning techniques and image processing utilities along with computer vision to address few of the issues which can be solved using the same.

**Keywords** Vehicular number plate recognition (VNPR) · Optical character recognition (OCR) · Artificial neural network (ANN) · Machine learning · Computer vision · Image processing · K-nearest neighbor

## 1 Introduction

Every vehicle is registered with a unique vehicle number containing a combination of letters and digits. These number plates distinguish vehicles from each other, where the number plate data is extracted from vehicle's picture or from grouping of pictures without direct human mediation [1]. Traffic activity control and vehicle proprietor distinguishing proof have progressed toward becoming a significant issue for each nation. Occasionally, it progresses in such a way that the entire process of VNPR becomes complicated and the probable reasons are if the vehicle proprietor abuses road safety rules and also if the driver drives too fast [2]. As advancements in the field of technology are continuously made, this process can also be migrated to the cloud, which can handle the computations and act like a database housing all the number plates being stored for security concerns or for future use [3, 4].

---

J. S. Nirmala · R. Banerjee (✉) · R. S. Bharadwaj  
Department of CSE, Nitte Meenakshi Institute of Technology, Bengaluru, India

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_41](https://doi.org/10.1007/978-981-15-2329-8_41)

VNPR is a very useful system which can not only help traffic police to deal with traffic violators but also can help to catch regular offenders. Countries like India in particular needs an effective system such as this to take care of the situations mentioned above. Therefore, it is important that such a system is implemented and deployed with high accuracy and efficiency [2, 4, 5].

The proposed system can be helpful for authorities looking for solutions to overcome the manual labor work of noting down vehicle number before they speed away.

## 2 Literature Review

The most important roles played in any video analysis of a number plate image are computer vision and character recognition and the algorithms that are used to recognize any number plate. These form the fundamental core modules of a VNPR system.

Optical character recognition (OCR) is one of the most widely used technologies which helps in obtaining machine-encoded text by translating the text present in photographs. The text present in an image can be either printed text or handwritten text. OCR works on a feed-forward neural network to detect non-overlapping sets of images to detect the characters [6, 7]. Artificial neural network (ANN) is also one of the most popularly used intelligent computing methods to recognize patterns. A simple internal architecture called multilayered feed-forward neural network is more commonly used, while making use of ANN classifies the available inputs into a set of target categories. In this technique, we use feature extraction and binary pixel values to get an organized input which is obtained from neural networks. Feature extraction helps to achieve good performance even under non-favorable conditions like the image being noisy or even in situations where the lighting conditions are poor [8, 9]. Common pattern matching technique is one of the rudimentary methods to recognize the text of a single font and of non-varying size. It is a well-suited approach for the implementation of VNPR systems. They require less computation power and have less execution time. The accuracy is less compared to ANN as incorrectly segmented characters or misplaced characters in the text affect the ability of the OCR [1, 7, 10]. Neural networks, on the other hand, are, however, more precise in recognizing the characters accurately as they have a strong memory and self-adaptation to assist. ANN requires more neurons to work efficiently which increases the computing power and execution time significantly [11].

Every system discussed has its own advantages and disadvantages. Based on user requirements or available resources, a suitable technique can be used to implement VNPR.

### 3 Methodology

The main objectives of the system are recognizing the number plates and extracting the data from the images taken. The extracted data can then be used to track down the owners of the vehicles.

This process includes the below-mentioned three major stages:

1. Number plate detection
2. Image extraction and pre-processing
3. Character segmentation and recognition.

#### 1. Number Plate Detection:

A number plate can be extracted by using image segmentation method. Images are taken in different illumination conditions and at various distances from the camera [12]. It is difficult to click pictures of moving vehicles; therefore, the process comes along with lower accuracy. Number plates can be of different shapes and different sizes and can contain a lot of font styles; font size might also vary, and hence, all these must be kept in mind while computing everything in this field [8].

#### 2. Image Extraction and Pre-processing:

When a vehicle comes into the frame, the algorithm is designed in such a way that the number plate is located, detected, and extracted from the image/video feed. This image, in turn, is sent for processing for the extraction of the features. Image processing includes few of the following steps [1, 3, 5, 7, 8]:

- (i) The image getting converted from RGB format to gray scale (infrared)
- (ii) Smoothing of the image by noise reduction
- (iii) Contrast enhancement by changing threshold values of the images
- (iv) Canny edge detection and feature extraction
- (v) Actual number plate extraction.

The features in turn are used to extract the number plate, read the characters, and make the process ready for OCR.

#### 3. Character Segmentation and Recognition:

Character segmentation is all about understanding and recognizing the characters present in the image; it can also be a series of characters.

Segmentation is also about extracting the attributes of the characters. It tells us about the type of the character, i.e., whether the character is a number or an alphabet. It will also help us classify whether the alphabet is in upper case or lower case. The number plate can be of different size which can also result in the font of the number plate to change as well [10, 11].

OCR makes use of pattern matching and recognition. It is an offline process where it targets a group of characters or sometimes even individual characters. It can be used to extract particular and specific characters from the images and can also be used for

extracting characters out of a handwritten document. OCR makes use of K-NN to locate the nearest possible character based on the process of matching. When there is confusion between two similar alphabets, the K-NN score decides which alphabet it should be classified as such [4, 6, 8, 11].

## 4 Computer Vision Process

### Associated Algorithm

Step 1 Import all the dependent libraries: cv2, Tesseract, NumPy, PIL, copy.

Step 2 **Pre-processing Procedure**

- Take 'img' as input
- Using 'cv2.GaussianBlur', change the blur factor of the 'img'; 'img' now becomes 'imgBlur'
- Change the 'img' to gray scale using 'cv2.cvtColor(imgBlur, cv2.COLOR\_BGR2GRAY)'
- Change the threshold of 'imgBlur' using any suitable threshold template

Step 3 **Noise Reduction Procedure**

- Detect the 'contours' in the plate 'img'
- If 'contours' are detected
  - Mark the area coordinates
  - Bind the area using cv2.boundingRect()
  - Apply any suitable threshold template to the bounded area

Step 4 **Extraction Process**

- The bounded area is cropped and is saved as a separate 'result\_img.png'

Step 5 Using the 'Tesseract' library; extract the characters with aid of the OCR process

Step 6 Store the time stamp and the vehicle number in the designated 'plate.csv' file.

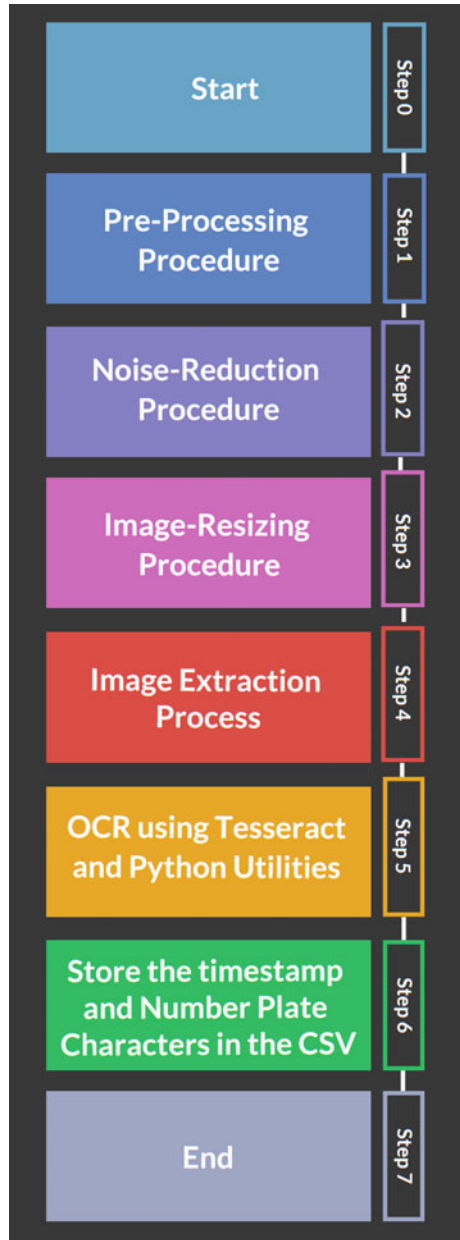
### Flowchart Depicting the Process

See Fig. 1.

## 5 Results

See Figs. 2, 3, 4, and 5.

**Fig. 1** Flowchart for the entire process of VNPR





**Fig. 2** Original number plate image\* (reference image has been taken from the Internet; hence, there is no chain of ownership in knowledge) <https://www.quora.com/How-are-number-plates-assigned-in-India>, <https://www.quora.com/How-are-number-plates-assigned-in-India>



**Fig. 3** Cropped out and enhanced number plate image\* (resultant of the reference image) <https://www.quora.com/How-are-number-plates-assigned-in-India>

## 6 Challenges Associated

There are numerous constraints that come in the way of high accuracy in detecting and reading the number plates. Below are a few of the limitations discussed [1, 2, 5, 7, 10, 11]:

- The first factor is the fact that every country has its own format for number plates, and there is no uniformity in terms of font or size of the text.
- The second factor is the distorted pictures with too much noise that hinders the ability of the OCR system to detect the number plate precisely. This can be due to a variety of reasons like low-resolution cameras, incomplete scope of vision, and high speed of cars.

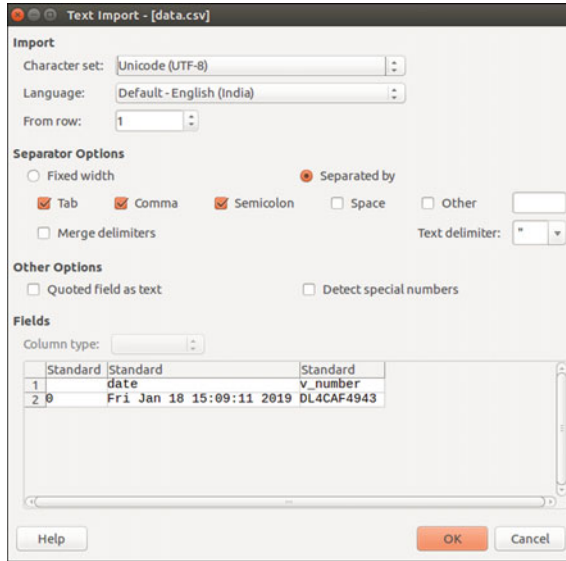


Fig. 4 CSV file handling the vehicle number plate and associated time stamp

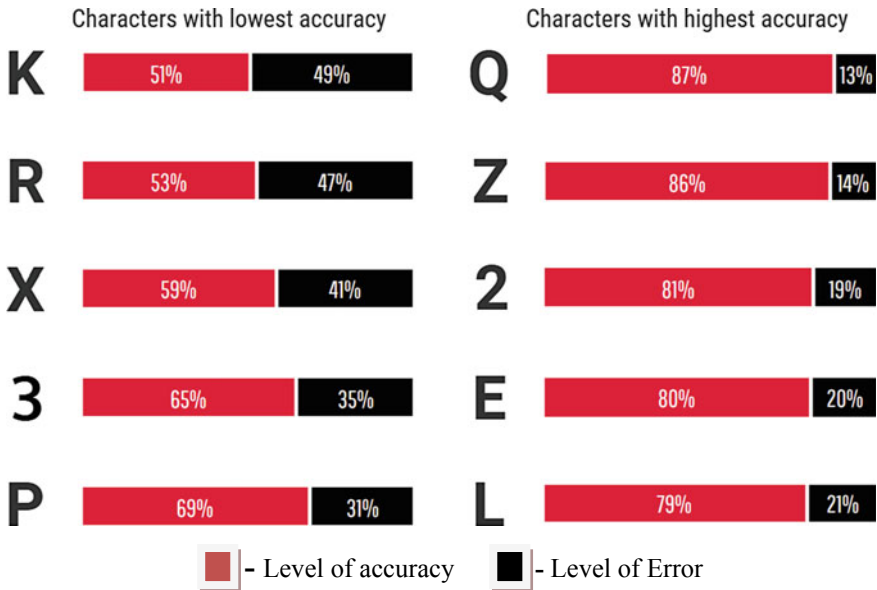


Fig. 5 Statistical representation of the data collected after conducting 100 tests

- The third factor is the computing power of the system that handles the OCR process. If the necessary software and hardware requirements are not met, the speed and correctness of the system to detect number plate reduce.
- The fourth factor is the visibility of the plate; if a car's number plate is covered with dust or mud, it is physically impossible to uncover the characters hidden below the dirt or mud.
- The fifth factor is the visibility condition. If the environment is too dark, the same process becomes complicated as the characters may not be properly visible.
- Other factors include the high cost associated with the acquisition of a high-quality camera for the process. This might be a one-time process but will still set the organization back a bit.

## 7 Conclusions and Future Work

This paper presents an efficient way of implementing an automatic number plate recognition System. The proposed system is mainly meant for vehicles registered in India, but it also works with number plates from selected countries. This system works well in reasonably low-lit environments and images with little noise in them. Rigorous tests have been done to check the accuracy of the system under different weather and lighting conditions. The limitations of this system are with respect to images with misplaced text and number plates of a few countries with no spacing between characters and numbers that look alike. The system has an accuracy of recognizing 81 number plates correctly in a sample set of 100. Future works include improving the efficiency by distinguishing similar looking numbers and characters and predicting partially visible text in a number plate.

## References

1. Malik SM, Hafiz R (2014) Automatic number plate recognition based on connected components analysis technique. In: IEEE 2nd international conference on emerging trends in engineering and technology (ICETET'2014), pp 30–31
2. Hongliang B, Changping L (2004) A hybrid license plate extraction method based on edge statistics and morphology. In: Proceedings of the 17th international conference on pattern recognition, vol 2 (ICPR'04), pp 831–834
3. Ghazal M, Hajjdiab H (2013) License plate automatic detection and recognition. In: IEEE international conference 2013
4. Daramola SA, Adetiba E, Adoghe AU, Badejo JA, Samuel IA, Fagorusi T (2011) Automatic vehicle identification system using license plate. *Int J Eng Sci Technol (IJEST)* 3(2)
5. Noor SSM, Tahir NM (2010) Car plate recognition based on UMACE filter. In: 2010 international conference on computer applications and industrial electronics (ICCAIE), IEEE
6. Jia Y, Gonnot T, Saniie J (2016) Design flow of vehicle license plate reader based on RGB color extractor. In: 2016 IEEE international conference on electro information technology (EIT), IEEE



7. Zhao Z, Ma Q (2012) A real-time processing framework for massive traffic sensor data. In: 2012 international conference on connected vehicles and expo (ICCVE), IEEE
8. Kumar A, Godara S (2008) A review: on number plate recognition, IEEE
9. Singh S, Kaur B (2016) Number plate recognition through image using morphological algorithm. In: 2016 3rd international conference on computing for sustainable global development (INDIACom), IEEE
10. Roth PM et al (2010) Automatic detection and reading of dangerous goods plates. In: 2010 seventh IEEE international conference on advanced video and signal based surveillance (AVSS), IEEE
11. Agarwal A, Goswami S (2016) An efficient algorithm for automatic car plate detection and recognition. In: 2016 second international conference on computational intelligence and communication technology (CICT), IEEE
12. Wang J, Gao G, Yang H (2009) The method research of vehicle image preprocessing and license plate location. In: 2009 international conference on computational intelligence and software engineering (CiSE), IEEE

# High-Frequency CNTFET-Based Voltage-Controlled Oscillator for PLL Application



Yogesh Kumar, Ashish Raman, Ravi Ranjan and R. K. Sarin

**Abstract** In this paper, a high-frequency differential ring dual-delay voltage-controlled oscillator (DR-VCO) is proposed. This 9 GHz DR-VCO is designed using 45 nm CNTFET technology. CNTFET is faster than MOSFET, and hence it provides a high tuning range. VCO with four-delay cells is designed in this work. This design consists of four-stage delay cells, and every delay cell is designed on dual-delay path topology. This topology results in high-output oscillation frequency which includes the range from 3 to 9 GHz. The observed phase noise at offset of 1 MHz frequency is  $-66$  dBc/Hz.

**Keywords** Differential delay cell · CNTFET (Carbon nanotube field-effect transistor) · Differential ring VCO (DR-VCO) · Phase-locked loop · Dual-delay path topology

## 1 Introduction

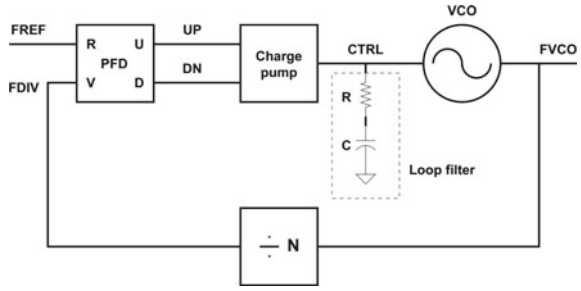
Almost in every electronic circuit, phase-locked loops are the essential block for frequency synthesizer and data recovery circuits. Critical parameters of wireless communication systems are tuning range of frequency and power dissipation [1–4]. VCOs are the essential building block of PLL circuits. PLL circuits are broadly used in various wireless communication systems. A phase detector (PD), low pass filter (LPF), charge pump (CP) and voltage-controlled oscillator (VCO) are the components of PLL. PLL's stability proportionally depends on VCO's performance. To improve the overall performance of PLL, the VCO design can be improvised to attain better results. Generally, there are two types of VCOs can be used in circuits which are LC oscillators and ring oscillators. In LC oscillators, the inductor is designed using a spiral winding method, so it takes a broader area with high power dissipation which is not desired in modern large scale circuits. LC oscillators are used where the main focus is to achieve a superior phase noise performance. But in the other aspect,

---

Y. Kumar (✉) · A. Raman · R. Ranjan · R. K. Sarin  
Department of ECE, NIT, Jalandhar, India  
e-mail: [yogesh.ece123@gmail.com](mailto:yogesh.ece123@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_42](https://doi.org/10.1007/978-981-15-2329-8_42)

**Fig. 1** PLL block diagram



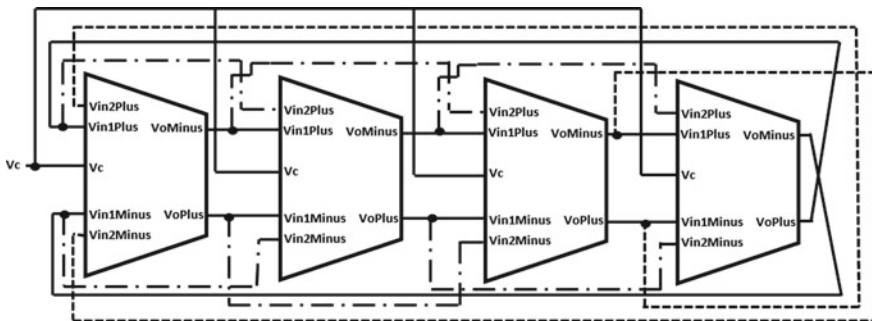
the tuning range of frequency for LC oscillator is very poor [5]. So, because of the above disadvantages, ring oscillators are used over LC oscillators Fig. 1.

The motivation of this work is to enhance the frequency range. In the proposed design, CNTFETs are used to enhance the operating frequency of VCO and to increase the frequency range. Due to dependency on chirality factor of CNTFET, its electronic structure can be controlled. To achieve high-performance characteristics, carbon nanotubes (CNTs) are best suited because the transport, as well as ballistic or near-ballistic transport, is determined with low-voltage bias. The proposed VCO which consists of four-stage delay cells is designed for application of high switching speed and low power.

## 2 Design Description of VCO

The VCO block diagram is shown in Fig. 2. In this DR-VCO with four stages, dual-delay cells are shown in the figure.

The oscillation frequency of output of the DR-VCO is obtained by the total count of dual-delay cells ( $N$ ) which are associated in a closed loop to create oscillations. In case, if the count of delay stages in the closed loop is fixed, then by controlling the



**Fig. 2** Block diagram of proposed DR-VCO [6]

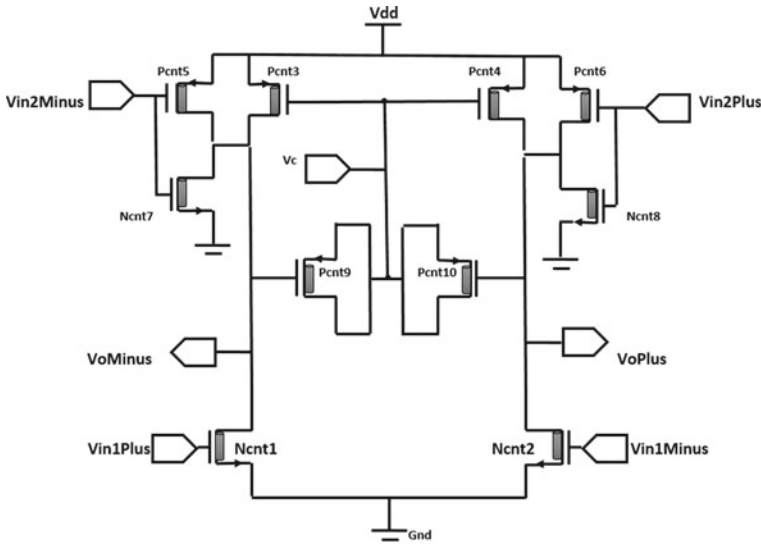


Fig. 3 Single delay cell of designed VCO [6]

control voltage, the output frequency can be controlled. The oscillation frequency of output of the DR-VCO can be evaluated as follow [7]:

$$f_o = \frac{1}{2t_d N} \tag{1}$$

where  $f_o$  is the output frequency and  $t_d$  is the delay time of a single cell and  $N$  is the count of the delay cell.

The single delay cell is shown in Fig. 3 [6].

The output frequency for conventional ring VCOs is limited due to the fact that the delay time of single inverter is always less than the overall circuit delay. The negative-skewed path design is used to remove that frequency limitation [8] by reducing overall delay time of the oscillator below the level of single inverter’s delay.

### 3 Delay Cell Working Methodology

In Fig. 3, the primary inputs are Vin1Plus and Vin1Minus. Two transistors PCNT3 and PCNT4 are used as a load. Transistors PCNT9 and PCNT10 are used as varactor by shorting their drain and source terminals. The output frequency can be controlled by changing the control voltage ( $V_c$ ), which is input to the third terminal of load transistor, i.e., gate terminal, PCNT3 and PCNT4 and also to the drain-source terminal of PCNT9 and PCNT10. When the  $V_c$  is decreased, the resistance( $R$ ) of load transistors and capacitance( $C$ ) of PCNT9 and PCNT10 decreases [6]. So when the

time constants  $R$  and  $C$  of the dual-delay cell decrease, it results in increase in the output frequency. When  $V_{in1Plus} < V_{th}$ , NCNT1 turns off and since the PCNT5 is connected to the feedback signal, it receives input signal earlier than CNT1, it results in conducting CNT5 transistor. This pre-charges the output signal by pulling it up and hence it decreases rise time of the output signal. NCNT7 and NCNT8 are used in the same way to pull down the fall time by pre-discharging the output. Hence, this methodology increases the oscillation frequency of output.

The power dissipation of a DR-VCO can be evaluated as follows [9]:

$$P = NI_dV_{dd} \tag{2}$$

To optimize the power dissipation of the DR-VCO, the delay cells ( $N$ ) can be chosen.

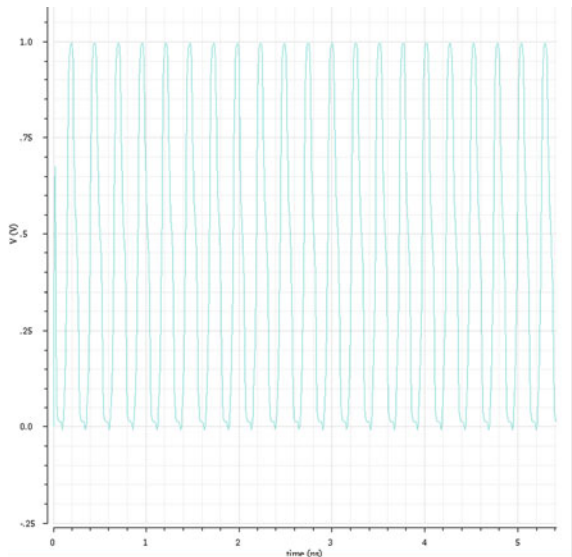
Phase noise is very crucial specification of VCO. The noise performance of a device can be evaluated by its phase noise. For this design, the obtained phase noise at offset frequency of 1 MHz is  $-66$  dBc/Hz.

### 4 Simulations

The output oscillation frequency ranges from 3.84 to 9.27 GHz. The following output is shown in Fig. 4 at  $V_c = 0.7$  V.

The output frequency ( $V_c = 0.7$  V) in Fig. 5 and power dissipation graph in Fig. 6 of designed VCO are shown.

**Fig. 4** Output waveform of designed VCO



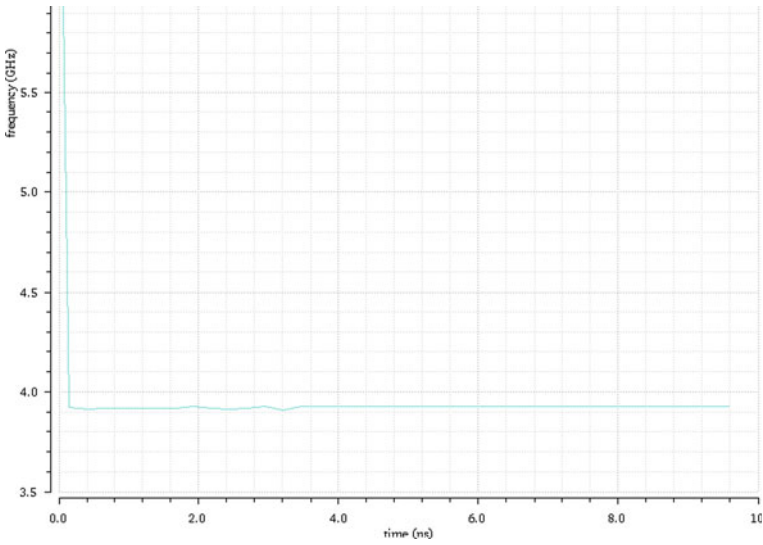
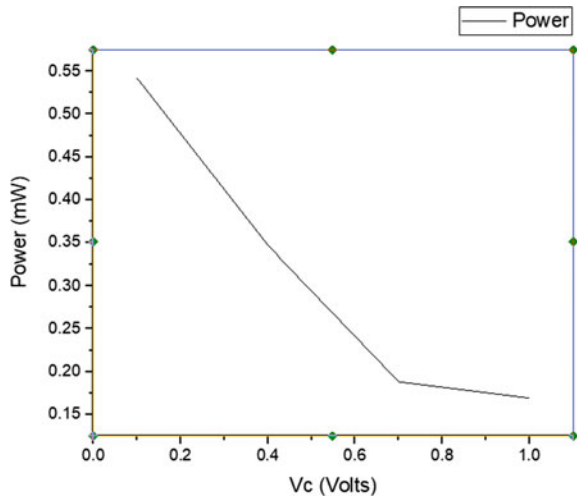


Fig. 5 Frequency output of designed VCO

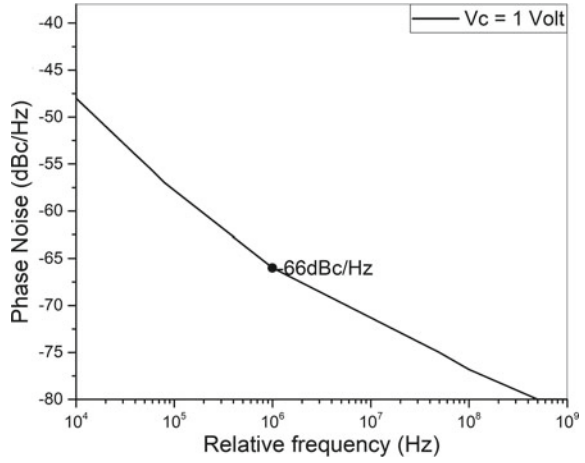
Fig. 6 Power dissipation of designed VCO



In Fig. 7, phase noise at 1 MHz offset is shown, and in Table 1, frequency, power dissipation and phase noise variation for the range of control voltage are shown.

In Table 2, the comparative analysis is shown with conventional CMOS VCOs and CNTFET-CMOS (hybrid) VCOs and the proposed work.

**Fig. 7** Phase noise performance of designed VCO



**Table 1** Frequency and power dissipation with  $V_c$  variation

Control voltage( $V_c$ )	Frequency (GHz)	Power dissipation (mW)	Phase noise (-dBc/Hz)
0.1	9.27	0.542	-61.5
0.4	5.39	0.347	-63
0.7	3.92	0.188	-64.8
1.0	3.84	0.169	-66

**Table 2** Comparative analysis of different designs

References	Tech ( $\mu\text{m}$ )	Tuning range (GHz)	Supply voltage (V)	Power (mW)	Phase noise (dBc/Hz)
[1]	180 nm CMOS	0.479–4.09	1	13	-93.3 @ 1 MHz
[10]	180 nm CMOS	Max5.5	-	17.5	-104 @ 4 MHz
[4]	180 nm CMOS	0.12–1.3	0.5	0.085	-72 @ 1 MHz
[11]	180 nm CMOS-CNTFET	3.12–5.26	1.8	0.625	-
This work	45 nm CNTFET	3–9	1	0.542	-66 @ 1 MHz

## 5 Conclusion

This design can be used for high-frequency application with low power consumption. The output frequency is high which lies between 3.84 and 9.27 GHz. The maximum power dissipation is 0.542 mW. The phase noise value at offset of 1 MHz frequency is  $-66$  dBc/Hz at a control voltage of 1 V.

## References

1. Sheu M-L, Tiao Y-S, Taso L-J (2011) A 1 V 4 GHz wide tuning range voltage-controlled ring oscillator in 0.18  $\mu\text{m}$  CMOS. *Microelectron J* 42:897–902
2. Jovanovic G, Stoj M, Stamenkovic Z (2010) A CMOS voltage-controlled ring oscillator with improved frequency stability. In: Scientific publications of the state university of Novi Pazar, vol 2, pp 1–9
3. Appenzeller J, Knoch J, Martel R (2002) Carbon nanotube electronics. *IEEE Trans Nanotechnol* 1(4):184–189
4. Li T, Ye B, Jiang J (2009) 0.5 V 1.3 GHz voltage controlled ring oscillator. In: ASIC, 2009. ASICON'09. IEEE 8th international conference. IEEE, Hunan, pp 1181–1184
5. Razavi B (1997) A 2 GHz 1.6 mW phase-locked loop. *IEEE J Solid-State Circuits* 32:730–735
6. Kumar N, Kumar M (2018) Design of CMOS-based low-power high-frequency differential ring VCO. *Int J Electron Lett*
7. Ashish R, Vashishtha JN, Sarin RK (2012) A RF low power 0.18  $\mu\text{m}$  based CMOS differential ring oscillator. In: Proceedings of the world congress on engineering, vol 2
8. Park CH, Kim B (1999) A low-noise, 900 MHz VCO in 0.6  $\mu\text{m}$  CMOS. In: *IEEE J Solid-State Circuits* 34:586–591
9. Parvizi M, Khodabakhsh A, Nabavi A (2008) Low-power high-tuning range CMOS ring oscillator VCOs. In: IEEE international conference on semiconductor electronics. Johor Bahru, pp 40–44
10. Cho TS, Lee KJ, Pan T, Kong J, Chandrakasan AP (2007) cDesign and characterization of CNT-CMOS hybrid systems. MTL annual research report
11. Rahane SB, Kureshi AK (2017) A low power and linear voltage controlled oscillator using hybrid CMOS-CNFET technology. *Int J Appl Eng Res* 12(9):1969–1973. ISSN 0973–4562



# Trajectory Tracking Control of Unmanned Aerial Vehicle for Autonomous Applications



Aditi Zear and Virender Ranga

**Abstract** Quadrotors are suitable aerial platform for carrying out agile flight maneuvers. Currently, quadrotors are handled by ignoring different aerodynamic effects such as rotor drag. Rotor drag is the main aerodynamic effect that causes trajectory tracking error in flight during high speed. Therefore, a control model considering the rotor drag effect is proposed in this research paper. Our proposed model exploits the differential flatness property of dynamic quadrotor model to reduce the trajectory tracking error during the flight of unmanned aerial vehicle (UAV) in the environment. Further, a geometric controller is used to stabilize the UAV in the midair. A trajectory publisher is also used to provide the stream of feed-forward control terms for the desired trajectory. The performance of proposed control method is checked with the benchmark controller by computing root-mean-square position error. The proposed solution is tested for predefined horizontal Geronno Lemniscate trajectory and circular trajectory. Further, the proposed model is also tested with change in angular velocity values to prove its robustness in the environment.

**Keywords** Quadrotor · Rotor drag · Trajectory publisher · Differential flatness

## 1 Introduction

Recently, quadrotors have been emerged as convenient aerial platform for carrying out agile flight maneuvers. These aerial rotorcrafts are being employed for robotics research such as for different control design analysis. They are well suited for robotics research work because of their hover capability and ability to carry considerable payload according to their size. They are easily repairable and economical with their increasing demand in different applications. Quadrotors have simple mechanical structure consisting of four rotors with elementary propellers and stiff frame, how-

---

A. Zear (✉) · V. Ranga

Department of Computer Engineering, NIT Kurukshetra, Kurukshetra, India

e-mail: [aditizear93@gmail.com](mailto:aditizear93@gmail.com)

V. Ranga

e-mail: [virender.ranga@nitkk.ac.in](mailto:virender.ranga@nitkk.ac.in)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication*

*Engineering*, Lecture Notes in Networks and Systems 106,

[https://doi.org/10.1007/978-981-15-2329-8\\_43](https://doi.org/10.1007/978-981-15-2329-8_43)

ever [1–4]. These are usually controlled by ignoring various aerodynamic effects like rotor drag which becomes important in non-hover conditions. The rotor drag effect basically originates from induced drag of rotors and blade flapping. The rotors induce aerodynamic drags during translational movement [4–7]. The aerodynamic effects, when considered as unrecognized disturbances, work well in controlling the quadrotor with low velocity near to hover conditions. However, with the increase in velocity, the quadrotor’s accuracy for trajectory tracking reduces. It is valuable to achieve accurate trajectory tracking for rapid obstacle avoidance [5, 8–12]. In order to achieve such accuracy, a solution is needed to be proposed for exact tracking of unknown trajectories before flying.

## 2 Related Work

Many solutions have been proposed in order to target the accuracy of trajectory tracking. Some solutions are also proposed while considering the obstacle avoidance [9–16]. However, less attentions have been paid in constructing nonlinear control systems to design trajectory tracking controllers [13–18]. Therefore, in this research paper, our work is focussed in developing a control system to improve trajectory tracking efficiency of quadrotors by taking into account rotor drag effects. We have cast off the property of differential flatness, i.e., the dynamic model of quadrotor likely to have linear rotor drag effects is differentially flat having flat outputs chosen to be its heading and position [5, 16, 18–20]. This property is further used in computing feed-forward control terms from the reference trajectory that needs to be tracked.

## 3 Control Law

The controller is used to track the reference trajectory. It uses feedback terms calculated from the tracking errors along with the feed-forward terms calculated from the reference trajectory by using the differential flatness property of the quadrotor [1–4]. The quadrotor’s control architecture does not permit to apply torque inputs exactly [5, 16]. Hence, a lower-level body rate controller is provided to accept the desired body rates. Hence, a control algorithm that consists of lower-level body rate controller along with higher-level position controller is proposed in this paper. A world frame and body frame fixed with quadrotor’s origin, coinciding with center of mass (COM) and represented in world coordinates, are used with their orthonormal basis  $x_W, y_W, z_W$  and  $x_B, y_B, z_B$ , respectively. The quadrotor subjects to acceleration  $g$  due to gravitation in  $-z_W$  direction. The quadrotor’s COM position is denoted as  $p$  and its velocity, orientation, and acceleration are denoted as  $v, R$  and  $a$ , respectively. The position controller calculates the required orientation  $R_d$ , the collective thrust  $C_{cm}$ , the required body rates  $\omega_d$ , and the angular acceleration  $\dot{\omega}_d$ . All these inputs are then feeded to low-level controller [5]. As initial step in high-level controller, the

acceleration of quadrotor's body can be calculated as

$$a_d = a_f + a_r - a_{rd} + gz_w \quad (1)$$

Here  $a_f$  is the proportional derivative feedback control terms which can be calculated from control errors in position and velocity.

$$a_f = -K_p(p - p_r) - K_v(v - v_r) \quad (2)$$

where constants  $K_p$  and  $K_v$  are constant diagonal matrices and  $a_r = -R_r D R_r^T v_r$  is acceleration of rotor drag. Here  $D$  is a constant diagonal matrix that consists of mass-normalized rotor drag coefficients, and  $a_r$ ,  $v_r$ ,  $p_r$ , and  $R_r$  are the quadrotor's acceleration, velocity, COM, and orientation calculated from reference trajectory, respectively. The desired orientation  $R_d$  with respect to reference heading  $\psi_r$  is calculated as

$$z_{B,d} = \frac{a_d}{\|a_d\|} \quad (3)$$

$$x_{B,d} = \frac{y_C \times z_{B,d}}{\|y_C \times z_{B,d}\|} \quad (4)$$

$$y_{B,d} = z_{B,d} \times x_{B,d} \quad (5)$$

The desired accelerations are projected into the actual body's  $z$ -axis, and the collective thrust is calculated as

$$c_c = a_d^T z_B - k_h (v^T (x_B + y_B))^2 \quad (6)$$

Similarly, the body rates are calculated as

$$\omega_d = \omega_f + \omega_r \quad (7)$$

where  $\omega_f$  and  $\omega_r$  are the feedback and feed-forward terms calculated from altitude controller and reference trajectory, respectively. The root-mean-square error over  $N$  control cycles of high-level controller for executing the given trajectory is used to analyze the trajectory tracking performance. The error can be calculated as

$$E_a = \sqrt{\frac{1}{N} \sum_{k=1}^N \|E_p^k\|^2} \quad \text{where } E_p^k = p^k - p_r^k \quad (8)$$

### 4 Working Model

The described control law is implemented using Robot Operating System (ROS) and PX4 flight stack. The Gazebo simulator is used to simulate the results. ROS nodes are created using C++. MAVROS is used to provide efficient communication between ROS and PX4 flight stack. The control law is implemented using two ROS nodes called geometric controller and trajectory publisher [19, 20]. The controller benchmark as ROS node is used to track the errors and RVIZ visualization tool is used to visualize this control model. Figure 1 gives the brief description about exchange of ROS messages between components, and Fig. 2 shows the control model components along with their published and subscribed topics. The major components of our proposed working model and their operations are discussed below.

#### 4.1 Geometric Controller

It is used to implement the differential flat system considering the rotor drag effect. The controller continuously publishes thrust and body rates at /mavros/setpoint\_raw/attitude (ROS topic) in order to satisfy the constraints described by PX4. The desired position and velocity are obtained from trajectory publisher at the intervals of 10 ms. The actual position and velocity are published

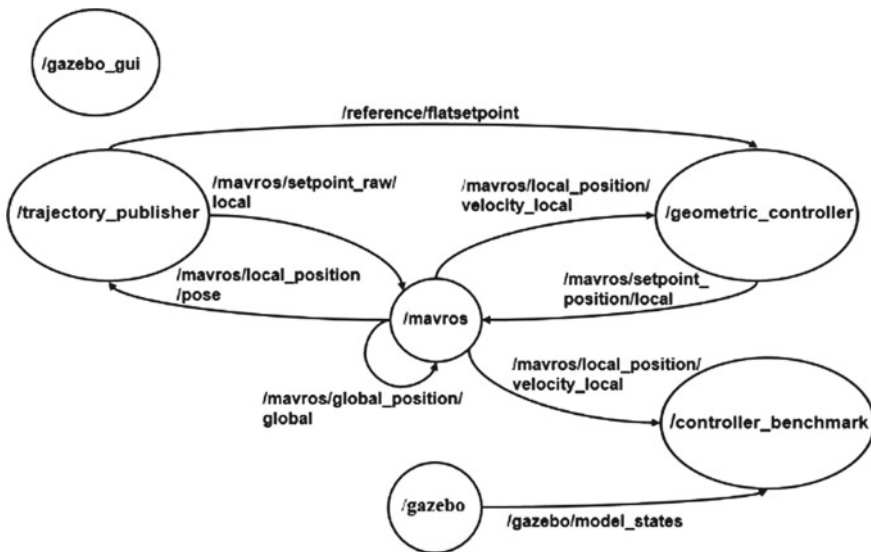


Fig. 1 Exchange of ROS messages between control model components

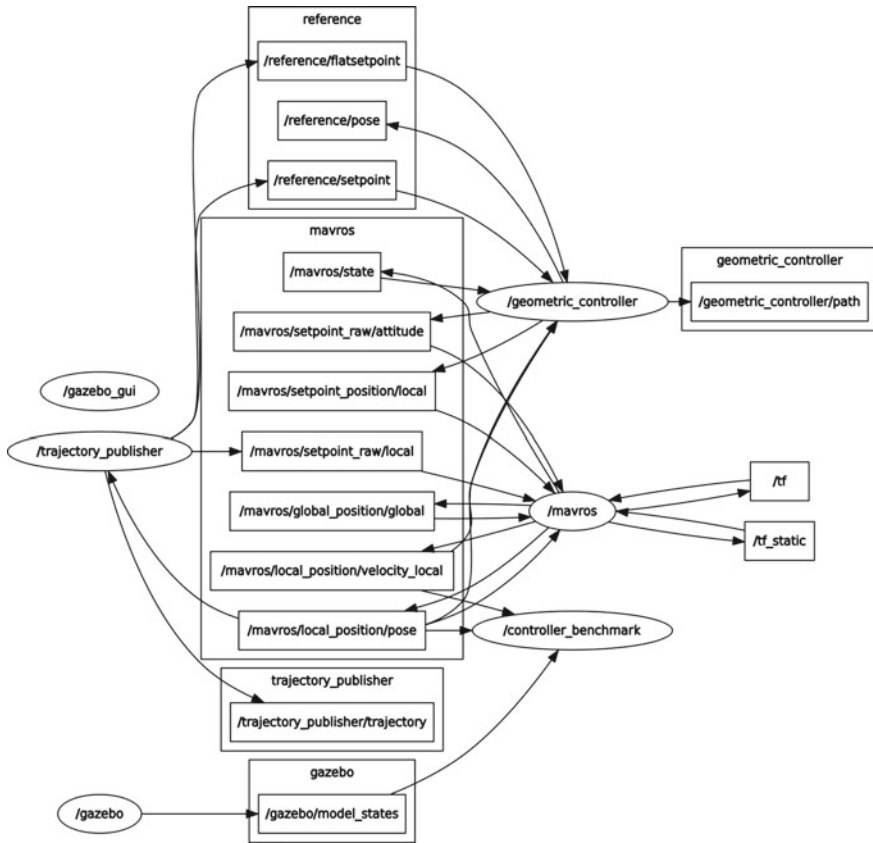


Fig. 2 Published and subscribed topics between ROS components

by MAVROS through MAVLink tunnel between PX4 and ROS. The geometric controller publishes and subscribes to the following topics:

1. *Published Topics:*

- mavros/setpoint\_raw/attitude(mavros\_msgs/AttitudeTarget): It consists of value for thrust and body rate calculated by controller.
- reference/pose(geometry\_msgs/PoseStamped): It describes the actual  $x, y, z$  coordinates of quadrotor.
- reference/pointPub(geometry\_msgs/Point): These are actual  $x, y, z$  coordinates published for error calculation.

2. *Subscribed Topics*

- reference/setpoint(geometry\_msgs/TwistStamped): It provides desired  $x, y, z$  coordinates of desired trajectory.
- mavros/state(mavros\_msgs/State): It shows the current state of the robot.

- `mavros/local_position/pose(geometry_msgs/PoseStamped)`: It provides the actual  $x$ ,  $y$ ,  $z$  coordinates of quadrotor.
- `gazebo/model_states(gazebo_msgs/ModelStates)`: It depicts the current state of the quadrotor in Gazebo environment.
- `mavros/local_position/velocity(geometry_msgs/TwistStamped)`: It consists of information regarding linear and angular velocities.

## 4.2 Trajectory Publisher

The trajectory publisher publishes continuous trajectories and provides them to the geometric controller and RVIZ tool. It includes already defined reference trajectories such as horizontal Geronon Lemniscate trajectory and horizontal circular trajectory. These trajectories are published at the rate of 10 ms. It publishes and subscribes the following ROS topics.

### 1. Published Topics

- `reference/trajectory(nav_msgs/Path)`: This is used to publish the desired trajectory to the rviz.
- `reference/setpoint(geometry_msgs/TwistStamped)`: It provides the desired position, velocity, and acceleration.
- `reference/targetPointPub(geometry_msgs/Point)`: It consists of desired  $x$ ,  $y$ ,  $z$  coordinates published for error calculation.

### 2. Subscribed Topics

- `mavros/local_position/pose(geometry_msgs/PoseStamped)`: It provides actual  $x$ ,  $y$ ,  $z$  coordinates of quadrotor.
- `mavros/local_position/velocity(geometry_msgs/TwistStamped)`: It contains desired position, velocity, and acceleration.

## 4.3 Controller Benchmark

It is used to evaluate the performance of the controller. It calculates the error using Eq. 8 as described above. The desired position and the actual position are provided to the controller benchmark by trajectory publisher and MAVROS. It publishes and subscribes the following ROS topics:

### 1. Published Topics

- `reference/error(reference/error)`: It is used to publish root-mean-square error.

## 2. Subscribed Topics

- mavros/local\_position/pose(geometry\_msgs/PoseStamped): It provides actual  $x$ ,  $y$ ,  $z$  coordinates of quadrotor.
- mavros/local\_position/velocity(geometry\_msgs/TwistStamped): It contains information regarding linear and angular velocities.
- reference/setpoint(geometry\_msgs/TwistStamped): It provides the desired position, velocity, and acceleration.

## 5 Simulations and Results Analysis

The simulations have been conducted in the Gazebo environment. The Gazebo environment helps to understand the complications and challenges for the proposed model, if it runs in the real environment. The RVIZ tool shows the better trajectory error tracking performance of the controller [19, 20]. Figures 3 and 4 show the simulation of quadrotor in Gazebo and visualization of our proposed control model, respectively. The trajectory tracking performance of the proposed control scheme is tested by using Geronon Lemniscate trajectory and circular trajectory.

The benchmark controller is used to analyze the performance of our proposed controller model by analyzing the root-mean-square (rms) position error (using Eq. 8). Figure 5a, b shows the plot of root-mean-square error ( $E_a$ ) with respect to time ( $t$ ) for circular trajectory and Lemniscate trajectory, respectively. The performance of the control model is further analyzed by increasing the velocity of quadrotor, and the rms position error with increased velocity is also computed. Figure 6a, b shows

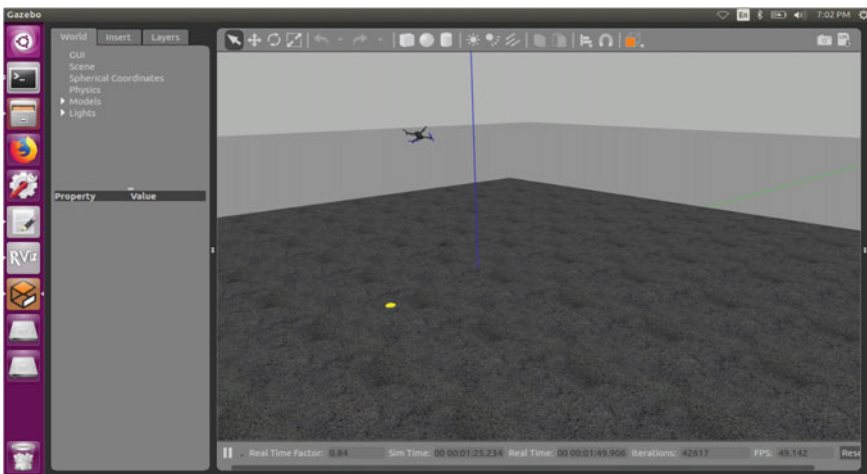


Fig. 3 Quadrotor's simulation in Gazebo

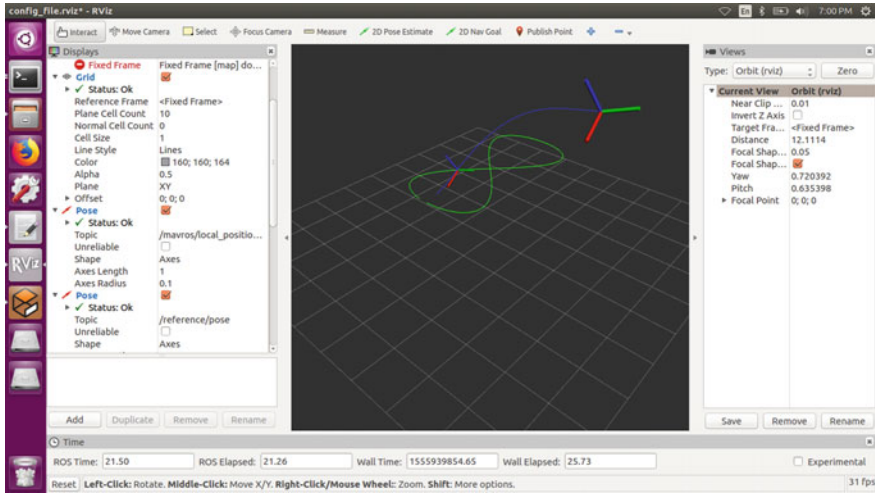


Fig. 4 RVIZ visualization of controller

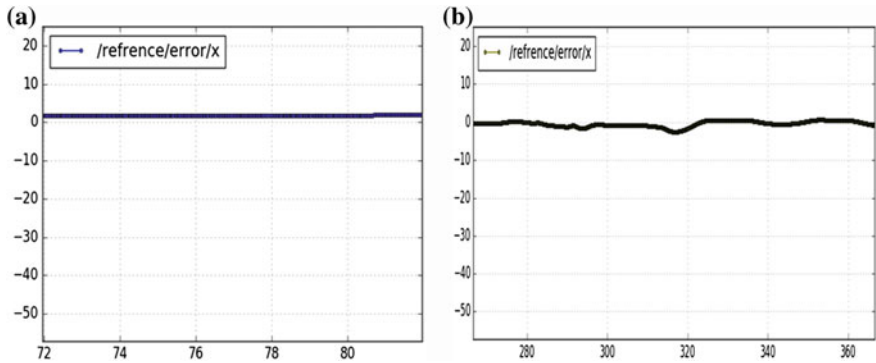
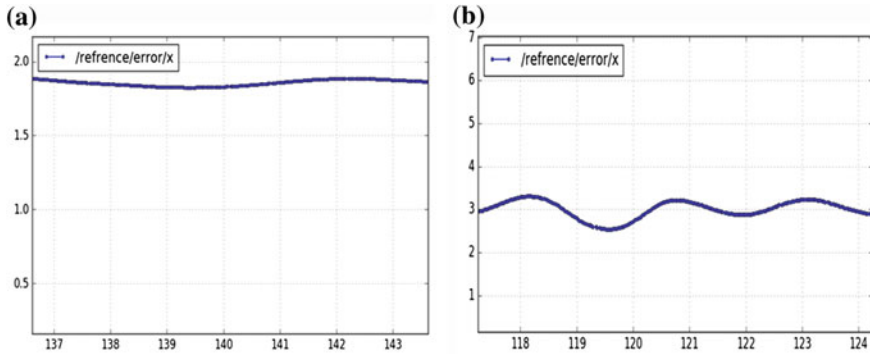


Fig. 5 a  $E_a$  w.r.t to  $t$  for circular trajectory and b  $E_a$  w.r.t to  $t$  for lemniscate trajectory

the plot of rms error ( $E_a$ ) with respect to time ( $t$ ) with change of velocity for circular trajectory (CT) and Lemniscate trajectory (LT), respectively.

From the plots, it has been observed that the proposed control model shows less stable behavior with increase in velocity and root-mean-square error for both circular and lemniscate trajectories which increase with time. There is an increase in the error value for lemniscate trajectory as compared to circular trajectory which shows that the controller becomes less stable while moving from simple to complicated trajectories.





**Fig. 6** **a**  $E_a$  w.r.t to  $t$  for CT with increased velocity and **b**  $E_a$  w.r.t to  $t$  for LT with increased velocity

## 6 Conclusion and Future Scope

In this paper, a control model is proposed to track the reference trajectory which uses feedback terms calculated from the tracking errors along with the feed-forward terms calculated from the reference trajectory using the differential flatness property of the quadrotor. The control law is implemented while considered rotor drag effects. The performance of the proposed control method is tested with benchmark controller which analyzes the root-mean-square position error with respect to time. The simulation results show that error value is increased for lemniscate trajectory compared with circular trajectory. For high velocity values, it is observed that the controller becomes unstable with increase in velocity and also for complicated trajectories. In future, we would like to improve the performance of the control model in order to achieve minimum trajectory tracking error with increase in velocity and for more complicated trajectories along with obstacle avoidance mechanism. We will further try to implement our controller as real-world working model.

## References

1. Burri M, Bloesch M, Taylor Z, Siegwart R, Nieto J (2018) A framework for maximum likelihood parameter identification applied on MAVs. *J Field Robot* 35(1):5–22
2. Mellinger D, Kumar V (2011) Minimum snap trajectory generation and control for quadrotors. In: 2011 IEEE international conference on robotics and automation, pp 2520–2525
3. Ferrin J, Leishman R, Beard R, McLain T (2011) Differential flatness based control of a rotorcraft for aggressive maneuvers. In: 2011 IEEE/RSJ international conference on intelligent robots and systems, pp 2688–2693
4. Bristeau P-J, Martin P, Salaün E, Petit N (2009) The role of propeller aerodynamics in the model of a quadrotor UAV. In: 2009 European control conference (ECC), pp 683–688
5. Faessler M, Franchi A, Scaramuzza D (2017) Differential flatness of quadrotor dynamics subject to rotor drag for accurate tracking of high-speed trajectories. *IEEE Robot Autom Lett*

- 3(2):620–626
6. Martin P, Salaün E (2010) The true role of accelerometer feedback in quadrotor control. In: 2010 IEEE international conference on robotics and automation, pp 1623–1629
  7. Mahony R, Kumar V, Corke P (2012) Multirotor aerial vehicles: modeling, estimation, and control of quadrotor. *IEEE Robot Autom Mag* 19(3):20–32
  8. Leishman RC, Macdonald JC, Beard RW, McLain TW (2014) Quadrotors and accelerometers: state estimation with an improved dynamic model. *IEEE Control Syst Mag* 34(1):28–41
  9. Burri M, Dätwiler M, Achtelik MW, Siegwart R (2015) Robust state estimation for micro aerial vehicles based on system dynamics. In: 2015 IEEE international conference on robotics and automation (ICRA), pp 5278–5283
  10. Bangura M, Mahony R (2017) Thrust control for multirotor aerial vehicles. *IEEE Trans Robot* 33(2):390–405
  11. Kai J-M, Allibert G, Hua M-D, Hamel T (2017) Nonlinear feedback control of quadrotors exploiting first-order drag effects. *IFAC-PapersOnLine* 50(1):8189–8195
  12. Omari S, Hua M-D, Ducard G, Hamel T (2013) Nonlinear control of vtol uavs incorporating flapping dynamics. In: 2013 IEEE/RSJ international conference on intelligent robots and systems, pp 2419–2425
  13. Svacha J, Mohta K, Kumar V (2017) Improving quadrotor trajectory tracking by compensating for aerodynamic effects. In: 2017 international conference on unmanned aircraft systems (ICUAS), pp 860–866
  14. Bangura M (2017) Aerodynamics and control of quadrotors. Ph.D. thesis, ANU
  15. Faessler M, Falanga D, Scaramuzza D (2016) Thrust mixing, saturation, and body-rate control for accurate aggressive quadrotor flight. *IEEE Robot Autom Lett* 2(2):476–482
  16. Faessler M, Fontana F, Forster C, Scaramuzza D (2015) Automatic re-initialization and failure recovery for aggressive flight with a monocular vision-based quadrotor. In: 2015 IEEE international conference on robotics and automation (ICRA), pp 1722–1729
  17. Nelder JA, Mead R (1965) A simplex method for function minimization. *Comput J* 7(4):308–313
  18. Hehn M, D’Andrea R (2014) A frequency domain iterative learning algorithm for high-performance, periodic quadrocopter maneuvers. *Mechatronics* 24(8):954–965
  19. [wiki.ROS.org/Documentation](http://wiki.ROS.org/Documentation), downloaded on 28/12/2018
  20. [dev.px4.io/autopilot](http://dev.px4.io/autopilot), downloaded on 02/01/2019

# Evaluation and Study of IoT Entrances



E. Sai Sravani, A. V. Sreehitha, A. Konda Babu and Durgesh Nandan

**Abstract** The people in the present scenario are leading a busy life which is filled with modern technology that changes rapidly. In such a rapid growth of the human race, technological developments were also increasing rapidly. At the beginning of the twenty-first century, technology turned its focus towards automation which leads to the development of new innovative technology called IoT. But it is a bit harder to implement because IoT is not just connecting hardware devices with the Internet, but it is the interconnection of devices with the Internet that should work with intelligence. To do that, we should require gateways and a cloud to store data. The main key element in the complete success of IoT is the gateway. A gateway is a device either a hardware device or a software program that connects the client and the server. The server or the cloud consists of data and this data from the cloud to server and vice versa can be done through these gateways. The functioning of a gateway is not only to relocate the information between the consumer and the server but also to maintain protocols and to ensure security to the continuous flow of data that is exchanged between the client and the server. In the above paper, we gave a complete gist regarding the IoT gateways and explain the working or functionalities of different gateways in different applications.

**Keywords** IoT gateways · Protocols · Sensors · Smart home · Automation

---

E. Sai Sravani · A. V. Sreehitha · A. Konda Babu  
Department of ECE, Aditya Engineering College, Surampalem, India  
e-mail: [saisravani819@gmail.com](mailto:saisravani819@gmail.com)

A. V. Sreehitha  
e-mail: [adurthisakunthala@gmail.com](mailto:adurthisakunthala@gmail.com)

A. Konda Babu  
e-mail: [kondababu.amaradi@aec.edu.in](mailto:kondababu.amaradi@aec.edu.in)

D. Nandan (✉)  
Accendere Knowledge Management Services Pvt. Ltd., CL Educate Ltd., New Delhi, India  
e-mail: [durgeshnandano51@gmail.com](mailto:durgeshnandano51@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_44](https://doi.org/10.1007/978-981-15-2329-8_44)

# 1 Introduction

Nowadays IoT is growing at a fast pace and the reason behind this success and growth is due to the automation of everything attracted the interest of everyone, and this emerging technology lays down the path for contracts technologies that are used to interpret data and produce an outcome [1]. The main important element in the working of the IoT system is IoT gateway. A gateway is nothing, but a device or software used to connect the client and the server (cloud). These IoT gateways are appreciated because there exists a control concerning the input and output [2].

Gateways are used to exchange the information among the clients and the server, and they are used to homogenize the protocols, data storage, process the data and also ensure secure data flow between the cloud and the customers. The gateways consist of many devices working together, i.e. software and hardware working together, and it needs to organize protocols and it needs to organize protocols to do this CoAp protocol [3] which is mainly used which ensures security and effective communication. In IoT, there should some connection between the physical world and the processing unit or the gateway to perform a certain task. In order to establish a connection with the physical world, we need sensors, GPS, scanners [4] (Figs. 1 and 2).

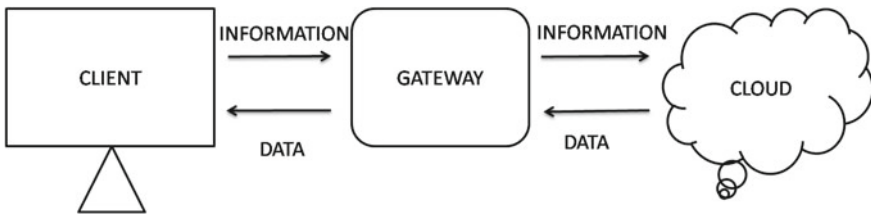


Fig. 1 Information passage through gateway

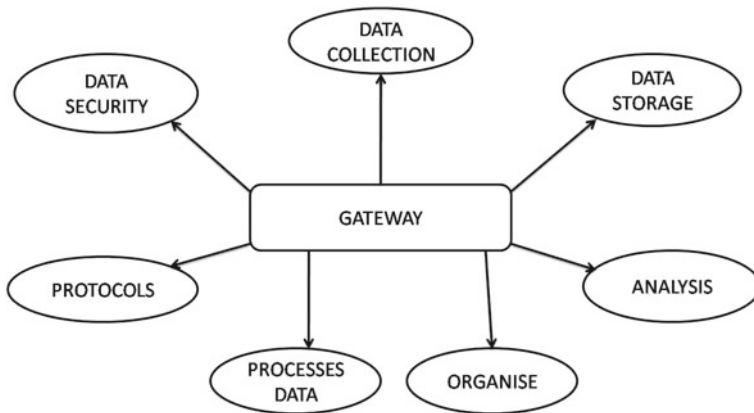


Fig. 2 Functionality of gateway

In IoT, everything must be automatic basing on the data present in the cloud. Sometimes, we even didn't have a client to transfer information to the gateway in such situation there should be some devices to transmit the information automatically by sensing the external environment. The sensing domain mainly consists of sensors to detect light, heat, temperature, humidity, moisture, etc. In the place of the client, these sensors will send the data to the gateways and further process is continued [5]. However, the technology of IoT involves continuous monitoring, and continuous communication must exist in between the device. This continuous interconnection may lead to continuous high-power usage [6].

In this paper, Sect. 2 focuses on the gateway protocols that are available depicting the functionalities of the IoT gateways. In Sect. 3, existing techniques in different fields such as agriculture, health and households were explained. Section 4 depicts completely the operations of IoT gateways. In the final section, the further execution of IoT gateways was listed. Finally, the theme of exploring IoT gateways is concluded.

## 2 Literature

There are many devices to transmit data and receive data such as Zigbee protocols and radio packet services. There are two modes or two types of gateways such as passive and semi-autonomic gateways. Router devices are used to route the data between devices. Firstly, information from sensors is taken and transmitted through routers and sent to further processing and output is generated by analysing the information. Libraries are maintained in the processing units of the devices for certain inputs. These IoT devices are connected to the Internet through Wi-Fi and GPS. However, the most preferred way of transferring data is through Wi-Fi mode. The main advantage of this technology is real-time monitoring can be done, that is, the owner/the key person can have the chance to take any action regarding the devices by continuous monitoring. The information in these IoT devices through a wireless network reduces complexity and it is economical [7].

Radio frequency identification system (RFID) is a smart technology that makes everything automatic. In this type of system detection, storage, analysis can be done through radio waves [8]. IoT needs automation, and that automation can be achieved through these RFID systems. In order to do that, we require sensors and microcontrollers to communicate and to follow the protocols [9]. The main task performed by the gateways is to process the data that means computation must be done and this can be achieved by a microprocessor, microcontroller, etc. [10]. The information in these gateways is through routers. Routers are those devices which are very helpful to pass the information from sensors to cloud or clients to server [11]. IoT which is increasing as a wildfire has faced many updates and development. The best development is the control of these IoT devices through smart phone from any place at any time [12].

### 3 Existing Techniques

**Gateway in smart homes:** Coming to the designing part, In HAN (home area network) if the devices are connected with the Internet, then the consumer can control these smart devices with the help of gateways at any time from any place. In HAN systems, the gateways should contain a circuit to control, to connect and to analyse. In HAN, sensors used to sense, actuators used to control, and after this, they are connected to the server. The sensors transmit data depending on the exact passwords or codes. Gateways cannot proceed directly without appropriate passwords. This type of techniques ensures high security. An Ethernet connection and high data rate enhance the working of HAN effectively [13]. In order to achieve automation, communication should be maintained in between the devices, i.e. machine-to-machine (M2M) communication [14].

**Gateway in agriculture:** IoT can be even applicable to agriculture to reduce human effort. In the agriculture sector, IoT can be used for various purposes such as nitrate monitoring in the soil. Since the nitrate level is very important for the growth of the plants. In agriculture in order to collect the data from the physical world, we use sensor networks which are wireless in this application the gateway used must have high-power capability. Similar to the gateways used in other applications, the gateways must be connected through Ethernet for nitrate monitoring in the soil [15].

There are many innovative techniques for continuous monitoring which increases productivity. A wireless sensor network (WSN) technique was developed to detect the chlorine level and water level in the soil [16]. An appropriate greenhouse is fully automatic and smart with reduced water wastage. During the period of growth of plants, there should be regular and continuous monitoring must be there. Sensing devices and actuators were used to gather and control data. The sensors and actuators work collaboratively to pass the information to the gateway where information management takes place and server, and it detects any unwanted situation generates securing signs like alarms, etc. [17]. In agriculture domain, the sensors were positioned in the soil since to observe the growth of the plants we should the moisture, salinity and PH level in the soil [10].

**Gateways in health:** IoT gateways development has increased the utility of the technology even into health for mobile monitoring of patients. In the field of health, IoT gateways must work along with cloud computing. One of the most popular methods for the transmission of the data triggering based transmission method for data transfer. This type of monitoring systems should have the capacity to act fast and smart because if there exists any delay that may result in loss of life of the patient. Whenever there is any serious change in the health of the patient the change must be communicated through the monitoring system to the gateway and from gateway, it must reach cloud where the next automated action is described [18].

Coming to the case of surgeries and bone fracture continuous monitoring is needed for the patients for the fast recovery by collecting the data from various sensors and by regular comparison with the information from the health specialist stored in the cloud. If there is any mismatch found, then alternative steps are taken [19]. The pain

of the patients can be even identified by the data of the facial expression of a patient and behavioural sign stored in the cloud [20].

### 4 Applications of IoT Gateways

IoT gateways have a broad range of applications, they are:

- Monitoring of climatic changes through an IIS [21].
- Applicable to the railway [22].
- Water management can be done.
- Can be extended to the service sector like hotels to exchange information regarding rooms when connected through the Internet [4].
- Transportation and vehicles automation [23].
- Lays down a path for new technologies such as cloud and fog computing [24].
- Disaster management can be achieved [25].
- Can be utilized in Industries [24, 26, 27].
- Used in the improvement of communication [12].
- Applicable to medical and biomedical for specific applications such as genetic cryptography [12]. All IoT structures are given in Fig. 3.

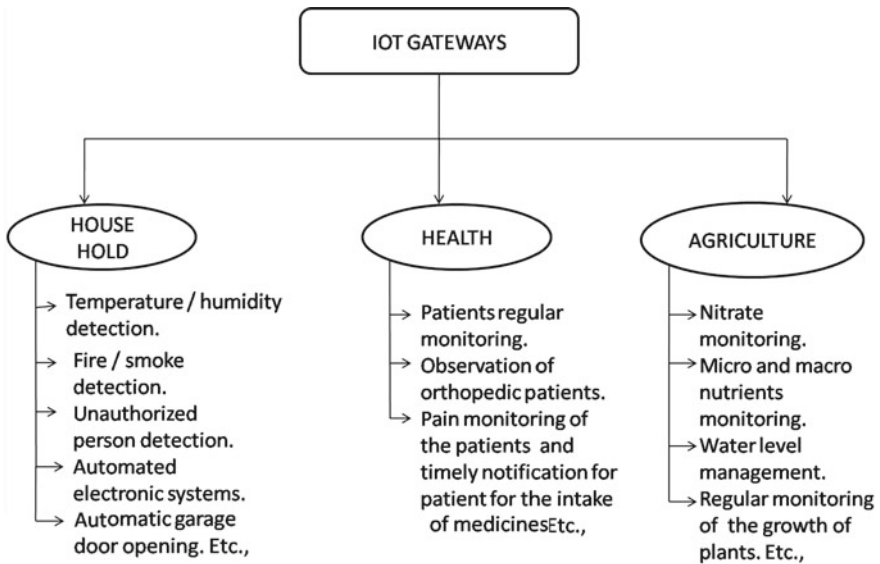


Fig. 3 IoT gateway structures

## 5 Result

In the vehicles, the monitoring system will help a lot to avoid the accidents and by using the GPS modules, we can easily trace the location of the vehicle. By using some sensors, we can easily prevent accidents in some major situation [14]. In this paper, we used to monitor the condition of the driver in different situations, and then, this is used to do preventions according to the situations, and checking the conditions that he/she is alcoholic or not, and he/she can drive the vehicle or not. And at the same time, the vehicle is to be trace by using the GPS and GPRS. The alcoholic detects sensor which is in front of the driver and it will check whether the person is consumed alcohol in a limit or not. If the person is alcoholic, then the alcohol sensor will detect it and it is not used to start the vehicle. And also, if the person is supposed to be sleepy, the vehicle is used to stop size and intimate the driver to take the rest for some time and also the engine is used to stop.

The final outcome of this paper is to explain the tasks performed by the IoT gateways in various areas such as household, health and agriculture. In the household, various detections are made. Such as temperature and humidity detection, fire or smoke detection, related to security unauthorized person detection and automation of electronic devices can be done. In health, regular and continuous observation of the patients which helps them to speed up their recovery. IoT is widely extended in the field of health and also take care regarding patients health through the identification of facial expression and gesture details that are stored in the cloud, and based on that info, the devices will automatically switch according to the requirement of the user. Spontaneous pain detection, especially in case of orthopaedic patients and regular and timely notifications for the patients for the regular intake of medicines in hospitals, will reduce the work of hospital staff even patients will easily understand the notification regarding medicines, and thus, IoT helps very effectively for the patients in the field of medicine. IoT gateways are used even in agriculture to monitor the elements such as chlorine, nitrate monitoring and to manage the water level in the soil and continuous monitoring of micro- and macro-nutrients can be done. All this is done by putting the sensors in the soil and comparing with the data stored in the cloud. By focusing more on the sensors position in the soil, we can easily detect even the growth of the weeds and pest on the side of the crop that damage the growth of crops. Security from the external environments such as animals which destroy crops for food, for example, pigs and buffaloes can be provided by detecting and notifying the farmer through alarm systems, etc. All these will definitely increase the product outcome by reducing the losses. Hence on extending IoT into agriculture, we can find solutions for many problems faced by the farmers during the time of sowing, harvesting, cutting, etc.



## 6 Conclusion

Development is the main leading idea behind the progress in the technologies which we are using up to date. Similarly, new technology has arrived named IoT emerged which attracted all the researchers and scientists. Many advancements are made in IoT gateways which lead to the development of various new technologies. The main theme of all these technologies is to reduce manpower. This paper gives an overview regarding the IoT which is the latest technology with the main idea of automation of everything in order to implement these gateways which play a prominent role through which the transfer of data takes place, and finally, the conclusion of this paper is to give an idea regarding the working of the gateway in various applications in various fields such as agriculture, health and smart homes and even many more applications were developed and are still developing through researchers. The concept of gateways always revolves around information or data of the main function of a gateway which is to integrate the data from IoT devices and to proceed with an output basing on the input.

## References

1. Wang K, Wang Y, Sun Y, Guo S, Wu J (2016) Green industrial internet of things architecture: an energy-efficient perspective. *IEEE Commun Mag* 54(12):48–54
2. Jutila M (2016) Adaptive edge router enabling internet of things. *IEEE Internet Things J* 3(6):1061–1069
3. Bormann C, Castellani AP, Shelby Z (2012) Coap: an application protocol for billions of tiny internet nodes. *IEEE Internet Comput* (2):62–67
4. Zhong CL, Zhu Z, Huang RG (2015) Study on the IoT architecture and gateway technology. In: 2015 14th international symposium on distributed computing and applications for business engineering and science (DCABES), IEEE, pp 196–199
5. Kang B, Kim D, Choo H (2017) Internet of everything: a large-scale autonomic IoT gateway. *IEEE Trans Multi-Scale Comput Syst* 3(3):206–214
6. Chen Y, Chiotellis N, Chuo LX, Pfeiffer C, Shi Y, Dreslinski RG, Grbic A, Mudge T, Wentzloff DD, Blaauw D et al (2016) Energy-autonomous wireless communication for millimeter-scale internet-of-things sensor nodes. *IEEE J Sel Areas Commun* 34(12):3962–3977
7. Jia X, Feng Q, Fan T, Lei Q (2012) RFID technology and its applications in internet of things (IoT). In: 2012 2nd international conference on consumer electronics, communications and networks (CECNet), IEEE, pp 1282–1285
8. Kumar S et al (2017) Slotted circularly polarized microstrip antenna for RFID application. *Radioengineering* 26(4):1025–1032
9. Manimuthu A, Ramesh R (2018) Privacy and data security for grid-connected home area network using internet of things. *IET Netw* 7(6):445–452
10. Gia TN, Rahmani AM, Westerlund T, Liljeberg P, Tenhunen H (2018) Fog computing approach for mobility support in internet-of-things systems. *IEEE Access* 6:36064–36082
11. Venkatesan VP, Devi CP, Sivaranjani M (2017) Design of a smart gateway solution based on the exploration of specific challenges in IoT. In: 2017 international conference on I-SMAC (IoT in social, mobile, analytics and cloud) (I-SMAC), IEEE, pp 22–31
12. El Kaed C, Khan I, Van Den Berg A, Hossayni H, Saint-Marcel C (2017) SRE: semantic rules engine for the industrial internet-of-things gateways. *IEEE Trans Ind Inf* 14(2):715–724

13. Huang J, Xing CC, Shin SY, Hou F, Hsu CH (2017) Optimizing M2M communications and quality of services in the IoT for sustainable smart cities. *IEEE Trans Sustain Comput* 3(1):4–15
14. Alahi ME, Pereira-Ishak N, Mukhopadhyay SC, Burkitt L (2018) An internet-of-things enabled smart sensing system for nitrate monitoring. *IEEE Internet Things J* 5(6):4409–4417
15. Harris N, Cranny A, Rivers M, Smettem K, Barrett-Lennard EG (2016) Application of distributed wireless chloride sensors to environmental monitoring: initial results. *IEEE Trans Instrum Meas* 65(4):736–743
16. Li X, Deng Y, Ding L (2008) Study on precision agriculture monitoring framework based on WSN. In: 2008 2nd international conference on anti-counterfeiting, security and identification, IEEE, pp 182–185
17. Verma P, Sood SK (2018) Fog assisted-IoT enabled patient health monitoring in smart homes. *IEEE Internet Things J* 5(3):1789–1796
18. Al-Fuqaha A, Guizani M, Mohammadi M, Aledhari M, Ayyash M (2015) Internet of things: a survey on enabling technologies, protocols, and applications. *IEEE Commun Surv Tutor* 17(4):2347–2376
19. Yang G, Jiang M, Ouyang W, Ji G, Xie H, Rahmani AM, Liljeberg P, Tenhunen H (2017) IoT-based remote pain monitoring system: from device to cloud platform. *IEEE J Biomed Health Inf* 22(6):1711–1719
20. Fang S, Da Xu L, Zhu Y, Ahati J, Pei H, Yan J, Liu Z (2014) An integrated system for regional environmental monitoring and management based on internet of things. *IEEE Trans Ind Inf* 10(2):1596–1605
21. Tsai WC, Zhu SX, Lu MH, Merzoug J, Yu C, Huang I (2017) An implementation of IoT gateway for home appliances control over cellular network. In: 2017 IEEE 8th international conference on awareness science and technology (iCAST), IEEE, pp 400–404
22. Jo O, Kim YK, Kim J (2017) Internet of things for smart railway: feasibility and applications. *IEEE Internet Things J* 5(2):482–490
23. Al Disi M, Djelouat H, Kotroni C, Politis E, Amira A, Bensaali F, Dimitrakopoulos G, Alinier G (2018) ECG signal reconstruction on the IoT-gateway and efficacy of compressive sensing under real-time constraints. *IEEE Access* 6:69130–69140
24. Fraile F, Tagawa T, Poler R, Ortiz A (2018) Trustworthy industrial IoT gateways for interoperability platforms and ecosystems. *IEEE Internet Things J* 5(6):4506–4514
25. Hakiri A, Berthou P, Gokhale A, Abdellatif S (2017) Publish/subscribe-enabled software defined networking for efficient and scalable IoT communications. arXiv preprint [arXiv:1711.05036](https://arxiv.org/abs/1711.05036)
26. Condry MW, Nelson CB (2016) Using smart edge IoT devices for safer, rapid response with industry IoT control operations. *Proc IEEE* 104(5):938–946
27. Hamici Z (2018) Towards genetic cryptography for biomedical wireless sensor networks gateways. *IEEE J Biomed Health Inf* 22(6):1814–1823

# Survey on the Impact of FSM Design for High-Performance Architecture Evaluation



K. Sowmya, P. Bujji Babu and Durgesh Nandan

**Abstract** In digital signal processing (DSP), the power consumption is more so, to decrease power and latency without affecting the other parameters, and mostly, the filters are designed using finite state machine (FSM). This paper gives a view of the multiplier architectures and its design issues for the expected level of performance. Literature states that the FSM approach is also a good choice in designing the multiplier architectures. In this paper, various design approaches are also described with the HDL modeling language, like in Verilog HDL, in building efficient multipliers. High-speed multipliers like Vedic multipliers are good in terms of speed and are considered as fastest and low-power multipliers.

**Keywords** Booth multiplier · Verilog HDL · Partial product generator

## 1 Introduction

In general, the hardware design of the multiplier is complex and takes more resources which are not desirable. So, many methods were introduced to overcome this problem. These methods give maximum frequency and reduce the required hardware resources. Combinational and pipelined implementations are two important methods to build hardware resources. The combinational approach is not suitable for hardware construction as it increases the critical path delay. So, most of the times, the pipelined method is used. The designers mainly implement the multipliers in FPGA (field programmable gate array) because the ASIC (application-specific integrated circuits) does not contain any DSP or dedicated multiplier block [1]. Multipliers are considered as the slowest elements of the system, so by the multipliers performance shows the performance of the system. Multipliers occupy large space compared to

---

K. Sowmya · P. Bujji Babu

Department of Electronics and Communication Engineering, Aditya Engineering College, Surampalem, Andhra Pradesh, India

e-mail: [bujjibabu\\_penumuchi@aec.edu.in](mailto:bujjibabu_penumuchi@aec.edu.in)

D. Nandan (✉)

Accendere Knowledge Management Services Pvt. Ltd., CL Educate Ltd., New Delhi, India

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication*

*Engineering*, Lecture Notes in Networks and Systems 106,

[https://doi.org/10.1007/978-981-15-2329-8\\_45](https://doi.org/10.1007/978-981-15-2329-8_45)

other components, so the speed, area and power have to be optimized for better performance of the system. Multiplication process is used in almost all applications such as communication, graphics, image enhancement, robotics, and machine vision. We get these multiplication results by adding two numbers [2]. Higher-order multiplications require a large number of adders to perform addition of two partial products. Mathematical problems can also be solved by Vedic sutras. The processors developed using these Vedic sutras are really fast and consume fewer hardware resources. This Vedic mathematics makes the complex formulae to simple one as it contains formulae according to the human mind [3]. Multiplication is carried out with signed and unsigned processes. For unsigned multiplication, there is no need to consider the sign, whereas the signed multiplication considers sign, and due to this, it may give the wrong results to avoid this we use an algorithm called Booth's algorithm. Multiplication in high-speed digital multipliers is carried out by Booth's algorithm. The conventional process produces number of iterations, so the Booth multiplier decreases the iterative steps [4]. As everyone knows, the signed/unsigned multipliers are implemented using two methods (a) tree-based architecture for programmable signed/unsigned redundant arithmetic and (b) signed/unsigned Booth multiplier to select the mode of operation [4]. HDL is a hardware description language and a coding language, and it is of two types VHDL and Verilog. Nowadays industries are using an updated version of Verilog called system Verilog. A state machine is one in which the mathematical models produces the values dependent on the input values and internal state. These state machines may give deterministic or non-deterministic behavior. The former is one in which the output is the same for internal state and input values, whereas it is different in later one [5]. This paper compares different multipliers. In Sect. 1, different methods and technologies of designing multipliers are discussed. Section 2 explains the design of different multipliers. Section 3 gives a comparison of different multipliers. Section 4 deals with the real-time applications of FSM-based multipliers and other multipliers. Section 5 gives the conclusion.

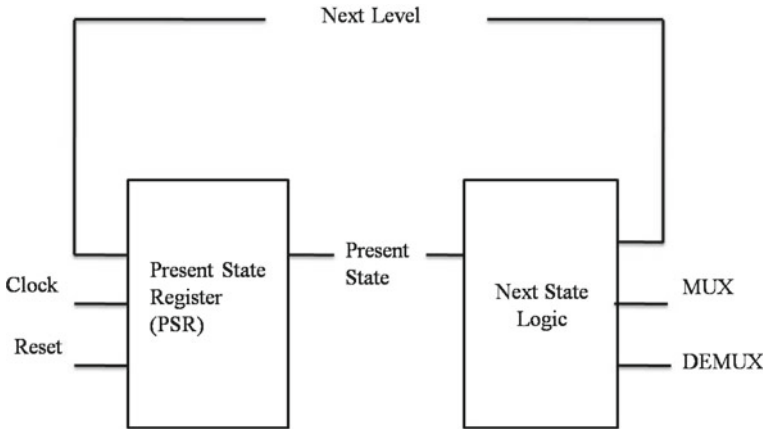
## 2 Literatures

Multiplication plays a major role in digital systems like DSP. Array multipliers are mostly used in VLSI but the computation time of these multipliers depends on the word length of operands. So, these operate at a low speed. At first, the  $n$ -PPs are generated, and these are summed up pairwise by a binary tree of RBAs by this  $n$ -bit multiplication can be done in time corresponding to  $\log 2n$ . The redundant binary representation consists of three values 0, 1, and  $T$ . The number elements required are  $n2$ . The addition can be done with constant time and does not depend on word length of operands, and as a result, it does not generate any carry propagation [6]. Previously used HDLs are available only to the highly qualified professionals and are complex and use only low-level description. It requires more time than a pure software solution. This HDL is modified and made it simple for software-based workers [7]. Compared to other components in ALU, the adder and multiplier make a large

amount of delay. Based on criteria like speed, power consumption and area, the ALUs are designed in one of the ECL, TTL or CMOS technologies [8]. The multipliers are of many types like Wallace, Radix-4, Radix-2, array, logarithm and conventional  $12 * 12$  multiplier, and array multiplier consumes less power and produces more delay, whereas the Booth multiplier (radix-4) will consume moderate power and is of speed [2]. As high-speed processors are increasing, the multipliers should also work as fast as possible. This can be achieved by Vedic mathematics as it contains formulae that are easily understood by the human mind. The multiplier is also a large source of power dissipation. Multiplication process needs more processing time than addition and subtraction [3]. FSM-based multipliers are further modified by adding carry look ahead adders than ripple carry adders. Due to this modification, the multiplier produces optimized performance [9]. High-speed multipliers are designed and developed to run in semi-custom environment. These are developed using both static and dynamic logic design techniques, and to get high-performance characteristics, both these logics must be combined [6]. Basic types of FSM are Mealy and Moore machines, and there are many other complex machines. Verilog and VHDL provide the capacity in writing various codes for FSM. Certain specified goals like performance and minimum resource utilization on chip are achieved by the choice of the code [9, 10]. Compared to other multipliers, Vedic multipliers are more efficient in terms of speed [11]. For the computational point view, there are two main building blocks such as adder and multiplier. Wallace signed multiplier circuit through modified Baugh-Wooley approach using standard conventional logic gates/cells, based on complementary pass transistor logic and have been validated with simulations, a layout vs. schematic check, and a design rule check. It is proved that the proposed multiplier is better and optimized, compared to its existing counterparts with respect to the number of gates, constant inputs, garbage outputs, hardware complexity and number of transistors required. The multipliers exhibit comparable faster, less power dissipation and smaller area [12–15].

### 3 Methodology

Multiplication can also be done by shift and add algorithm. During this shift and add algorithm process, the  $n$ -bit multiplicand can be multiplied with  $n$ -bit multiplier then it produces  $2n$ -bit product. This process results  $n$  number of PPAs and each PPA will shift toward left by one bit and then the addition is done directly. This addition is easy in simple multipliers, but it gives pointless results in large multipliers. This process involves five stages and three major parts, namely (1) partial product generator block, (2) data path block and (3) FSM block [1]. These three blocks are crucial to build a modified multiplier. The first block generates the required partial products. These partial products are added later to get the multiplication. PARTIAL PRODUCT GENERATOR BLOCK: This block consists of different types of adders, and it generates partial product from multiplicand and multiplier. The 36-bit adder performs all additions. Every 36-bit adder is a CLA. There are different techniques for



**Fig. 1** Partial product generator blocks

different multipliers, and some of them are discussed below. Array multiplier: This multiplier is very easily understood and two multiply two numbers we first calculate the partial products either from multiplicand or multiplier. A ripple carry adder is used to gather all the partial products. But this causes a delay in time because the carry is carried from the first. The array multiplier circuit uses add and shift algorithm. Wallace tree multiplier: Wallace tree multiplier is devised by Chris Wallace in 1964 and is used to increase the speed of the multiplication in a parallel multiplier. Instead of using ripple adders it uses a set of counters which adds all the partial products in a parallel fashion. To produce the final result a fast adder is used at the end of the multiplier. Wallace tree multiplier uses the carry save addition technique to decrease the latency [18]. Wallace multiplier consists of only  $O(\log n)$  reduction layers with  $O(1)$  propagation delay (Figs. 1 and 2; Table 1).

The below table describes the performance of different multipliers. Even though array multiplier is a simple multiplier, it has high delay, power consumption, area-delay product but it occupies less area compared to other multipliers (Fig. 3).

Radix multiplier which is having large area requirement is mostly used because of its low delay, low power consumption and low area-delay product. The graph above shows the performance of different multipliers.

## 4 Applications

It is used to perform multiplication in large multipliers. These are used where there is a delay in terms of time because using these, we can reduce the delay time. These are used where there is the use of high-frequency multiplication. These are also used in high speed, low power dissipation and low area applications. Mealy machines are used where there is the use of faster results, but this makes the design complex [17].

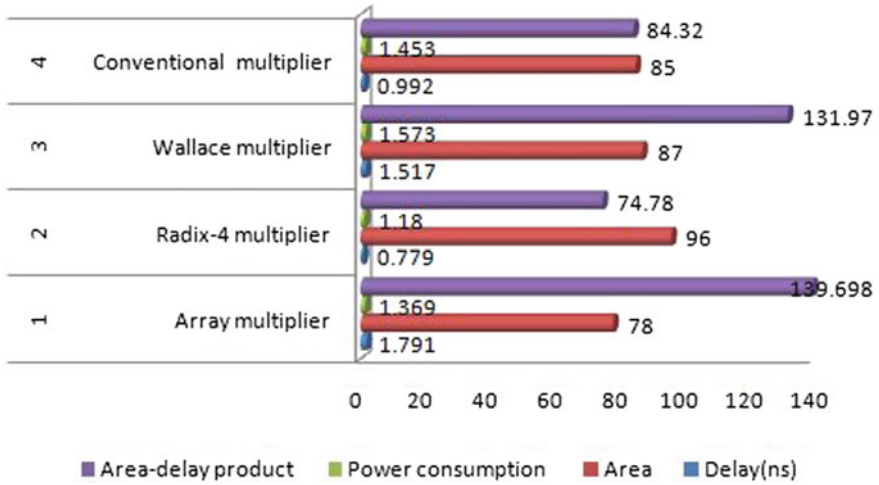
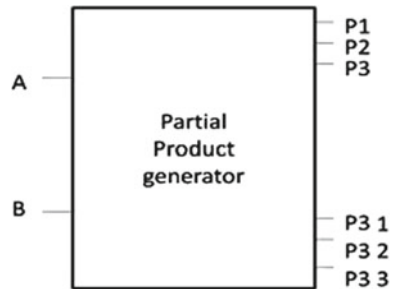


Fig. 2 Graphical representation of performance of different multipliers

Table 1 Comparison of different multipliers

S. No.	Multiplier	Delay (ns)	Area	Power consumption	Area-delay product
1	Array multiplier	1.791	78	1.369	139.698
2	Radix-4 multiplier	0.779	96	1.18	74.78
3	Wallace multiplier	1.517	87	1.573	131.97
4	Conventional multiplier	0.992	85	1.453	84.32

Fig. 3 Graphical representation of performance of different multipliers



The real-time applications of FSM-based multipliers are vending machines, ATM, trace signaling, etc.

## 5 Result and Conclusion

In the above table, different multipliers are compared with different parameters like delay, area, power consumption and area-delay product. Compared to remaining multipliers, the Radix-4 multiplier gives less delay, less power consumption and area-delay product, but it occupies more area which is the only drawback in the Radix-4 multiplier. Array multiplier, as it is the first ever multiplier discovered it gives more delay and more area-delay product than all other multipliers. Coming to Wallace multipliers although it is used to increase the speed it consumes more power compared to all other multipliers. So finally, Radix-4 multiplier is the most used multiplier as it gives less delay, occupies less area and consumes less power. In the above, we discussed different types of multipliers and their design using different techniques and different parameters are compared. From the above discussion, we also came to know about HDL language and different types of HDL languages. We also came to know that these FSM's can be designed using these languages and their efficiency in terms of better design constraints. These types of designs can give maximum performance in both FPGA and ASIC. These type of designs gives maximum frequency when they are designed at the gate level. It also uses less output latency to calculate the result. This latency can be increased by adding a greater number of adders. Though there exist a number of design methods, the pipelined structure gives the better performance of the system.

## References

1. Ka AA, Rahman A, Mahjabeen B, Rahman M (2014) An efficient design of FSM based 32-bit unsigned high-speed pipelined multiplier using VERILOGHDL, pp 164–167
2. Kaur J, Sumit K (2016) Analysis and comparison of different multiplier. *Int J Recent Innov Trends Comput Commun*, ISSN 2321-816975
3. Maiti A, Chakraborty K, Sultana R, Maity S (2016) Design and implementation of 4-bit vedic multiplier. *Int J Emerg Trends Sci Technol* 3(5):3865–3868
4. Li Q, Liang G, Bermak A (2010) A high-speed 32-bit signed/unsigned pipelined multiplier. In: 2010 fifth IEEE international symposium on electronic design, test and applications, pp 207–211
5. Chiuchisan I, Potorac AD, Graur A (2010) Finite state machine design and VHDL coding techniques. *Dev Appl Syst* 75
6. Takagi N, Yasuura H, Yajima S (1985) High-speed VLSI multiplication algorithm with a redundant binary addition tree. *IEEE Trans Comput* 9:789–796
7. Ogoubi E, David JP (2004) Automatic synthesis from high level ASM to VHDL: a case study. In: The 2nd annual IEEE Northeast Workshop On Circuits And Systems, NEWCAS 2004, pp 81–84
8. Raahemifar K, Ahmadi M (2000) Fast 32-bit digital multiplier. In: 2000 IEEE international symposium on circuits and systems. Emerging technologies for the 21st century. Proceedings (IEEE Cat No. 00CH36353), vol 5, pp 625–628
9. Bokade S, Dakhole P (2016) CLA based 32-bit signed pipelined multiplier. In: 2016 international conference on communication and signal processing (ICCSP), pp 0849–0852



10. Kernhof J, Beunder M, Hoeffinger B (1988) High speed CMOS adder and multiplier modules for digital signal processing in a semicustom environment. In: ESS-CIRC'88: fourteenth European solid-state circuits conference, pp. 62–65
11. Sinthura SS, Begum A, Amala B, Vimala A, Aparna VV (2018) Implementation and analysis of different 32-bit multipliers on aspects of power, speed and area. In: 2018 2nd international conference on trends in electronics and informatics (ICOEI), pp 312–317
12. Nandan D, Kanungo J, Mahajan A (2016) An efficient VLSI architecture design for antilogarithmic converter by using the error correction scheme
13. Nandan D, Kanungo J, Mahajan A (2017) An efficient VLSI architecture for iterative logarithmic multiplier. In: 2017 4th international conference on signal processing and integrated networks (SPIN), pp 419–423
14. Nandan D, Mahajan A, Kanungo J (2017) An efficient antilogarithmic converter by using 11-regions error correction scheme. In: 2017 4th international conference on signal processing, computing and control (ISPCC), pp 118–121
15. Poornima M, Patil SK, Shivukumar SK, Sanjay H (2013) Implementation of multiplier using vedic algorithm. *Int J Innov Technol Explor Eng (IJITEE)* 2(6):219–223
16. Nandan D, Kanungo J, Mahajan A (2017) An efficient VLSI architecture design for logarithmic multiplication by using the improved operand decomposition. *Integration* 58:134–141
17. Nandan D, Kanungo J, Mahajan A (2018) 65 years journey of logarithm multiplier. *Int J Pure Appl Math* 118:261–266
18. Nandan D, Kanungo J, Mahajan A (2018) An efficient architecture of iterative logarithm multiplier. *Int J Eng Technol* 7(2–16):24–28
19. Nandan D, Kanungo J, Mahajan A (2018) An error-efficient gaussian LTER for image processing by using the expanded operand decomposition logarithm multiplication. *J Amb Intell Human Comput* 1–8

# A Taxonomy and Survey of Data Partitioning Algorithms for Big Data Distributed Systems



Quadri Waseem, Mohd Aizaini Maarof, Mohd Yazid Idris and Amril Nazir

**Abstract** Data partitioning is a backbone of distributed systems that boost the performance of big data applications, especially in distributed systems. In past years, many data partitioning algorithms have been developed which had improved the big data management and its processing for the real-time applications of the big data stores. Furthermore, the feature of “elasticity” to the data partitioning has removed the need for human interaction while handling the big data applications on the distributed system during the high workloads and skews. In this survey, a taxonomy is proposed that characterizes and classifies various types of data partitioning algorithms, which will help to identify the current limitations in the state of the art and will extend the state of the art to improve the enhancements for the effective and efficient performance of the big data stores on distributed systems. The taxonomy not only highlights the design, the similarities, and the differences within state of the art for different types of data partitioning algorithms but also identifies the areas that need further research.

**Keywords** Big data · Distributed systems · Data partitioning · Elasticity

---

Q. Waseem

Independent Researcher, Srinagar 190005, Jammu and Kashmir, India

M. A. Maarof (✉)

School of Computing, Universti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

e-mail: [aizaini@utm.my](mailto:aizaini@utm.my)

M. Y. Idris

Media and Game Innovation Centre of Excellence, Institute of Human Centered Engineering,

Universti Teknologi, Johor Bahru, Malaysia

e-mail: [yazid@utm.my](mailto:yazid@utm.my)

A. Nazir

Department of Computer Science and Engineering, American University of Ras al Khaimah, Ras

al Khaimah PO Box: 10021, UAE

e-mail: [mohd.nazir@aurak.ac.ae](mailto:mohd.nazir@aurak.ac.ae)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication*

*Engineering*, Lecture Notes in Networks and Systems 106,

[https://doi.org/10.1007/978-981-15-2329-8\\_46](https://doi.org/10.1007/978-981-15-2329-8_46)

## 1 Introduction

Big data's "5 V's" [1] means a huge amount of heterogeneous data [2, 3] which is generated from many sources and activities [4] and is therefore used by various industries for analytic purpose. Big data has valuable knowledge, less cost, least processing time, and real-time features [5]. Big data management is a way of centralizing storing, managing, processing, and querying the huge volume of different available data in any format [1, 2, 5, 6], which is really a big challenge [7]. Two types of processing can be done on big data. The first one is batch processing [8]. Here, the MapReduce framework (Hadoop) supports the batch processing of bulk datasets because of its advantages like high-throughput I/O and its suitability of processing large volumes of data for a longer time [9]. However, they fail to process the non-batch (real-time data) [4, 10], due to its limitations which include slow processing, not designed for high performance and their high latency [2, 11]. The second one is non-batch processing (real-time processing) [2, 12], which includes the real-time OLTP online transaction processing database management DBMS systems with the variable workload and spike in traffic. They are dependent on shared-nothing architecture and use main memory for the processing and scalability while maintaining the acid guarantee of the transactions. They provide data parallelism through partitioning for the OLTP distributed DBMSes [2, 11, 13–16]. In addition to this, the advantages include low cost, high extensibility, and high availability [17]. OLTP DBMS uses the H-Store, which is an OLTP system that operates on a distributed cluster of shared-nothing machines using in-main memory [7, 11, 18, 19]. The OLTP databases use the elasticity concept to scale up during the changing workloads for load balancing. [13, 20]. Furthermore, the OLTP databases are tuned manually for changing workloads and they require periodic configurations for scaling out during need basis. None of the databases provides the full "auto-scaling" including acid guarantee assured. Thus, elasticity is the available option which scales the database in such a way that it automatically adds or removes the servers on the need basis to meet the throughput and latency requirements of the OLTP databases with a controlled acid guarantee [4, 13].

## 2 Literature Review

There is substantial literature available for the data partitioning algorithms related to big data applications. A survey of this kind gives an insight into not only a comparative analysis of the performance for the available strategies but also suggests the suitability of these strategies for the big data applications.

Das et al. [21] proposed a "ElasTras" elastic, scalable, and self-managing transactional database for the cloud. It is helpful during on-demand live database migration technique, used for load balancing. Pavlo et al. [22] proposed cost-based horticulture-skew-aware automatic database partitioning for shared-nothing OLTP systems. It

minimizes the number of distributed transactions while maintaining the load. Quamar et al. [23] proposed a “Sword” scalable workload-aware data placement for transactional workloads, which is based upon a hypergraph compression technique for load balancing and availability. Schall et al. [24] proposed adaptive dynamic physiological partitioning on a shared-nothing database cluster. It handles repartitioning events using the lowest query runtimes. Taft et al. [25] proposed a two-tiered partitioning fine-grained elastic partitioning “E-Store” for distributed transaction processing systems—it takes care of reconfiguration by taking care of hot and cold tuples accordingly. It automatically scales resources during load spikes through tree schema (single partitioning key). Xu N et al. [26] proposed a “LuTe” an efficient fine-grained partitioned technique for distributed systems, which uses lookup tables. It minimizes the number of nodes during load and increases the throughput. Serafini et al. [6] proposed a coarse-grained “Accordion”—elastic scalability for database systems which supports ACID transactions (local or distributed). Pinar Tözün et al. [27] proposed a physiological partitioning (“PPL”) approach that partitions logically, the physical data of shared-everything systems during dynamic load balancing using repartitioning. Fetai et al. [16] proposed a Cumulus-workload-driven adaptive data partitioning and distribution—it uses adaptive static partitioning mechanism. It works with local transactions to minimize the repartitioning. Kaiji et al. [28] proposed a system for scheduling repartitioning of OLTP workloads with an acid guarantee with fast executing the repartitioning. Serafini et al. [29] proposed a “clay”—a fine-grained adaptive (two-tier) online partitioning that supports fine-grained partitioning using clump of hot and cold tuples. It is elastic and reactive which follows (both tree and non-tree multi-key transactions. Asad et al. [30] proposed an “AdaptCache” graph partitioning technique for distributed systems. It is a cache framework for load balancing. John et al. [31] proposed a “Redynis”—a heuristic-based dynamic repartitioning for tuples based on shared something architecture—it is best for a geo-distributed set of key-value, which reduces network latency.

### 3 Data Partitioning

Data partitioning is a fragmentation of a logical database into separate independent blocks for the scaling of data. It is an important requirement for the performance of in-memory systems. The partitioned data provide the parallelism. The shared-nothing scale-out parallelism (distributed computation) should have the best data partitioning strategy to obtain performance, load balancing, and least cross-partition coordination [9].

### ***3.1 Data Partitioning Taxonomy***

Many authors have contributed to the state of the art. The most and the latest classification is mentioned in [11]. The authors proposed a classification of data partitioning algorithms based on the strategies of their approaches as well as on unit of operation applied to them.

The contribution in this survey is more toward the proposed taxonomy, which classifies the data partitioning strategies into different categories like schema based, nature based, transaction based, and action based.

### ***3.2 Different Categories of Data Partitioning Techniques***

Here the data partitioning algorithms are divided into two basic modes, i.e., online and offline data partitioning algorithms. The online data partitioning algorithms are further divided into four subcategories: (1) schema-based category, which includes a tree schema-based algorithm, hence accessing a single “tree schema.” It is used to minimize the number of multi-partition transactions. It works without considering the distributed transactions [6, 13, 25] and another one as non-tree-based algorithms, which include non-tree schemas. It considers distributed transactions minimally. (2) Nature-based category, which includes elastic and non-elastic algorithms, representing the nature of the databases to expand and contract its cluster-based on demand. (3) Transaction-based category, which includes the local and the distributed transaction-based algorithms, representing the distribution of transaction within the single server or distribution of partitions over different servers. (4) Action-based category, which includes the reactive and the proactive algorithms, representing the effectiveness of the database in responding before and after the load changes (Fig. 1).

Because of space limitation, detailed explanation of categories cannot be presented here but Table 1 provides the comparison of the algorithms for a better understanding.

### ***3.3 Evaluation Criteria for Data Partitioning Strategies***

The first evaluation criteria for the data partitioning strategies are “load balancing” that indicates how the strategy can balance the load among multiple servers during reconfiguration. The second is the “Consistency” (acid guarantee), which includes the transaction uniformity during the repartitioning. The third is the “Scalability,” which includes adding or removing of data partitions according to workload variations. The fourth is “Elasticity,” which includes online adding of servers or data partitions based upon the load variations without human interaction. The fifth is the “Transaction Handling” (distributes/local), which means the distribution of a transaction within a single server or among different servers. The sixth is the “Cost,” which measures

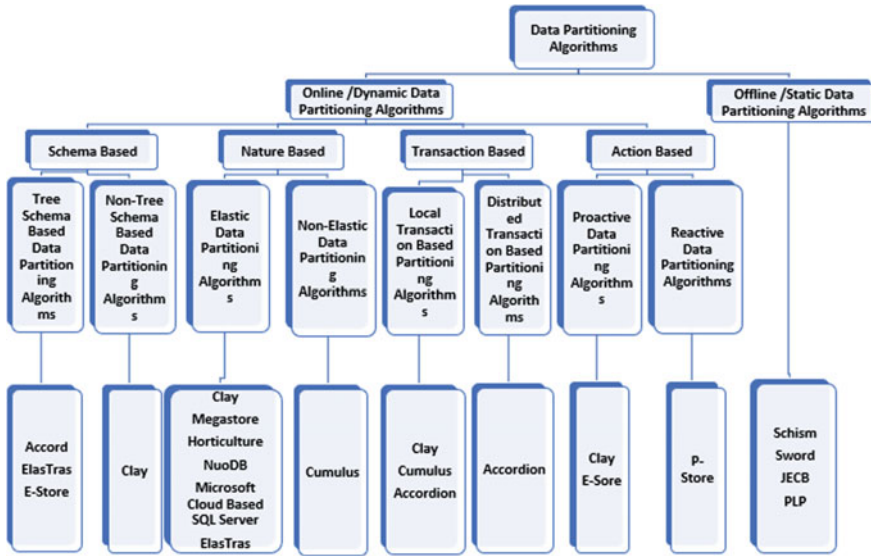


Fig. 1 Taxonomy of data partitioning algorithms

the expenses of migration and reconfiguration. The seventh is the “Fault Tolerance,” which measures the capacity to work smoothly during the failure.

The eighth is the “Service Level Agreement,” which includes the agreement between the service provider and the client. The ninth is the “Recovery, which includes the availability of data before and after the reconfiguration. The tenth is the “Performance,” which considers both high throughput and the low latency.

In Table 2, it is observed that some of the strategies maintain the load balancing, some maintain acid guarantee, and some decrease cost. However, none of the data partitioning strategies had covered all “Evaluation Criteria” (parameters). Unfortunately, so far none of the strategies have talked about the “Fault Tolerance,” the service level agreement (SLA) and the “Recovery” (availability).

### 3.4 Evaluation of Performance for Different Data Partitioning Strategies

Each strategy mentioned in Table 2 has its performance contribution, the details of which are mentioned as: (1) ElasTras strategy increased the throughput to 0.2 million transactions/minute). (2) Sword strategy increased the throughput from 165 transactions/second to 904 transactions/second for 8 partitions. (3) Dynamic physiological strategy performance details are unavailable in the paper. (4) E-store strategy

**Table 1** Classification of data partitioning algorithms

Online reconfiguration/dynamic		Offline reconfiguration static		
Tree schema (single partitioning key)	General schema/Non-tree (multi-key transactions)	Elastic/Non-elastic	Local/Distributed transactions	Reactive/Proactive
<p>Accordion—coarse-grained elastic partitioning</p> <p>ElasTras—achieve elastic scaling on complex structures by limiting transactions to a single partition</p>	<p>Clay (both tree and general schema) fine-grained adaptive partitioning clay adopts a two-tiered approach that supports fine-grained mapping for single tuples</p>	<p>Clay-elastic</p> <hr/> <p>Megastore-elastic (Automatic entity—grouping for OLTP workloads) similar to clay—elastic</p>	<p>Clay-local Cumulus-local</p> <hr/> <p>Accordion-local and distributed</p>	<p>Clay-reactive</p> <hr/> <p>E-store-reactive</p>
<p>E-store-fine-grained and coarse-grained partitioning</p> <p>Technique: these three systems assume that the database is partitioned a priori into a set of static blocks, and all tuples of a block are moved/migrated together at once</p> <p>Doesn't consider distributed transactions</p>	<p>Technique: uses dynamic blocks, called dumps that are created on-the-fly by monitoring the workload when a reconfiguration is required. It is incremental, thereby minimizing the cost of data migration by reducing/minimize the distributed transactions(reduce/minimize the amount of data migrated and cost) balance load</p>	<p>Horticulture-elastic</p> <hr/> <p>NuoDb—elastic achieve elastic scaling on complex structures by limiting transactions to a single partition</p>		<p>Schism—it builds an access graph similar to heat graph of clay and uses Metis</p> <p>Sword performs incremental partitioning adjustments</p> <p>JECB provides a partitioning strategy to handle complex schemas, but the focus is on scalable partitioning for large clusters</p> <p>PLP-technique to address partitioning in a single-server, shared memory system</p>

(continued)

**Table 1** (continued)

Online reconfiguration/dynamic		Offline reconfiguration static
	<p>Microsoft's cloud-based SQL server (adapting Microsoft SQL server for cloud computing—elastic)</p> <ul style="list-style-type: none"> <li>– Achieve elastic scaling on complex structures by limiting transactions to a single partition</li> </ul> <p>Cumulus-non elastic</p> <p>ElasTras-elastic and fault tolerance and recoverable</p>	



**Table 2** Evaluation criteria for data partitioning strategies

Evaluation criteria (parameters)	Strategy names and the year of publication											
	ElasTras 2010	Sword 2013	Dynamic physiological 2014	E-store 2014	LuTe 2014	Accordion 2013	PPL 2015	Cumulus 2015	SOAP 2015	Clay 2016	AdaptCache 2016	Redynis 2017
Load balancing	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Consistency	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Scalability	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Elasticity	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✗	✓
Transactions (distributed or local)	-	Local	-	Local	Local	Distributed	-	Local	-	Local	Distributed	-
Cost	✓	✗	✗	✗	✗	✗	✗	-	✗	✗	✗	-
Fault tolerance	-	-	-	-	-	-	-	-	-	-	-	-
Service level agreement	-	-	-	-	-	-	-	-	-	-	-	-
Recovery	-	-	-	-	-	-	-	-	-	-	-	-
Performance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

✓ Increased  
 ✗ Decreased  
 - Not discussed

increases the throughput of up to 130%, and the latency was reduced to 80% as compared to single-tier approaches. (5) LuTe strategy increased the throughput about 20–70% in comparison with other state of the art. (6) Accordion strategy was increased as compared to MILP-SP, Kairos-SP, greedy, and equal algorithms. (7) PLP strategy increased the performance of 50%. (8) Cumulus strategy increased the performance percentage when compared to the static partitioning approaches. (9) SOAP strategy performance details are unavailable in the paper. (10) Clay strategy increased the throughput  $15 \times$  better and 99% lower latency. (11) AdaptCache strategy performance details are unavailable in the paper. (12) Redynis strategy performance details are unavailable in the paper.

## 4 Advantages, Disadvantages, and Future Directions of Different Data Partitioning Strategies

“E-store” is designed for transactions that access a single partitioning key (with a “tree schema”), i.e., locally. They are not suitable for distributed transactions [25]. It can be extended to support more complex workloads and multi-partition transactions while considering their cost and the memory consumption [13, 25]. “Clay” [29] generalizes the E-Store approach to multi-key transactions (non-tree schemas). It balances the workloads without creating new distributed transactions. Its future work includes combining the predictive modeling to scale these complex workloads proactively. “Cumulus” is like clay, which attempts to minimize distributed transactions through adaptive repartitioning, but it does not support elasticity [16]. “Accordion” is a dynamic data placement system for partition-based DBMSes that support ACID transactions (local or distributed). It has future work for elasticity mechanisms that instead of considering split and merge partitions at run-time based on load rather than considering partitions as pre-defined partitions [6]. “ElasTras” support large single-tenant database instances and many small multi-tenant databases. It does not support databases that are partitioned across several nodes [21]. Its implementation with model-based elasticity can be investigated in future rather than using the current rule-based elasticity. Many partitioning techniques use the “squall” which is a live migration technique. This technique can be modified with automatic self-tuning so as to find the optimal chunk size and remove end delays between the chunks during the live migration of the data. Future work should also investigate whether other design alternatives to squall are better suited for moving large amounts of data with high-throughput workloads or not [13].

In the future, it would be great to see the implementation of the machine learning algorithms for the data partitioning strategies. Other suggestions for the future will be the implementation in such a way that the complex workload handling, hybrid form of general and tree schema, hybrid form of reactive and proactive methods, cost, load balancing, and memory utilization should be considered as an enhancement in big data distributed systems. Another future direction would be the utilization of the

machine learning techniques for the data migration between the distributed networks while maintaining the memory and the cost.

## 5 Conclusion

Parallelism and elasticity are the two best solutions to handle the high demand for huge real-time online requirements of big data in distributed environments (cloud). The parallelism provides the strong performance, load balancing, and the least cross-partition coordination while as elasticity provides the online feature to add or move the data chunks based on the need purpose. Enhancing these two areas will provide efficient in-memory processing and storage for the big data OLTP databases. This paper had outlined some of the research work (taxonomy, compare, evaluation, and future work) that needs to be discussed to determine the best optimal solutions for the effective and efficient performance of big data stores on distributed systems.

## References

1. Emani CK, Cullot N, Nicolle C (2015) Understandable big data: a survey. *Comput Sci Rev* 17:70–81
2. Hu H, Wen Y, Chua T, Li X (2014) Toward scalable systems for big data analytics: a technology tutorial. *IEEE Access* 2:652–687
3. Sharma S, Tim US, Wong J, Gadia S, Sharma S (2014) A brief review on leading big data models. *Data Sci J* 14–041
4. Shahrivari S (2014) Beyond batch processing: towards real-time and streaming big data. *Computers* 3:117–129
5. Kaur P, Monga AA (2016) Managing big data: a step towards huge data security. *Int J Wirel Microw Technol* 2:10
6. Serafini M, Mansour E, Abounaga A, Salem K, Rafiq T, Minhas UF (2014) Accordion: elastic scalability for database systems supporting distributed transactions. *Proc VLDB Endow* 7(12):1035–1046
7. Siddiqua A, Hashem IAT, Yaqoob I, Marjani M, Shamshirband S, Gani A, Nasaruddin F (2016) A survey of big data management: taxonomy and state-of-the-art. *J Netw Comput Appl* 71:151–166
8. Assunção MD, Calheiros RN, Bianchi S, Marco AS, Netto MA, Buyya R (2015) Big data computing and clouds: trends and future directions. *J Parall Distrib Comput* 79:3–15
9. Zhang H, Chen G, Ooi BC, Tan KL, Zhang M (2015) In-memory big data management and processing: a survey. *IEEE Trans Knowl Data Eng* 27(7):1920–1948
10. Dean J, Ghemawat S (2008) Mapreduce: simplified data processing on large clusters. *Commun ACM* 51(1):107–113
11. Phansalkar S, Ahirrao S (2016) Survey of data partitioning algorithms for big data stores. In: 2016 fourth international conference on parallel, distributed and grid computing (PDGC), IEEE, pp 163–168
12. Singh D, Reddy CK (2015) A survey on platforms for big data analytics. *J Big Data* 2(1):8
13. Taft R (2017) Elastic database systems. PhD thesis, MIT, Cambridge
14. Tatarowicz AL, Curino C, Jones EPC, Madden S (2012) Lookup tables: fine-grained partitioning for distributed databases. In: ICDE: IEEE Computer Society

15. Stonebraker Michael (1986) The case for shared nothing. *IEEE Database Eng Bull* 9(1):4–9
16. Fetai I, Murezzan D, Schuldt H (2015) Workload-driven adaptive data partitioning and distribution—the Cumulus approach. In: 2015 IEEE international conference on big data (big data), IEEE, 2015, pp 1688–1697
17. Valduriez P (2009) Shared-nothing architecture encyclopedia of database systems. In: Liu L, Özsu MT (eds) pp 2638–2639
18. H-Store: A next generation OLTP DBMS. <http://hstore.cs.brown.edu>. Online. Accessed 21 Aug 2017
19. Fang H, Zhang Z, Wang CJ, Daneshmand M, Wang C, Wang H (2015) A survey of big data research. *IEEE Netw* 29(5):6–9. <https://doi.org/10.1109/MNET.2015.7293298>
20. Galante and L. C. E. D. Bona. 2012. A survey on cloud computing elasticity. In *Proceedings of the IEEE/ACM 5th international conference on utility and cloud computing*, pp 263–270
21. Das S, Agrawal D, El Abbadi A (2009) ElasTraS: an elastic transactional data store in the cloud. In *USENIX HotCloud*
22. Pavlo A, Curino C, Zdonik S (2012) Skew-aware automatic database partitioning in shared-nothing, parallel OLTP systems. In: *SIGMOD*
23. Quamar A, Kumar KA, Deshpande A (2013) SWORD: scalable workload-aware data placement for transactional workloads. In: *Proceedings of the 16th international conference on extending database technology*, pp 430–441
24. Schall D, Harder T (2014) Dynamic physiological partitioning on a shared-nothing database cluster. *CoRR*, abs/1407.0120
25. Taft R, Mansour E, Serafini M, Duggan J, Elmore AJ, Aboulnaga A, Pavlo A, Stonebraker M (2014) E-store: fine-grained elastic partitioning for distributed transaction processing systems. *Proc VLDB Endowment* 8:245–256
26. Xu N (2014) Fine-grained data partitioning framework for distributed database systems. In: *Proceedings of the 23rd international conference on world wide web, Seoul, Korea—07 Apr 2014*
27. Tözün P, Pandis I, Johnson R, Ailamaki A (2012) Scalable and dynamically balanced shared-everything OLTP with physiological partitioning. *VLDB J*
28. Kaiji Chen, Yongluan Zhou, Yu Cao, “Online Data Partitioning in Distributed Database Systems,” 18th International Conference on Extending Database Technology (EDBT), March 23–27, 2015, Brussels, Belgium
29. Serafini M, Taft R, Elmore AJ, Pavlo A, Aboulnaga A, Stonebraker M (2016) Clay: fine-grained adaptive partitioning for general database schemas. *Proc VLDB Endowm* 10(4):445–456
30. Asad O, Kemme B (2016) Adaptcache: adaptive data partitioning and migration for distributed object caches. In: *Proceedings of the 17th international middleware conference, ACM*, p 7
31. John V (2017) Redynis: traffic-aware dynamic repartitioning for a distributed key-value store. In: *The 17th international middleware conference*, 24 Mar 2017

# Voice-Based Automation Control Platform for Home Electrical Devices



Constance Izuchukwu Amannah and Promise Nlerum

**Abstract** Voice control automation system is a system that enables the control of appliances with your voice through which an android application and an Arduino mega board is used to control the relay through which an appliance is switched on/off. The aim of the research was to develop a voice support platform, design an interface for a Bluetooth medium and a microcontroller board (Arduino board), activate communication between an android device and the Bluetooth-enabled microcontroller board wirelessly via Bluetooth communication, detect and transfer speech data via Bluetooth to the microcontroller board, prepare a software to run on the microcontroller that will perform the necessary control, and prepare the software to run on the android device in other to get speech data and transferring over Bluetooth. The methodology used in this project was waterfall model. The output design of the proposed system consists of the Bluetooth module, Arduino, and the relays. The Bluetooth module receives the input data and sends it to the Arduino, and then it processes the input command and sends an output signal to a particular relay required for switching. The complete application software was achieved successfully by using Android, C Language, Bluetooth module, microcontroller, and relay.

**Keywords** Voice · Arduino · Bluetooth

---

C. I. Amannah (✉)

Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt P.M.B. 5047, Nigeria  
e-mail: [aftermymisc@yahoo.com](mailto:aftermymisc@yahoo.com)

P. Nlerum

Rivers State University, Nkpolu-Oroworukwo, Port Harcourt, Rivers, Nigeria

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_47](https://doi.org/10.1007/978-981-15-2329-8_47)

# 1 Introduction

## 1.1 *Background and Statement of the Problem*

Technologists and scientists are everyday looking for ways to minimize the stress and risk faced in our everyday life. Scientists seek to improve life by taking it from manually controlled equipment's and appliances which could fault and cause accidents in the home to automatically controlled system which would prevent most of these accidents from occurring as one could be so stressed out from work. Usually, conventional wall switches are located in different corners of the home and, and thus, we are faced with the need of locating and manually operating these controls; the physically challenged finds it difficult to access the light and fan switches especially when they are the only ones at home and could not be assisted by available neighbors. Another group of people that may be the elderly, like the physically challenged, also suffer from the ordeal of old age, thereby inhibiting their access to electrical provisions in the room. To address the issue, we interface a mobile device (android phone) with a control unit (Arduino board) to communicate wirelessly. This will bring about the ease of control as well as save one from stress as what the user will be doing is simply to operate through his phone which is not a difficult task. It also prevent direct contact between the user and the electrical appliances and by so doing bring to zero percent the risk of electric shock.

## 1.2 *Purpose and Specific Objectives*

The aim of the study was to develop a voice-based automation control platform for home electrical devices. The study was designed with the following objectives:

- i. Design an interface for a Bluetooth medium and a microcontroller board (Arduino board).
- ii. Activate communication between an android device and the Bluetooth-enabled microcontroller board wirelessly via Bluetooth communication.
- iii. Detect and transfer speech data via Bluetooth to the microcontroller board.
- iv. Prepare a software to run on the microcontroller that will perform the necessary control.
- v. Prepare the software to run on the android device in other to get speech data and transferring over Bluetooth.

### ***1.3 Significance of the Study***

The developed system will be of immense benefit to the elderly and the physically challenged. In the event of sickness and tiredness from daily tasks, the physically fit persons will also benefit from the system, as they can conveniently access any electrical devices from any location of their room sometimes from the comfort of their beds.

### ***1.4 Scope of the Study***

The study is limited to the use of the voice recognition system together with the Bluetooth functionality contained in an android device in order to turn on and off household power points.

## **2 Literature Review**

### ***2.1 Actor–Network Theory***

The theory of actor–network was developed by Latour [4]. The theory explains a link to voice control home automation system because different microprocessors like chips will be designed to interact with the computer programs in order to achieve the aim and objectives of the study. In Latour’s view, networks are formed by the combination of the actor and the networks. This combination makes for the successes achievable by both the actor and the network. Hence, no single entity can account for the advancement of the relationship between the actor and the network except the duo.

### ***2.2 Application of Voice Control and Automation Devices***

**Vocal light:** This feature does not require any installation, no setup, and no Wi-Fi is required. All you have to do is simply screw any bulb at home into VOCCA adapter, connect that into any light socket, and say the magic word—“go VOCCA light” [3].

**Jibo:** This is an application-based JavaScript platform that determines the robot’s features for voice recognition, touch screen, and camera work [6]. Jibo is well prepared for its task and includes high-resolution cameras, built-in speakers; modules for Bluetooth and Wi-Fi LCD touch screen, touch sensors, microphones, and processor.

### ***2.3 Principles of Voice Control Systems***

The principles of voice control systems Meng et al. [5] are highlighted in as follows:

- i. Speech signal
- ii. Feature selection
- iii. Feature extraction
- iv. Acoustic models
- v. Language models
- vi. Modeling.

### ***2.4 Voice-Based System for Home Communication***

Anand et al. [2] did a work on voice recognition. Their work focused on the following as their objectives:

- i. Training the Voice Recognition Kit.
- ii. Coding the LCD to display messages on it using CUBE SUITE+.
- iii. Interfacing and monitoring speed and direction control of DC motor using L293D (DC motor driver).
- iv. Configuring ZigBee for both transmitter and receiver section.
- v. Programming the Renesas microcontroller and dumping the codes using Flash Magic tool.
- vi. Assembling all the parts together and testing.

Their work was based on the following methodology: The design specifications for the hardware components (microcontroller, LCD, DC motor, etc.) and software tool (Cube Suite + IDE) with performance requirements and reliability. Their study achieved a proposed system which is combined with the software and hardware manipulations resulting in two sections—transmitter section and receiver section which give the final model for the home automation system. Their work did not focus on the activation of communication between an android device and the Bluetooth-enabled microcontroller board wirelessly via Bluetooth communication. The current study is therefore poised to close this gap which the work of Anand et al. [2] could not address.

### ***2.5 Voice-Controlled Smart Home***

In the work, Amrutha et al. [1] did not focus on preparing the software to run on the android device in other to get speech data and transferring via Bluetooth. The proposed system of the current study is designed to address this gap in knowledge



which was left open in the work of Amrutha et al. [1]. In the work of Mathew, Pranav et al. [7], it is evident that their work did not address the detection and transferring of speech data via Bluetooth to the microcontroller board. The current research identifies this obvious gap and is poised to create a mechanism for the detection and transfer of speech data via Bluetooth across a microcontroller board, especially at homes.

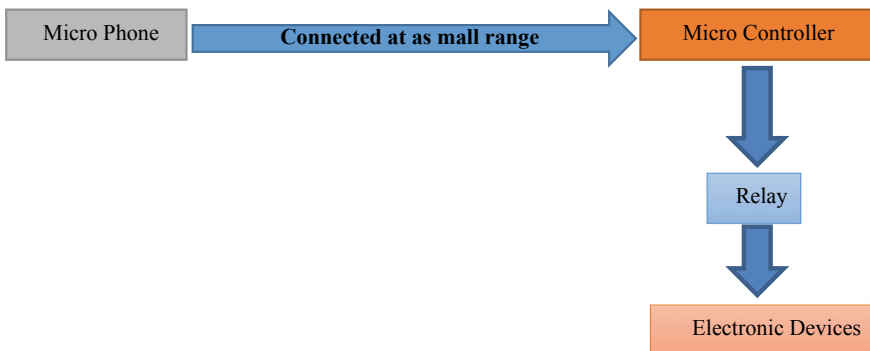
### 3 System Analysis and Design

#### 3.1 Analysis of Existing System

The existing system is voice recognition system for home automation and the system is made up of a V3 voice recognition module, an Arduino board, and relay modules. For the system to work, the voice recognition module must be trained first before it can recognize voice commands (Fig. 1).

#### 3.2 Constraints of Existing Systems

The existing system though designed to work wirelessly has some constraints with portability, mobility, and accessibility; as they require some sort of the presence of the operator at the speech recognition unit as well as the immobility of the speech recognition section of the individual systems. This has made it that there must be extra hardware designed and used alongside the entire system.



**Fig. 1** Block diagram of existing system

### ***3.3 Analysis of the Proposed System***

The proposed system is voice-controlled, and it consists of a microcontroller board, a Bluetooth module, and an android phone acting as the communication channel between the user and the device. The system is built based on wireless communication between the user and the device. When the user sends an audio (voice) command through the android phone, it is then received by the microcontroller wireless via the Bluetooth module, and the microcontroller further decrypts the commands to a form of code, processes this coded information, and based on the information triggers the output signal through its output pins to the relays connected to its pins. The relay(s) then does a switching action. The proposed system enables the interface for a Bluetooth medium and a microcontroller board and also enables activation of communication between a user of an android device and a Bluetooth-enabled microcontroller board wirelessly via Bluetooth communication and also detects and transfer speech data via Bluetooth to the microcontroller board.

### ***3.4 Justification of the Proposed System***

The proposed system is necessary in view of the following reasons:

- i. Communication is made between the user and the microcontroller system unit with the aid of an android device.
- ii. Communication identification is done wirelessly using the Bluetooth communication protocol.
- iii. The input control unit is mobile just as the android phone is mobile.
- iv. There is no need for extra accessories as the only thing the user needs to implement a control is his/her android device.

### ***3.5 Method of the Proposed System***

The waterfall model was used in this work. The choice of the model was because it is procedural nature and implementation style. It allows for step-by-step algorithm during the design and also during simulation. The model supports effectively the voice-controlled system of this study.

### ***3.6 Design of the System***

This section explains how the hardware constituents of the system are interfaced together for the actualization of the aim.

### 3.7 Architecture of the Proposed System

The system consists of basically six blocks: the relay board which does the switching action, the Arduino—which act as the controlling/processing device, the Bluetooth block which receives the commands wirelessly via radio wave at 2.4 GHz frequency and sends it to the connected microcontroller (Arduino), the open-air block which act as a medium for radio wave communication, the android phone which has Bluetooth application installed for interfacing with the voice command in form of sound waves which is converted to electrical energy by the android phone’s microphone. All these blocks put together to make the actual definition of the system (Fig. 2).

**Input Design:** This section of the design describes the processes through which the command gets to the microcontroller. The speech command is the desired command made by the user maybe “switch to inverter”; this is in row form as sound wave; this row sound wave is then converted to needed electrical wave by the android phones and then through the phone’s Bluetooth application the command in electrical waveform is further sent to via radio waves using air as a medium to the Bluetooth; the Bluetooth module then decrypts the information/command received and finally sends the command as input to the main controller (microcontroller) (Fig. 3).

**Process Design:** The main processing device for the proposed system is the micro-controller, and it handles processing of the command sent by the Bluetooth module and depending on the command received it sends an output signal to trigger a particular relay switching action, the microcontroller being Arduino UNO (Arduino one) has (Fig. 4).

**Output Design:** The Bluetooth module receives the input data and sends it to the Arduino, and then it processes the input command sends an output signal to a particular relay required for switching (Fig. 5).

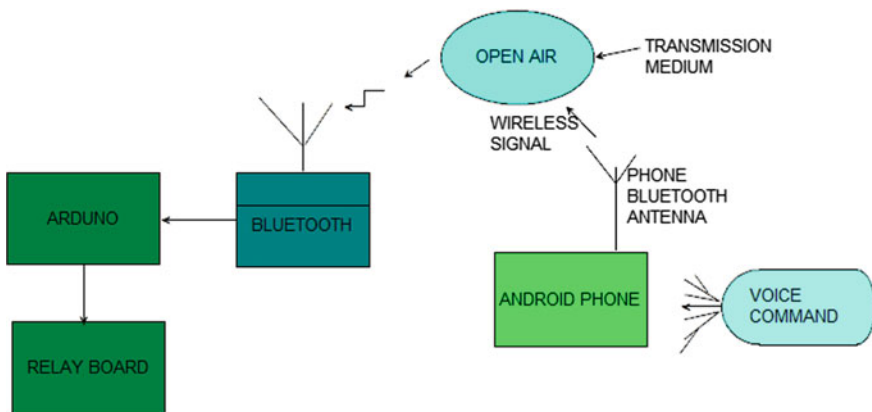


Fig. 2 Architecture of the system



Fig. 3 Input design of the proposed system

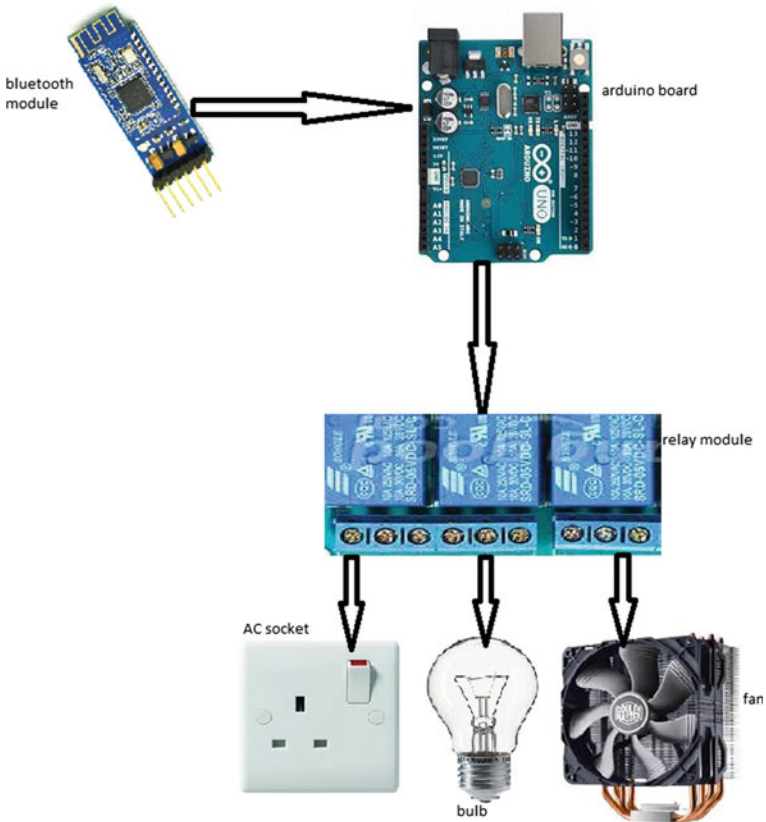


Fig. 4 Process design of the proposed system

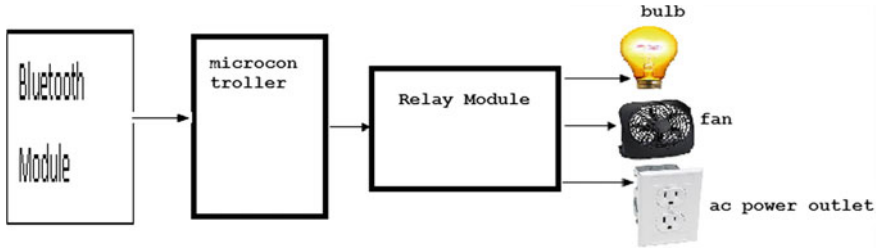


Fig. 5 Output design

Table 1 Database table

Variable	Data type	Size	Description
Turn on the light	Logical	32 bits	Switch on the light
Turn off the light	Logical	32 bits	Switch of the light
Turn on the fan	Logical	32 bits	Switch on the fan
Turn off the fan	Logical	32 bits	Switch off the fan
Turn on the power outlet	Logical	32 bits	Switch on the power outlet
Turn off the power outlet	Logical	32 bits	Switch off the power outlet

**Database Design:** This section describes the various ways in which data and data instructions can be stored. It highlights the logical and physical designs of the data elements (Table 1).

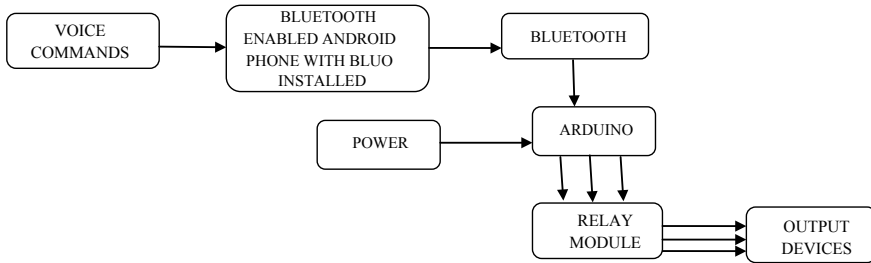
The security checks are listed below:

- i. If you do not have the application on your android phone you cannot access it.
- ii. If you do not know the address of the Bluetooth module you cannot access it.
- iii. If you do not have access to the database you cannot access it.
- iv. The connectivity can be reduced in the sense that only one or two users can connect to it.

## 4 Implementation Architecture

The implementation architecture runs through every constituent of the system from the android application to the final output devices. The diagram below shows the implementation architecture (Fig. 6).

Upon system turn on, the user is expected to connect his/her phone to the system via Bluetooth; once the Bluetooth connection is active, the LED on the Bluetooth module stops blinking. To send a command, the user should touch the microphone icon on the Blu app interface. The user is expected to say the command once the



**Fig. 6** Implementation architecture

“speak now” write up appears on the screen. The spoken command is then transferred to the microcontroller unit for processing and control. The microcontroller receives the command over Bluetooth and check what it is for and acts accordingly, i.e., turn on or off the light, the same goes for the fan and AC power outlet.

**User Manual:** In order to operate the system, the following procedure should be followed:

- i. Switch on the android device.
- ii. Locate Bluo android application and tap it, a new Android activity pops up, on the top right-hand side of your android screen, tap the search, button, a dialogue box pops up requesting the user to turn on the Bluetooth functionality in your phone (if the Bluo android application is not in your device go to Google play store and download it).
- iii. After turning on your Bluetooth functionality in your android device, IP addresses of available Bluetooth modules pop up.
- iv. Tap the IP address of your Bluetooth module, after then it will show connected.
- v. Finally make the voice command through your phone to achieve your desired result.

## 5 Summary and Conclusion

### 5.1 Summary

The aim of the study was achieved. The following objectives were designed to aid the achievement of the aim, design an interface for a Bluetooth medium and a microcontroller board (Arduino board), activate communication between an android device and the Bluetooth enabled microcontroller board wirelessly via Bluetooth communication, detect and transfer speech data via Bluetooth to the microcontroller board, and prepare a software to run on the microcontroller that will perform the necessary control. The complete application software was achieved successfully by using android,

C Language, Bluetooth module, microcontroller, and relay. The waterfall model was adopted as the methodology employed to actualize the systematic working of the system.

## 5.2 Conclusion

The study achieved its aim and objectives. It realized a voice-controlled home automation using the Bluetooth technology. The distance variation requisite for the full implementation of the voice-controlled system is 10 m. The coverage area can be increased using GSM modules.

## 6 Recommendations

For effective and efficient improvement of the system's output, the following needs to be done.

- i. Add LEDs for error checking if the right signal comes into the system.
- ii. Add moving components like motors to enable the system control doors.
- iii. Add a real-time clock module and SD card module for timekeeping of event occurrence.

## References

1. Amrutha S, Aravind S, Ansu M, Sugathan S, Rajasree R, Priyalakshmi S (2015) Voice controlled smart home. *Int J Emerg Technol Adv Eng* 5(1):272–273
2. Anand NL, Kumr BY, Gupta MK, Sauray FH (2016) Voice controlled home automation. *Int J Adv Res Comput Sci Softw Eng* 6(3):161–162
3. Gadgetreview S (2005) Voice controlled smart home. *Int J Emerg Technol Adv Eng* 5(1):34–56
4. Latour B (2005) Actor-network theory, CSI-Paris/science student-San. *J Sci Theor* 25(3):47–64
5. Meng J, Zhang J, Zhao H (2012) Overview of the speech recognition technology. *Int Conf Comput Inf Sci* 4(1):58–60
6. Park L (2008) What we talk about when we talk about context. *Pers Ubiquit Comput* 8(1):19–30
7. Pranav B, Vedant R, Aditya R (2016) Voice controlled home automation. *Int Res J Eng Technol* 3(4):2717–2718

# Modeling and Simulation of Inertial Navigation System



Madhavi Vedpathak, Prachi Mukherji and Balkrishna Prasad

**Abstract** There are many systems used to find out the location of the object or vehicle. The most widely used location tracking system is GPS that is the global positioning system. To find out the location of the missile, the inertial navigation system (INS) is used. Inertial measurement unit (IMU) performs the main role in this system, which consists of microelectromechanical system (MEMS) sensors. Accelerometer and gyroscope are used to give linear acceleration and angular rotation. Integrate the rates obtain from accelerometer and gyroscope twice to get velocity and position. To obtain the exact position of the missile, it is necessary to reduce the Coriolis effect from the rates. This survey paper elaborates on the modeling and simulation of the INS (Brown in Test results of a GPS/inertial navigation system using a low cost MEMS IMU, 04 2004 [1]).

**Keywords** GPS · INS · MEMS · IMU

## 1 Introduction

Similar to GPS, INS is also used to find out the position of the vehicle. GPS works on the principle of sending and receiving the signal from the GPS receiver, whereas INS works by using MEMS sensors, which are rigidly mounted on the object. Hence, it is fast as compared to GPS. But because of some sensor errors, it is less accurate than GPS. To overcome this disadvantage of INS, make the fusion of GPS and INS [4].

Inertial measurement unit (IMU) is rigidly mounted on a missile or vehicle which consists of accelerometer and gyroscope to get linear acceleration and angular

---

M. Vedpathak (✉) · P. Mukherji · B. Prasad  
MKSSS's Cummins College of Engineering for Women, Pune, Maharashtra, India  
e-mail: [madhavi.vedpathak@cumminscollege.in](mailto:madhavi.vedpathak@cumminscollege.in)

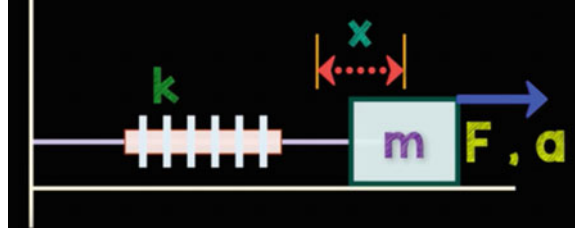
P. Mukherji  
e-mail: [prachi.mukherji@cumminscollege.in](mailto:prachi.mukherji@cumminscollege.in)

B. Prasad  
e-mail: [bkpnit1942k2@gmail.com](mailto:bkpnit1942k2@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_48](https://doi.org/10.1007/978-981-15-2329-8_48)



**Fig. 1** Accelerometer



velocity. Integrate translational and rotational rates twice to obtained velocity and position.

We know that [2],

$$\int \text{Acceleration} = \text{Velocity} \tag{1}$$

$$\int \text{Velocity} = \text{Position} \tag{2}$$

**A. Accelerometer**

Acceleration force is measured by an electromechanical device called accelerometer. There are many types of sensors such as MEMS, optical, mechanical. Accelerometer used in INS is MEMS sensors because it is very simple reliable and cost-effective. Here three accelerometers are used in *x*-, *y*-, *z*- directions to obtain the linear acceleration in all three directions.

Capacitive-type MEMS accelerometer is used widely to calculate the linear acceleration. The high sensitivity and the accuracy at high temperature of the capacitive MEMS accelerometer make it different. The device changes its value only when capacitance changes which depends on the distance between capacitive plates. From Fig. 1, we will see the working of the accelerometer [7].

Using the spring and mass system, we will understand the working of the accelerometer. In this Fig. 1, a body of mass '*m*' is attached to the wall having spring with spring coefficient '*k*'. When force '*F*' is applied to mass *m*, it is displaced with '*x*' in the direction of applied force with an acceleration '*a*'.

According to Newton law, we know that

$$F = ma = Fs \tag{3}$$

*F<sub>s</sub>* is nothing but force due to spring tension, where

$$F_s = kx \tag{4}$$

By equating the above equation, we get

$$ma = kx \tag{5}$$

Thus, we see that acceleration is the function of displacement  $a = f(x)$ . Thus by measuring the displacement, the acceleration of the body can be measured. To measure the displacement  $x$ , different methods are used, such as resistive, capacitive, and inductive techniques. Here we will see the capacitive technique.

From basic physics

$$C = f(A/d) \tag{6}$$

where  $d$  = distance between capacitor plates. As distance ' $d$ ' increases capacitance ' $C$ ' decreases. This capacitance can be measured electronically by a signal conditioning system. So by knowing the capacitance, we can obtain the acceleration. From two capacitor plates, one is attached to the movable body  $m$  and another is fixed to the initial position [7].

### B. Gyroscope

The angular velocity and the orientation of the object can measure by using gyroscope. The working of the gyroscope is based on the Coriolis force. Consider the body of mass ' $m$ ' as shown in Fig. 2, and it is moving in  $x$ -direction with velocity ' $v_x$ '. Suppose angular velocity ' $w$ ' acts on it in  $y$ -direction. The body of mass  $m$  experiences Coriolis force in  $z$ -direction. The magnitude of Coriolis force is directly proportional to the body of mass  $m$ , velocity  $v_x$ , and angular velocity  $w$ . The direction of the Coriolis force can find out using right-hand rule [6].

From this, by measuring the Coriolis force we can find out the angular velocity. Gyroscope uses two mass system keeps moving in opposite direction continuously

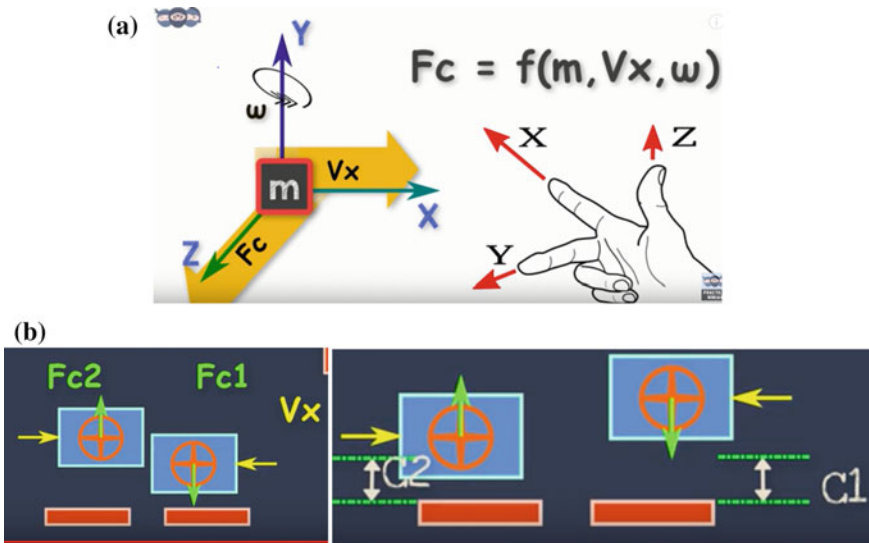


Fig. 2 a Gyroscope, b Axis representation for force experienced

having the same magnitude of velocity. When the whole system experiences angular rotation, mass experiences Coriolis force, they have the opposite direction as per the right-hand rule. As the angular rotation applied on the mass Coriolis force  $F_{c1}$  and  $F_{c2}$  acts, and they move in the opposite direction that is upward and downward. So as the distance changes the capacitance changes [6].

By adding these two forces, we get total Coriolis force. As they are moving in upward and downward direction, their capacitance changes as the distance ' $d$ ' changes,  $C1 = F(d1)$  and  $C2 = F(d2)$ ,  $D = d1 - d2$

$$F_c = f(C) \quad (7)$$

where  $C = c_1 - c_2$

Calculate the angular rotation by using

$$W = f(F_c) \quad (8)$$

### C. Frames

When missile is navigating in space, the MEMS sensors take all the measurements in the body frame [8]. To analyze it and to perform operations on it, we have to convert it from body frame to north–east–down (NED) frame and earth-centered earth fixed (ECEF) frame.

**Body Frame** In an inertial navigation system, all the measurements of linear acceleration and angular rotation are taken into the body frame.

**NED Frame** NED frame has its origin at the location of the navigation system. The axes of the NED frame are aligned with north, east, and vertical down.

**ECEF Frame** Earth-centered earth fixed frame is a geographic and Cartesian coordinate frame, and center of mass of the earth is the origin of this frame.

The center of the mass of earth is origin (0, 0, 0) of the frame.

Through the Americas at lat long (0,  $-90$ ), positive  $x$ -axis passes.

Through the South–North pole, positive  $y$ -axis passes.

Through the prime meridian (0 longitudes) at lat long (0, 0), positive  $z$ -axis passes.

### D. DCM

DCM is the direction cosine matrix [3] used to perform the frame conversion from one frame to another frame.

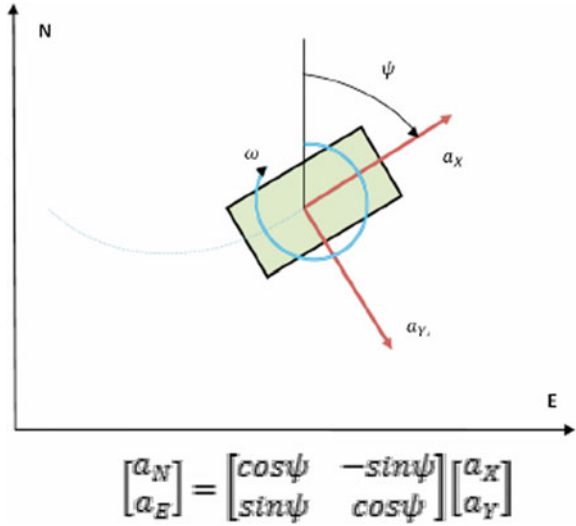
Figure 3 shows the direction cosine matrix for 2D.

$a_N, a_E$  = acceleration in NED frame

$a_X, a_Y$  = acceleration in body frame

$W$  = angular velocity

Fig. 3 DCM for 2D



Consider the body is moving in  $x$ - and  $z$ -directions with acceleration 2 and 1 ( $m/s^2$ ), respectively, and rotating along  $y$ -axis with 0.1(deg/s). Now we have to convert these measurements from the body frame to earth frame. This is 3 degree of freedom because body moving along  $x$ -,  $z$ -directions and rotating along  $y$ -direction. By using the above matrix, we can perform this conversion. Figure 4 shows the conversion from the body to earth frame [2].

**E. Gravity Compensation** (Fig. 5)

Suppose the missile is moving in space with height 10 km above the earth. Now we have to land this missile on earth. For that we have to compensate gravity on the missile, i.e., on IMU. IMU does not feel any gravity because it is rigidly mounted on missile so we are compensating the gravity on it.

But the problem is where to compensate the gravity in body frame or earth frame.

An equation to calculate gravity at any height is

$$g(h) = g(\text{sea level}) + \frac{dg(0)}{dh} * h(m) \tag{9}$$

whereas

$$g(\text{sea level}) = 9.780318 [1 + (5.3024 \times 10^{-3} \times \sin^2 \times L) - (5.9 \times 10^{-6} \times \sin^2 \times 2L)] m/s^2 \tag{10}$$

$$\frac{dg(0)}{dh} = -3.0877 \times 10^{-6} [1 - (1.39 \times 10^{-3} \times \sin^2 \times L)] \frac{m/s^2}{m} \tag{11}$$

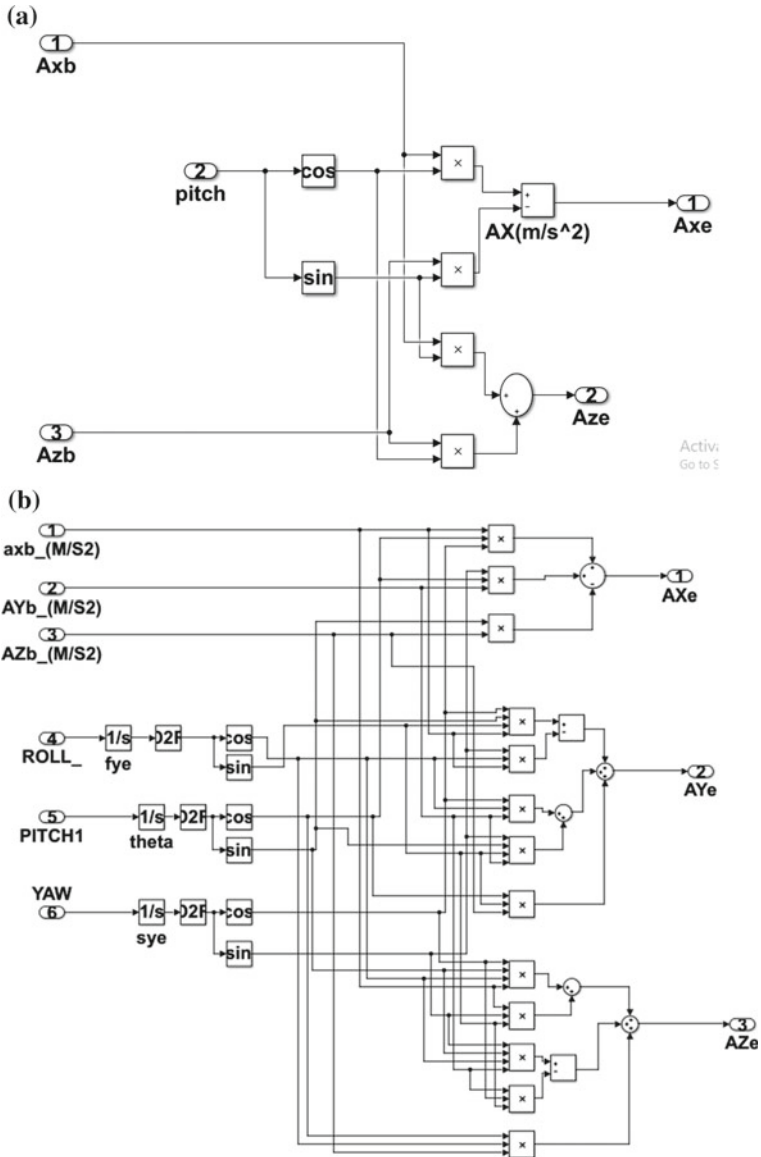


Fig. 4 a DCM 2D, b DCM 3D

If we want to compensate it in body frame, then compensate on all three axes because there is no gravitational force in space but if we compensate it in earth frame, then compensate only on z-axis because we know that in earth frame z-axis is directed downward [2].

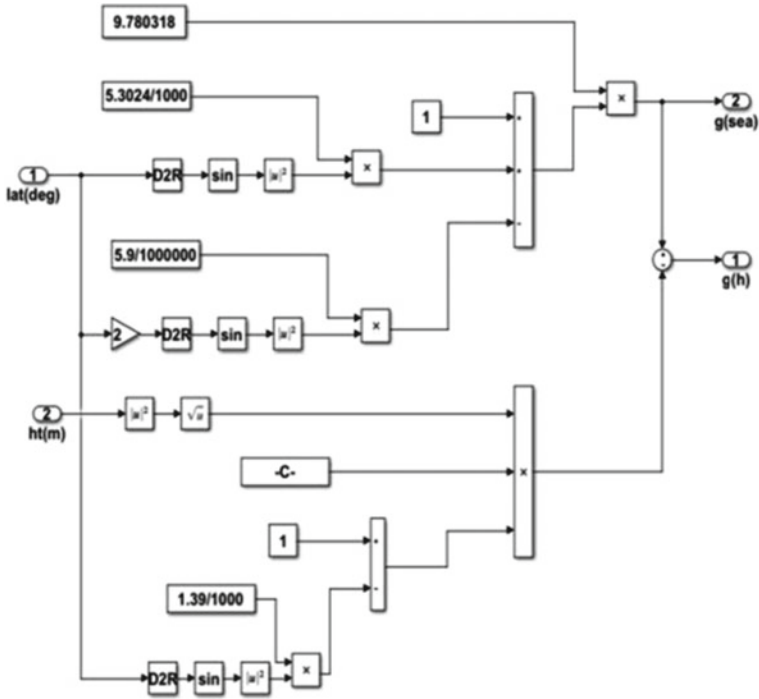


Fig. 5 Gravity compensation

**A. Coriolis Effect**

Coriolis force always acts on the object that is in motion.

The Coriolis force is acting perpendicular to the direction of motion and the axis of rotation. On the earth, the deflection of the object occurs due to this force. In the northern hemisphere, the deflection is toward right side and in the southern hemisphere the deflection is toward left side. Equations 12 and 13 are used to calculate latitude rate and longitude rate, and the values of  $M_E$  and  $N_E$  are taken from WGS 84 model [5].

$$\lambda' = \frac{u}{M_E + h} \tag{12}$$

$$u' = \frac{v}{(N_E + h) \cos \lambda} \tag{13}$$

The navigation equation for translational motion will be,

$$u' = -(u' + 2W_E) \sin(\lambda)V + \lambda' * W \tag{14}$$

$$v' = (u' + 2W_E) \sin(\lambda)U + (u' + 2W_E) \cos(\lambda)W \tag{15}$$

$$w' = -(u' + 2W_E) \cos(\lambda)V - \lambda' * U \tag{16}$$

$U, V, W$  NED velocity  
 $W_E$  earth rotational velocity in rad/s  
 $u$  longitude  
 $\lambda$  latitude

Navigation equation for rotational motion is

$$\begin{matrix} P & (W_E + u') \cos \lambda \\ Q = T_E^B * & -\lambda' \\ R & -(W_E + u') \sin \lambda \end{matrix} \tag{17}$$

## 2 Result

The data sensed from an accelerometer and gyroscope is given to the module to obtain the exact position of the object (Fig. 6).

All the data of the accelerometer and gyroscope are given to the INS module to obtain the exact position of the object.

The final result that is the position of missile along  $x$ -,  $y$ -,  $z$ -directions is given as (Figs. 7, 8, and 9).

t	ax	ay	az	p	q	r	psi	phi	theta
0.006	-0.70833	-4E-06	0.129986	-0.00396	0.147391	0.001322	0.000008	-2.4E-05	0.000378
0.026	-0.70764	-1.5E-05	0.106637	-0.00393	0.610674	0.001313	0.000034	-0.0001	0.00778
0.046	-0.70685	-2.7E-05	0.079127	-0.00391	1.040078	0.001315	0.00006	-0.00018	0.024133
0.066	-0.70596	0.000001	0.047591	-0.0041	1.43166	0.001231	0.000087	-0.00026	0.04872
0.086	-0.70503	0.000237	0.011013	-0.00429	1.779622	0.000866	0.000109	-0.00034	0.080739
0.106	-0.70394	0.000752	-0.03057	-0.00427	2.074452	-0.00078	0.000114	-0.00042	0.119228
0.126	-0.70299	0.002163	-0.07693	-0.00531	2.305951	-0.00516	0.000063	-0.00052	0.163029
0.146	-0.70185	0.004197	-0.12667	-0.00934	2.466017	-0.0157	-0.00013	-0.00066	0.21079
0.166	-0.70072	0.00526	-0.18016	-0.01399	2.54731	-0.03241	-0.0006	-0.00089	0.261019
0.186	-0.69953	0.007366	-0.24086	-0.01947	2.535663	-0.05414	-0.00144	-0.00123	0.312026
0.206	-0.69813	0.009745	-0.3024	-0.02278	2.411004	-0.08521	-0.0028	-0.00167	0.36174
0.226	-0.69714	0.009595	-0.35258	-0.02394	2.18605	-0.12316	-0.00487	-0.00215	0.407954
0.246	-0.69639	0.007086	-0.39057	-0.02639	1.902505	-0.15908	-0.00768	-0.00267	0.449047
0.266	-0.69555	0.004442	-0.4233	-0.03011	1.589079	-0.18884	-0.01116	-0.00326	0.484161
0.286	-0.69463	0.001481	-0.45439	-0.03123	1.249531	-0.21268	-0.01517	-0.00392	0.512762
0.306	-0.69414	-0.00102	-0.48317	-0.02847	0.882703	-0.2314	-0.01961	-0.00456	0.534312
0.326	-0.69337	-0.00413	-0.50719	-0.02495	0.492532	-0.24594	-0.02439	-0.00513	0.548291
0.346	-0.69268	-0.00798	-0.52598	-0.02326	0.087781	-0.25344	-0.02939	-0.00566	0.554312
0.366	-0.69228	-0.01243	-0.53952	-0.02221	-0.32231	-0.25239	-0.03447	-0.00617	0.552172
0.386	-0.69188	-0.01691	-0.54812	-0.02009	-0.72815	-0.24135	-0.03943	-0.00664	0.541856
0.406	-0.69142	-0.0212	-0.55303	-0.01876	-1.12399	-0.22053	-0.04407	-0.00707	0.523512
0.426	-0.69138	-0.02418	-0.55572	-0.0204	-1.5085	-0.19153	-0.04821	-0.00749	0.49736

Fig. 6 Data from IMU

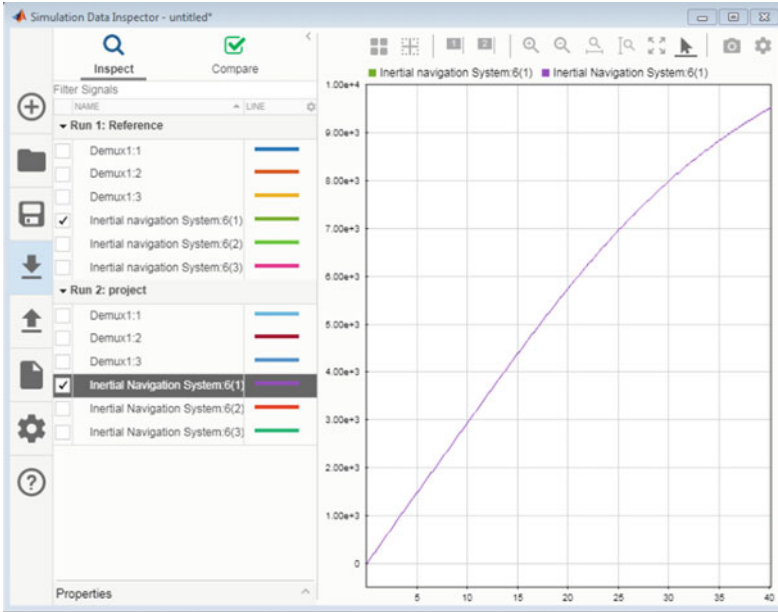


Fig. 7 X-direction

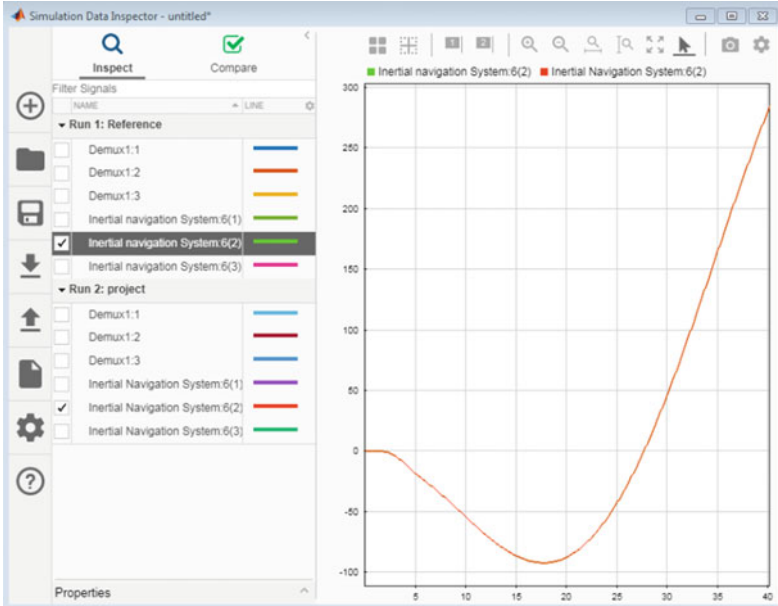


Fig. 8 Y-direction



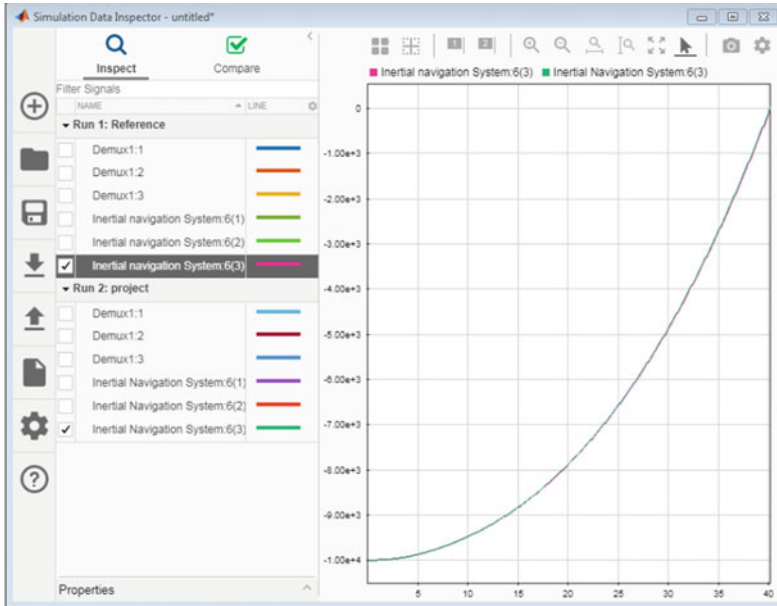


Fig. 9 Z-direction

### 3 Conclusion

In this paper, it is observed that the inertial navigation system (INS) is used to find out the location of the object using microelectromechanical system (MEMS) sensors. To find out the velocity and position, integrate twice the accelerometer and gyroscope rates obtain from inertial measurement unit (IMU).

The earth is moving with a specific speed, because of its rotation Coriolis effect occurs. The position of the object can be changed because of the Coriolis effect. To reduce these Coriolis effects, navigation equations are used.

### References

1. Brown A (2004) Test results of a GPS/inertial navigation system using a low cost MEMS IMU, 04 2004
2. Feng LC, Li SH, Liu Y, Fu QW (2018) Advantages of ECEF-frame based global navigation mechanization for INS/GNSS integrated navigation system, 05 2018
3. Qin F, Chang L, Jiang S (2018) Inertial navigation system calibration based on direction cosine matrix matching. In: 2018 IEEE international conference of intelligent robotic and control engineering (IRCE, Aug 2018)
4. Tan TD, Ha L, Thang N, Duc N, Nguyen T (2007) Integration of inertial navigation system and global positioning system: performance analysis and measurements, 12 2007

5. Vau B, Ponceau D, Bussutil M (2018) A control strategy for coriolis and centrifugal effects reduction in an inertial system test equipment, 03 2018
6. Zhang Q, Niu X (2018) Research on accuracy enhancement of low-cost MEMS INS/GNSS integration for land vehicle navigation, 04 2018
7. Zhou J, Edwan E, Knedlik S, Loffeld O (2010) Low-cost INS/GPS with nonlinear filtering methods, 08 2010
8. Zihajehzadeh S, Lee TJ, Lee JK, Hoskinson R, Park EJ (2015) Integration of MEMS inertial and pressure sensors for vertical trajectory determination. *IEEE Trans Instrum Measur* 03 2015

# Dual Watermarking for Colour Images' Copyright and Authentication Using DWT Technique



Rajesh Thakare, Sandeep Kakde and Prashant Mani

**Abstract** The watermarking technique is a very vital method to identify the originality of the object. This paper describes the dual watermarking method for the digital images which is invisible. In this technique, the watermark is added in host image, and specifically, the dual watermarks provide more security compared with the single one. The robust and fragile watermarks are used in this scheme for image authenticity and the copyright protection purpose. Firstly, the colour image is converted into YCbCr space separated into each space; the Y colour space is used as the embedding watermark section. The DWT is applied to the Y space from which the LL section is embedded with the watermark. The improved values of PSNR and MSE can define the quality of image.

**Keywords** DWT · Watermarking · PSNR · MSE · MATLAB

## 1 Introduction

In today's era, protecting digital information is an important subject to be discussed. To fix this, digital watermarking arises so as to protect the information. The watermarks are bounded with the contents such as the video, image and the text documents to prevent piracy and duplication. The watermarking scheme is utilized so as to yield the image authentication and copyright protection for the digital images, to prove the authenticity of the image. The user can extract the watermark maintaining the integrity of the image, thus the image can be recovered easily.

---

R. Thakare · S. Kakde (✉)

Electronics Engineering Department, Y C College of Engineering, Nagpur University, Nagpur, India

e-mail: [sandip.kakde@gmail.com](mailto:sandip.kakde@gmail.com)

R. Thakare

e-mail: [rdt2909@gmail.com](mailto:rdt2909@gmail.com)

P. Mani

Electronics and Communication Engineering Department, SRM University, Ghaziabad, India

e-mail: [prashanm@srmist.edu.in](mailto:prashanm@srmist.edu.in)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication*

*Engineering*, Lecture Notes in Networks and Systems 106,

[https://doi.org/10.1007/978-981-15-2329-8\\_49](https://doi.org/10.1007/978-981-15-2329-8_49)

## 2 Related Work

Choudhary et al. focuses on the technique to embed the binary data or the watermark in the host image, which consists of DWT method for watermarking the image; the user can extract the watermark maintaining the integrity of the image. Thus, the image can be recovered easily using 2-level. Also, the invisibility of the watermark is produced. The paper implements the digital watermarking algorithm using combination of dwt and SVD 2D is applied to decompose the host image in 4 sub- band focusing on watermarked image the attacks like blurring, contrast adjustment, histogram [1] equalization etc. the original watermark which is inserted is extracted from all bands. After the extraction the MSE and PSNR values are compared this technique makes the watermarked image remains unaffected by attacks and can also be recovered from the sub-bands efficiently. Furqan et al. present the digital watermarking method which is invisible, and it uses the singular value decomposition [2]. DWT-based method on dual watermarking for shading images is proposed which greatly uses for cryptographic purpose applications [3]. Watermark is inserted in red shading component, and DWT scheme is used in paper [4, 5]. DCT is applied on  $8 \times 8$  square which shows low-recurrence sub-band by utilizing discharge sort to enhance the installing power [6, 7].

## 3 Proposed Methodology

The proposed method involves the embedding of dual invisible watermarking and extraction at the output side. Figure 1 shows the block diagram of the watermarking scheme. The input to the system should be a digital colour image, and the output of the system is the colour image with the embedded watermark. Firstly, the colour image is converted into YCbCr colour space, and then it is split into individual colour space Y, Cb, Cr. Y space is separated which is further used to embed the watermark.

### A. Embedding the Watermark

Single-level DWT is implemented on the cover image that splits the image into four sub-bands: the frequency approximation, elevated frequency diagonal, low-frequency horizontal and low-frequency vertical sub-bands. It is entrenched into cover image (Fig. 2).

The product of a specific scaling factor with decomposed components of cover image and watermark is done and then it is added. While the embedding process is carrying out, the cover image size is greater than the size of watermark.

### B. Extraction of Watermark

See Figs. 3 and 4.

Steps implemented in embedding process are implemented in a descending pattern in the extraction procedure. The process initially involves the conversion of the colour

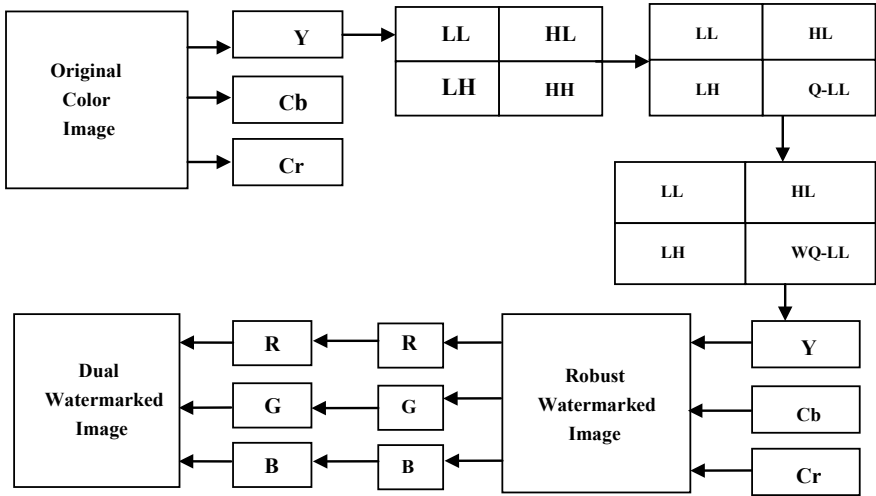


Fig. 1 Proposed method

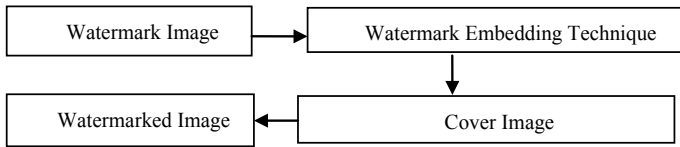


Fig. 2 Watermarking embedding process

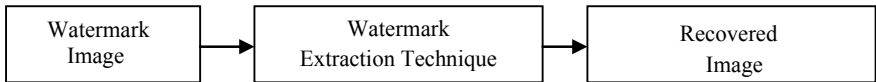


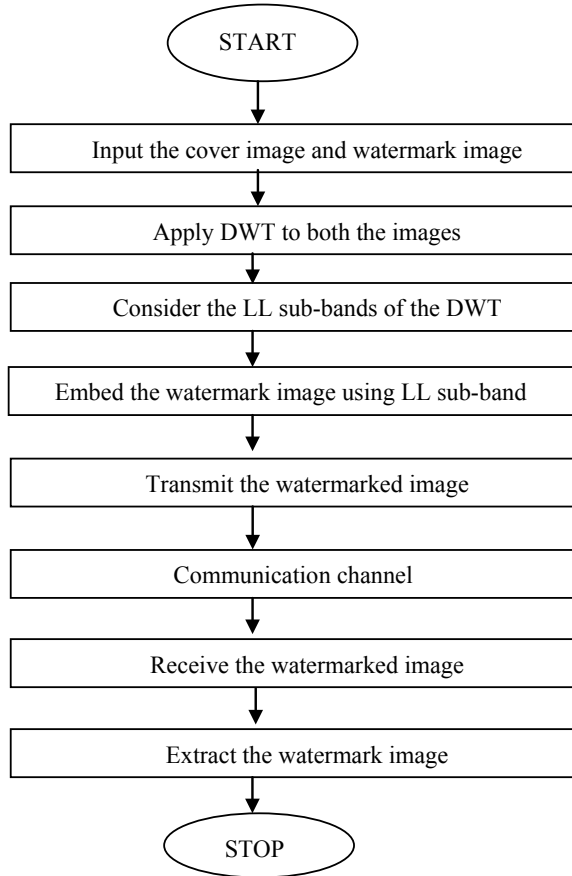
Fig. 3 Watermarking extraction process

image into YCbCr colour space, and then it is split into individual Colour space Y, Cb, Cr. Y space is separated which is further used to entrench the watermark into it. After the separation, the DWT is applied for the decomposition of the original images into four sub-bands.

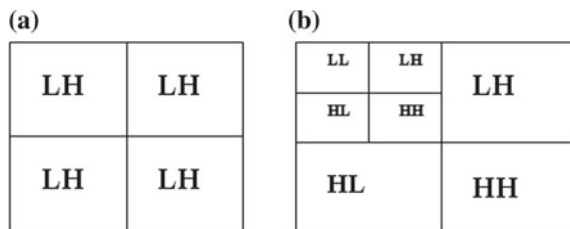
**DWT Algorithm**

DWT domain helps for wavelet transformation and it is performed in the pattern and discrete sample, all this procedure is performed in numerical and functional evaluation (Figs. 5 and 6).

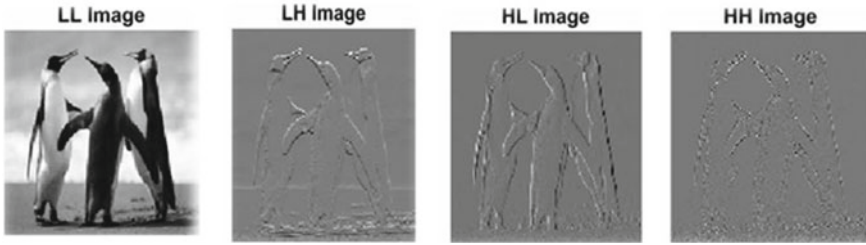
**Fig. 4** Flowchart of watermark methodology



**Fig. 5** a 1-level DWT and b 2-level DWT



The need for using DWT is to divide the frequency parameters resembles the decomposition of an image. In this system, the image is first of all deteriorated into four parts. The DWT is applied on the Y space. To define the quality of the image which is recovered, we used the parameters namely PSNR and MSE. The discrete wavelet in the Y channel of the YCbCr shading space is utilized for vigorous



**Fig. 6** DWT of input image

watermarking. After one dimension of DWT deterioration, the low-low (LL) sub-band of Y is quantized by the luminance quantization slab. The vigorous watermark is inserted in the high-high (HH) sub-band by expertly supplanting it with the after effect of the LL quantization. The delicate watermark for picture verification is inserted freely on each RGB shading channel as per an enhanced LSB substitution approach.

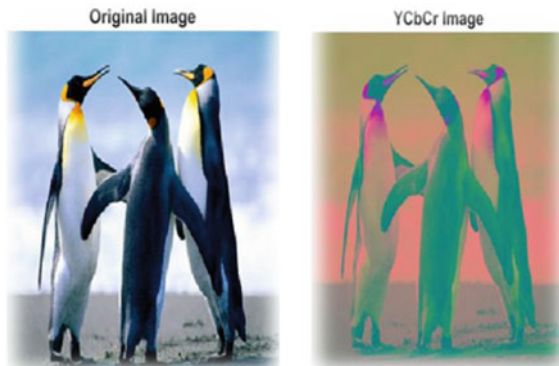
## 4 Results and Discussion

The results are obtained by converting the original image into the YCbCr by using the MATLAB tool (Figs. 7 and 8).

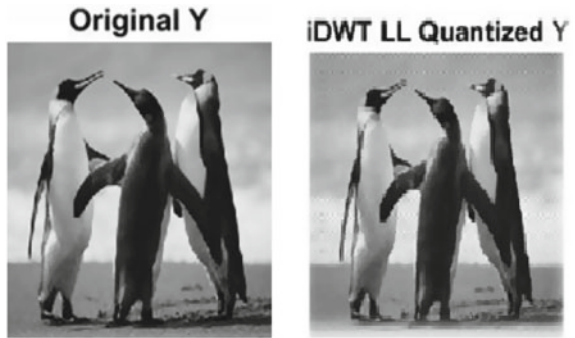
The Y space is separated from the YCbCr space and further quantized to the lower level as shown in Fig. 3. IDWT is applied to the Y space to embed the watermark. Test results for images at  $k = 0.1$  and size  $128 \times 128$  (Figs. 9, 10 and 11).

Table 1 shows the PSNR and MSE value after embedding the watermark into the original image at  $k = 0.1$ , and for the size  $128 \times 128$ , the watermark seems to be

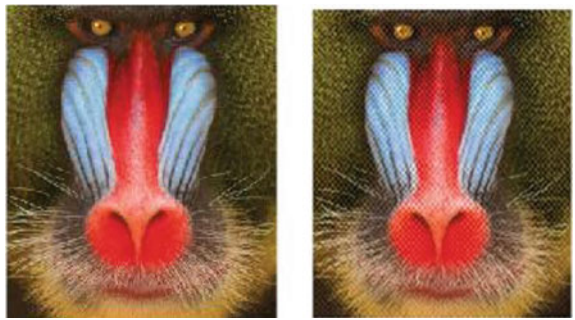
**Fig. 7** Conversion of original image to YCbCr



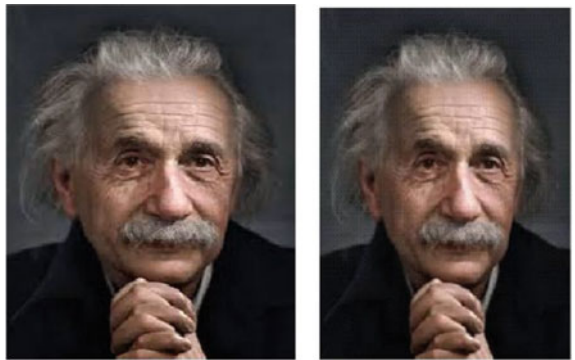
**Fig. 8** Quantization of Y space (lower level)



**Fig. 9** Watermarked image of baboon



**Fig. 10** Watermarked image of Albert Einstein

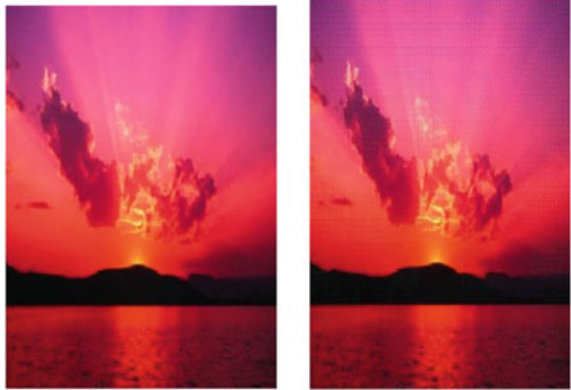


invisible with better PSNR and MSE values. Test results for images at  $k = 0.5$  and size  $128 \times 128$  are as shown in Figs. 12, 13 and 14.

Table 2 shows the PSNR and MSE value after embedding the watermark into the original image at  $k = 0.5$ , and for the size  $128 \times 128$ , the watermark seems to be visible



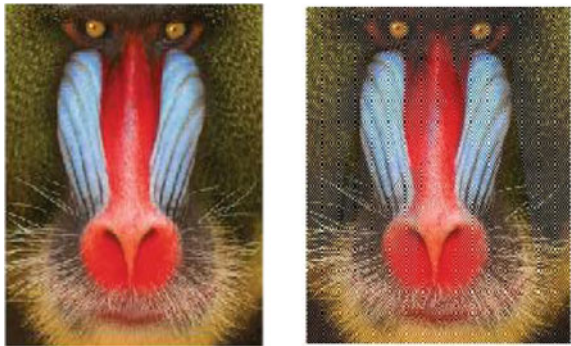
**Fig. 11** Watermarked image of sunset



**Table 1** Test result for MSE and PSNR values at  $k = 0.1$

S. no.	MSE PSNR results		
	Test images ( $k = 0.1$ )	MSE	PSNR (db)
1	Baboon	0.002	60.956
2	Albert Einstein	0.002	61.37
3	Lena	0.002	62.50
4	Sunset	0.002	61.133
5	Peppers	0.001	66.79

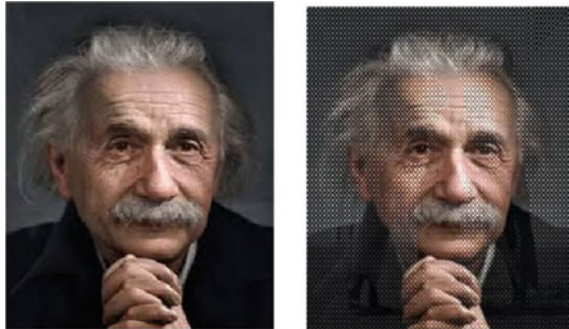
**Fig. 12** Watermarked image of baboon



## 5 Conclusion

A hybrid methodology is needed which combines both the advantages of encryption and watermarking techniques to design a better authentication as well as secure system. The proposed model provides a reliable way to authenticate images or protect copyrights protection, in which a watermark is embedded invisibly in the digital image to avoid attracting the attention of malicious attackers. In this paper, the results

**Fig. 13** Watermarked image of Albert Einstein



**Fig. 14** Watermarked image of sunset



**Table 2** Test result for MSE and PSNR values at  $k = 0.5$

S. no.	MSE PSNR results		
	Test images ( $k = 0.1$ )	MSE	PSNR (dB)
1	Baboon	0.004	43.55
2	Albert Einstein	0.003	57.3
3	Lena	0.005	53.41
4	Sunset	0.007	55.68
5	Peppers	0.004	55.482

obtained from the five test images, it is observed that increasing the order of  $k$  from 0.1 to 0.5 will make the watermark more visible and also the PSNR value decreases. So, in order to make watermark invisible, the  $k$  factor should be kept minimum 0.1 which also improves the PSNR value resulting improvement in the quality of the image. At  $k = 0.1$  and for image size  $128 \times 128$ , the results obtained are better than  $k = 0.5$ . The scheme proposed in this paper works on the dual watermarking principle that is robust and fragile watermarks which gives the image more security in terms of copyright protection and image authentication. Depending on the PSNR and MSE values, these techniques give improved results.

## References

1. Choudhary R, Parmar G (2016) A robust image watermarking technique using 2-level discrete wavelet transform (DWT). IEEE 2nd international conference on communication, control and intelligent systems (CCIS)
2. Furqan A, Kumar M (2015) Study and analysis of robust DWT-SVD domain based digital image watermarking technique using MATLAB 2015. IEEE international conference on computational intelligence and communication technology
3. Su Q, Niu Y, Liu X, Zhu Y (2012) A blind dual colour images watermarking based on IWT and state coding. *Opt Commun* 285(7):1717–1724
4. George J, Varma K, Chatterjee M (2014) Colour image watermarking using DWT-SVD and Arnold transform. In: 2014 Annual IEEE India conference (INDICON)
5. Liu X-L, Lin C-C, Yuan S-M (2016) Blind dual watermarking for colour images authentication and copyright protection. *IEEE Trans Circ Syst Video Technol*
6. Bansal D, Mathuria M (2017) Colour image dual watermarking using DCT and DWT combine approach. In: International conference on trends in electronics and informatics ICEI
7. Bhanuse SS, Kamble SD, Kakde SM (2016) Text mining using metadata for generation of side information. *Proced Comput Sci* 78:807–814
8. Naaz H, Rathkanthiwar S, Kakde Sandeep (2016) Implementation of hybrid algorithm for image compression and decompression. *Int J Eng Res* 5(5):398–403
9. Hatwar RB, Kamble SD, Thakur NV, Kakde S (2018) A review on moving object detection and tracking methods in video. *Int J Pure Appl Math* 118(16):511–526
10. Awaghate A, Thakare R, Kakde S (2019) A brief review on: implementation of digital watermarking for color image using DWT method. In: 2019 international conference on communication and signal processing (ICCSP), IEEE, pp 0161–0164
11. Nirmalkar N, Kamble S, Kakde S (2015) A review of image forgery techniques and their detection. In: 2015 international conference on innovations in information, embedded and communication systems (ICIIECS), IEEE, pp 1–5
12. Priyanka S, Pawar M, Kakde S (2019) A brief review on: mri images reconstruction using GAN. In: 2019 international conference on communication and signal processing (ICCSP), IEEE, pp 0139–0142
13. Rutuja C, Ambatkar S, Kakde S, Kamble S (2019) A brief review on: implementation of lossless color image compression. In: 2019 international conference on communication and signal processing (ICCSP), IEEE, pp 0131–0134

# A Transition Toward Green IT: An Initiative



Yashika Sharma and Sachin Lakra

**Abstract** The computers and hence computing have become an integral part of human lives in the present Information Age. The increased use of computers enables tedious tasks to be performed in a hassle-free and faster manner. But, this ever-increasing use of computing devices is taking its toll on the environment both in terms of resource utilization and its quality. Green computing is an environmental-friendly way and the current trend in the field of computing. Also, known as green IT, it paves the path to a greener version of computing. A slight transition from the conventional IT or non-green IT can make a huge difference and reduce carbon footprint to a great extent.

**Keywords** Green computing · Green IT · Energy efficiency

## 1 Introduction

Green computing is a study mainly concerned with the efficient use of resources. The primary goals are to reduce carbon footprints, to make computing more energy efficient, and minimize e-waste disposal. Computing involves various phases from its design [1], manufacturing to its end use and finally disposal [2]. Incorporating small changes at each phase can bring a huge difference to the adverse effects of IT on the environment. Minimizing power usage is another aspect that can substantially contribute in making computing less hazardous for the environment. Thus, green computing is not a technology but simply a transition from a conventional way to a new greener way. Modern IT or modern computing involves a lot more people, infrastructure, organizations [3], and money as compared to what it used to be 10 years back, and there is a continuous increase in these figures. Therefore, this

---

Y. Sharma (✉) · S. Lakra  
Manav Rachna University, Faridabad, India  
e-mail: [yashika.sharma85@gmail.com](mailto:yashika.sharma85@gmail.com)

S. Lakra  
e-mail: [sachin@mru.edu.in](mailto:sachin@mru.edu.in)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_50](https://doi.org/10.1007/978-981-15-2329-8_50)

transition must be systematic and must address many issues such as end user satisfaction, management restructuring, regulatory compliance, disposal of electronic waste, telecommuting, energy use, thin client solutions, and return on investment (ROI). The spectrum of IT industry is very huge as compared to any other industry and so are its effects. There is a need of transition so as to minimize the adverse effects of the IT industry on the environment. Green computing involves developing greener hardware as well as greener software.

## 2 Phases of Computing System

Hardware involves different phases from its designing to its disposal which is known as the life cycle of a computing system. Figure 1 depicts the typical steps in the life cycle of a computing system.

The major steps in the life cycle of a computing system are:

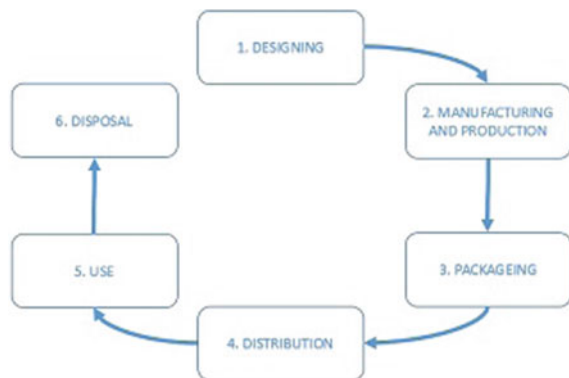
### 2.1 Designing [3]

It is an important phase to be considered in the life cycle of computer systems in general. The system should be designed such that it can be used, retired, and disposed of in an environmental-friendly way. The following points are required to be kept in mind to design such a system:

**Support, repair, and upgradability:** A system should include as many elements as can be repaired. Instead of replacing the complete system, it can be upgraded as and when required.

**Less power consumption:** Power consumption is a factor that affects both users' expenses and the environment. Systems running on less power definitely lead to a win-win situation for the users' pockets and the environment as well.

Fig. 1 Life cycle of computing systems



## ***2.2 Manufacturing and Production***

Materials used in the production of the system should be such that these can be recycled [4]. Moreover, the materials should be non-toxic so that an environmental-friendly disposal is possible.

## ***2.3 Packaging***

Multiple computers can be packed together rather than individual packing. Moreover, there can be a provision of taking back the packing material by the vendors or shipping agencies so as to reuse at least non-recyclable material. A user manual is also a part of the packaging which adds up to the overhead of disposal; manuals can be made available online to reduce the burden.

## ***2.4 Distribution***

Smaller and lighter products should be encouraged as it reduces CO<sub>2</sub> emissions and transportation costs. Transportation by sea is a greener way of transportation than air transportation.

## ***2.5 Use***

Efficient product design and power management reduce power consumption. Server center optimization also reduces energy used by these servers.

## ***2.6 Disposal [3]***

Many parts of the IT equipment can be reused as it is or with slight modification. Some other parts can be made ready for a new life cycle by recycling. Machines can also be refurbished or upgraded to be used again rather than disposing it off.

### 3 Minimizing Power Consumption

#### 3.1 Hardware

Power is the most crucial component of the IT industry. It is as important to IT equipment as food to us. Every IT device needs power in the form of electricity to function. The total power available to us is limited and over-usage may have a huge impact on the environment and the budget of the user as well. The cost factor is huge for organizations with a large number of machines. Thus, power consumption management is required to make optimized use of power.

**Reducing redundant data:** In larger organizations where the data is distributed over more than one sites, data is backed up regularly. In this process, the same piece of data is backed up several times making redundant copies of the data, consuming large amounts of bandwidth and storage space and hence consuming more power. A tool to eliminate duplicate data can be implemented. This will reduce the burden on the bandwidth and data would require less storage space. Moreover, fewer amounts of data can be dealt with by using a lower performance, energy-efficient disk.

**Virtualization:** Virtualization [5] is an energy-efficient technique which is both robust and cost-effective [6, 7]. It enables the users to work without worrying about the underlying hardware. Users need not be worried about the failure of the infrastructure; even if some components fail, users can still be able to use the services of the server. It also eliminates the need for individual hardware disks for every user. Lesser number of disks directly implies less energy consumption.

**Storage area network (SAN):** Storage area network provides a huge power saving over directly attached storage. Scaling is efficient and logical in SAN. Only a disk needs to be added to increase the storage capacity in SAN in contrast to adding a file server in case of directly attached storage.

#### 3.2 Software

**Improving network algorithm:** In big organizations where large numbers of computers are connected over a network, network delay and bandwidth utilization is a matter of concern. By optimizing the network/routing algorithms, these delays can be minimized and bandwidth utilization can be maximized. With improved bandwidth utilization, more amount of data can be transferred in less time thus minimizing overall “ON” time of the network and hence less power consumption.

**Improving scheduling algorithm:** There are various types of resources involved in the computing environment. These resources are to be scheduled among different users. The scheduling becomes critical when the system involves a large number of

users. Long idle time and poor throughput result in longer waiting state of the processes which directly implies more power consumption. Resource scheduling algorithms can be optimized using various approaches, hence, optimizing/minimizing power consumption.

### 3.3 Basic

**Switch off the system:** The system should be shut down and power should be switched off when the user is away for a longer period of time. This approach is very basic and simple but user's ignorance to this basic approach wastes a significant amount of energy.

**Low power mode:** The system should be kept on low power mode when not in use. This simple step contributes in saving considerable amount of energy.

**Maintain screen resolution:** A higher screen resolution consumes more battery power as compared to lower screen resolutions. Battery-enabled devices can have a longer battery life by choosing an optimum screen resolution.

## 4 Conclusions

The world has started moving slowly toward green IT, and there is a need of a transition from the conventional IT to a greener way on a global scale. The small steps, as mentioned in the paper, taken at the root level may curb the problem of carbon footprints to a greater extent and will contribute toward a better environment. More stringent rules [8] are required to implement these rules in a stricter way so as to attain a positive transition.

## References

1. Ranganathan P Recipe for efficiency: principles of power-aware computing. <https://doi.org/10.1145/1721654.1721673>
2. Dahiya V (2018) Green computing—an inside analysis. *Int J Adv Eng Res Dev* 5(04) Apr 2018
3. Dutta S, Gupta AK (2016) Green computing: a greener approach towards IT. In: International conference on computing for sustainable global development, IEEE, pp 50–53
4. Kurp P (2008) Green computing. *Commun ACM*. 51(10):1–19. <https://doi.org/10.1145/1400181.1400186>
5. Shakeel F, Sharma S (2017) Green cloud computing: a review on efficiency of data centres and virtualization of servers. In: International conference on computing, communication and automation, IEEE, pp 1264–1267



6. More NS, Ingle RB (2017) Challenges in green computing for energy saving techniques. In: International conference on emerging trends and innovation in ICT (ICEI), IEEE, pp 73–76
7. Saha B (2018) Green computing: current research trends. *Int J Comput Sci Eng* 6(3):467–469
8. Akpan AG (2018) Policies for green computing and e-waste in Nigeria. *Int J Comput Appl Technol Res* 7(10):386–389, ISSN:-2319–8656

# Vision-Based Real-Time Human–Computer Interaction on Hand Gesture Recognition



Poorvika Singh Negi, Riya Pawar and Roshan Lal

**Abstract** Gesture recognition technology has evolved greatly over the years. It is a field in language innovation and development with the aim of deciphering human motions by the use of certain algorithms. In this age of technology, we have fabricated boundless techniques and subsequently seen their drawbacks making us aware of our own limits in terms of speed and naturalness of the human body. The intellect and inventiveness of human beings have led to the development of many tools, gesture recognition technology being one of them. They help us extend the capabilities of our senses by combining natural gestures to operate technology, thereby reducing human efforts and going beyond human abilities. Gestures can be viewed as a route for PCs to start to comprehend human non-verbal communication, in this manner fabricating a more extravagant extension among machines and people than crude content UIs or GUIs, which still breaking point most of contribution to console and mouse and connect normally with no mechanical gadgets. Utilizing the idea of motion acknowledgment, it is conceivable to point a finger now will move as needs be. With the expeditious inventions of three-dimensional applications and the upcoming hype of virtual environments in systems, there is a need of new devices that can sustain these interactions. The advancement of user interface molds and develops the human-computer interaction (HCI). The paper aims to present a review of vision-based hand gesture recognition techniques for HCI and using it to tally finger count using the same.

**Keywords** Gesture · Computer vision · Real-time video · Human–computer interaction (HCI) · Contour extraction · Convex hull · Euclidean distance · Region of interest (ROI)

---

P. S. Negi (✉) · R. Pawar · R. Lal  
Amity University, Noida, UP, India  
e-mail: [negikpoorvika@gmail.com](mailto:negikpoorvika@gmail.com)

R. Pawar  
e-mail: [riyapanwar1999@gmail.com](mailto:riyapanwar1999@gmail.com)

R. Lal  
e-mail: [rhhokar@amity.edu](mailto:rhhokar@amity.edu)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_51](https://doi.org/10.1007/978-981-15-2329-8_51)

## 1 Introduction

Gestures are an important and striking means of communication. A natural movement caused by a body or a part of the body does not have a significant meaning and therefore differentiates gestures as the later transmits significant meaning to the observer that is receiving them [1]. Human and computing device interaction, which is inspired by the natural human-to-human interactions, can be achieved by hand gesture technology [2].

Wrapping this technology with computer vision makes it more dynamic and gives a detailed interpreted approach to gesture recognition technology [3, 4]. It is an interdisciplinary science, which is making its way in innumerable sectors, from retail and security to automotive, health, agriculture, and banking industry. From a technical point of view, we try to automate the tasks that the human vision system can do [5]. In simple words, it can be explained as when a computer and/or machine is given the capability of sight.

Picking up the accelerated speed that technology is attained today, when vision is hooked with certain algorithms, it transforms and takes a new step into another highly advanced field like machine and deep learning [6]. It grants the machine power of recognition of images, interpretation of images and solutions, and even learns in some cases [7].

This project gives a detailed analysis and review of gesture recognition technology when approached by computer vision, focusing on creating hand gesture recognition and concepts like background subtraction, motion detection, thresholding, and contours [8]. These concepts are implemented using OpenCV and python and show how to segment hand region from a real-time video sequence. Furthermore, the program recognizes the number of fingers shown in the real-time video sequence and shows the output on the screen.

## 2 Literature Review

The ever-changing nature of data requires us to have different kinds of methodology for recognizing and deciphering a signal in numerous ways [9]. However, many methods depend on key pointers spoke to the three-dimensionally arranged framework [10]. A signal can be identified with precision in view of the overall movement of the gesture.

The first step to translate development of the body will be to order them according to their basic properties and the message that the development may express [11]. Considering the conversations that take place with the help of gestures, every word or expression is communicated via them. Quek proposed in “Towards a Vision Based Hand Gesture Interface” a scientific classification that goes hand in hand with human–computer interaction (HCI) [12]. He divides them into the following so as to broadly categorize all kinds of motions:

## ***2.1 Manipulative Gesture System***

These kinds of system follow the traditional approach that was given by Richard Bolt, which is commonly known as “Put-That-There Approach.” Here, the system permits direct communication and allows the user to interact with big display objects that are in motion around the display screen [12]. The major characteristic of this system is that the gesture and the entity being controlled are coupled together and have a tight response in between each other. It is very similar to the direct interfaces that have been manipulated. However, the only difference in both is that there is a “device” present in this gesture system [3].

## ***2.2 Semaphore Gesture System***

This gesture system approach may be called “communicative” as the gestures here can be said to be equivalent to numerous gestures that are used to communicate with a machine. Each gesture/movement/pose may have a particular and designated significance [12]. However, it is important to note that unlike sign language, which has their own syntax and dynamics, the semaphore gesture system approach only comprises of isolated symbols [13].

Semaphores amount to a very small portion in the amount of gestures that are used in our day-to-day lives. But it is to be noted that they are frequently used in literature as they can most easily be achieved [14].

## ***2.3 Conversational Gesture System***

The gestures performed in due course of time by humans in their day-to-day life are known as conversational gestures. They come to us naturally, and the person making these gestures gives not a lot of thought [2, 12]. They do not have a particular or specifically fixed meaning and can be used in one more context unlike the semaphore gesture system where all the signs and gestures have fixed syntax, grammar, and meanings. They are isolated symbols and are generally accompanied by language or speech [1]. Even though they are not consciously constructed by the human mind, they can still be determined by disclosure text, personal style, culture, social presence, etc.

### **3 Methodology**

In this paper, a hand signal is perceived from a real-time video succession. To perceive the motions from a live arrangement, the first step would be to take out the hand area distinctly, evacuating the undesirable bits in the video succession [8]. In the wake of portioning the hand district, we at that point include the fingers appeared in the video succession to teach the robot dependent about the finger tally.

#### **3.1 Hand Segmentation**

The initial phase close by motion acknowledgment is clearly to discover the hand locale by taking out the various undesirable segments in the video arrangement [15]. After an amount of processing has already taken place, it is determined what areas of the image or what particular pixels are relevant for the process to move further.

#### **3.2 Background Subtraction**

In the first place, we need an effective technique to separate closer view from foundation. To do this, we utilize the idea of running midpoints [14]. We make our framework to investigate a specific scene for 30 outlines.

During this period, we register the running normal over the present casing and the past edges. By doing this, we basically make the framework aware of the foundation [16] (Fig. 1).

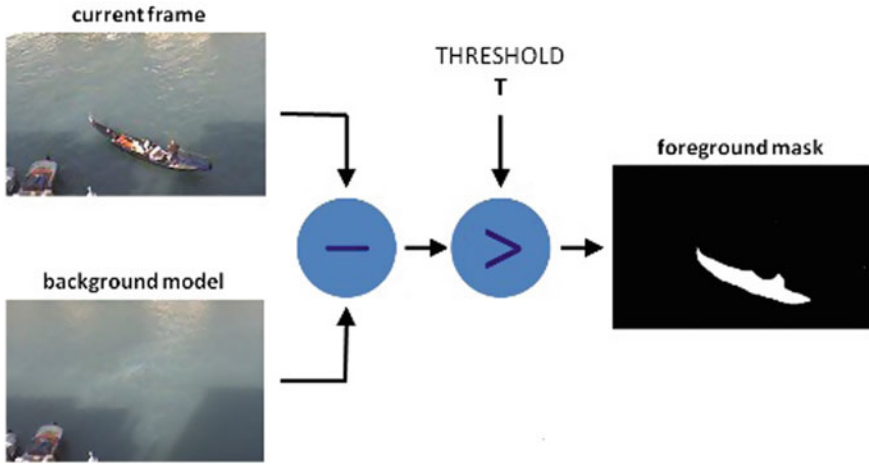
#### **3.3 Threshold**

To distinguish the hand district from this distinction picture, we have to limit the distinction picture, with the goal that lone our hand area winds up noticeable and the various undesirable areas are painted as dark [16]. This is called motion detection.

Threshold is the assignment of pixel powers to 0s and 1s based on a specific edge level with the goal that our object of intrigue alone is caught from a picture [6, 15].

#### **3.4 Contour Extraction**

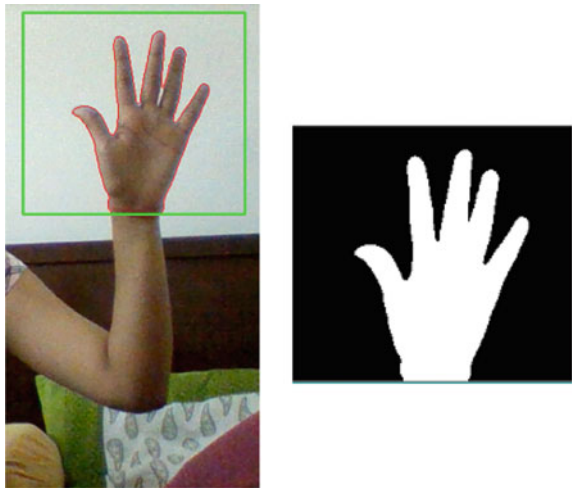
The next phase is the contour extraction. In simple words, it is basically a technique with the help of which we extract the boundary of the required part in the digital



**Fig. 1** Procedure to extract foreground mask. Sourced from <https://gogul.dev/software/hand-gesture-recognition-pl>

image in which we are applying it [9]. It gives us information about the shape of the detected part [13]. Characteristics are examined after the extraction and then further used as features in pattern classification (Fig. 2).

**Fig. 2** Picture acquired after segmentation, subtraction, threshold, and contour extraction








### 3.5 Finger Tally

The hand region has been obtained from the real-time video sequence (by the steps written above). To tally the fingers, the foremost requirement is the accessibility of a webcam or a camera attached to the system [2]. We have obtained the segmented hand region by assuming it as the most important contour (i.e., contour with the maximum location) inside the frame, and therefore, it is pertinent to that the hand occupies majority of the area inside the frame [17].

The next step is to construct an intermediary circle around the palm [13]. This is done by first detecting and computing the extreme points of the convex hull of the obtained hand region, taking them as the points joining to be the perimeter of the circle. Taking radius as the maximum Euclidean distance, the circle is constructed and henceforth we apply bitwise AND on the region of interest (ROI) and the frame [4] (Table 1).

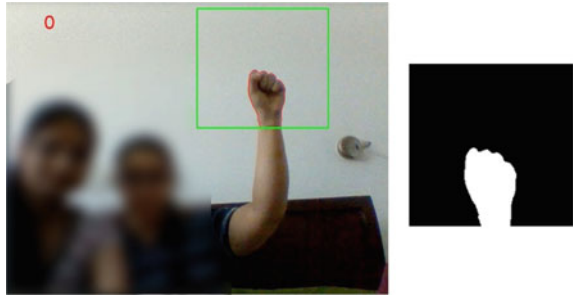
**Table 1** Algorithm to obtain finger tally count

	Step 1: obtain hand region
	Step 2: detect convex hulls and obtain extreme points
	Step 3: find center of palm
	Step 4: construct circle
	Step 5: perform bitwise AND between ROI and frame
<b>5</b>	Finger tally obtained

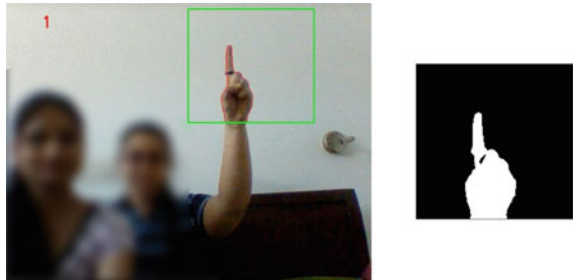
## 4 Result

When the hand is brought within the bounding box such as it is occupying the majority region, the hand is detected and a red dotted trace outlining the hand appears on the real-time video. As the fingers are gestured in the green bounding box, finger count is displayed on the top-left corner of the window in red color (Figs. 3, 4, and 5).

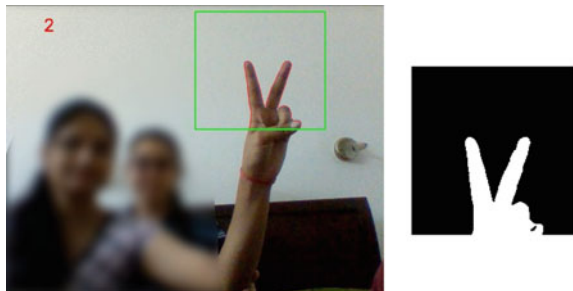
**Fig. 3** Video feed and threshold for finger tally with output 0



**Fig. 4** Video feed and threshold for finger tally with output 1



**Fig. 5** Video feed and threshold for finger tally with output 2





## 5 Conclusion

In this paper, we presented a review of vision-based hand gesture recognition techniques for HCI. With the help of different gesture systems, we classified the basic kinds of gestures. After reviewing the basic comebacks and difficulties that are faced in gesture recognition technology, we approached the same using computer vision. With the help of segmentation, background subtraction, motion detecting, thresholding, and contour extraction, we were successfully able to detect and note down the characteristics of a hand in the real-time video.

Using the algorithm that was discussed in the paper, we were successfully able to detect the number of fingers or finger count and it was displayed on the screen.

## References

1. Murao K, Terada T, Yano A, Matsukura R (2011) Evaluating gesture recognition by multiple-sensor-containing mobile devices. In: 15th Annual International Symposium on Wearable Computers, San Francisco, USA
2. Sarkar AR, Sanyal G, Majumder S (2013) *Int J Comput Appl* 71(15):0975–8887
3. Saeed A, Bhatti MS, Ajmal M, Waseem A, Akbar A, Mahmood A (2013) Android, GIS and web base project, emergency management system (EMS) which overcomes quick emergency response challenges. In: *Advances in intelligent systems and computing*, vol 206. Springer, Berlin
4. Nanaware T, Sahasrabudhe S, Ayer N, Christo R (2018) Finger spelling-Indian sign language. In: *IEEE 18th international conference on advance learning technologies*
5. Chiang T, Fan C-P (2018) 3D depth information based 2D low-complexity hand posture and gesture recognition design for human computer interactions. In: *3rd international conference on computer and communication systems*
6. Lin M, Mo G (2011) Eye gestures recognition technology in human-computer interaction. In: *4th international conference on biomedical engineering and informatics (BMEI)*, Shanghai, China
7. Geer D (2004) Will gesture recognition technology point the way. *Computer* 37(10):20–23
8. Phade GM, Uddharwar PD, Dhulekar PA, Gandhe ST (2014) Motion estimation for human-machine interaction. In: *International symposium on signal processing and information technology (ISSPIT)*
9. Yu C, Wang X, Huang H, Shen J, Wu K (2010) Vision-based hand gesture recognition using combinational features. In: *Sixth international conference on intelligent information hiding and multimedia signal processing*
10. Konwar P, Bordoloi H (2015) An EOG signal based framework to control a wheel chair. IGI Global, USA
11. Murao K, Yano A, Terada T, Matsukura R (2012) Evaluation study on sensor placement and gesture selection for mobile devices. In: *Proceedings of the 11th international conference on mobile and ubiquitous multimedia*. New York, USA
12. Quek FKH (1994) *Toward a vision-based hand gesture interface*. *Virtual Reality Software and Technology*, pp 17–31
13. Chayapathy V, Anitha GS, Sharath B (2017) IOT based home automation by using personal assistant. In: *International conference on smart technology for smart nation*
14. Shukla J, Dwivedi A (2014) A method for hand gesture recognition. In: *Fourth international conference on communication systems and network technologies*

15. Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. In: ACM computing surveys, Dec 2006
16. Bao PT, Binh NT (2009) A new approach to hand tracking and gesture recognition by a new feature type and HMM. In: Sixth international conference on fuzzy systems and knowledge discovery
17. Chaikhumphat T, Chomphuwiset P (2018) Real-time two hand gesture recognition with condensation and hidden Markov models. In: International workshop on advanced image technology (IWAIT)

# Investigations of Rectangular Dielectric Resonating Antenna Excited by CPW Feed



Alina Khan and Sovan Mohanty

**Abstract** In this paper, coplanar waveguide-based rectangular dielectric resonating antenna has been investigated by employing dielectric waveguide model (DWM) method. It is found that with CPW, feed line back radiation can be minimized and a strong capacitive matching can be obtained to achieve resonance in quad band. Proposed DRA provides efficiency as high as 90% with sufficient undesired band rejection and stable gain.

**Keywords** Dielectric resonating antenna (DRA) · Coplanar waveguide (CPW) · Dielectric waveguide model (DWM)

## 1 Introduction

In the late nineteenth century, Lord Rayleigh demonstrated that an infinitely long hollow cylindrical rod made of dielectric material could behave as a waveguide. In 1939, Robert D. Ritchmyer showed that dielectric resonator structure acts similar to cavity resonators made of conducting metallic material and named them as dielectric resonators. By this era, extensive research started in the field of dielectric resonators. By the early twentieth century, these resonators replaced bulky microwave frequency oscillators and filters. ‘Dielectric resonating antenna’ is a resonating structure which is opened up with proper modes along with perfect input characteristics, and feed line could actually behave as an efficient radiator [1]. Its resonant frequency and intrinsic impedance are function of aspect ratio and permittivity of the used dielectric material [2]. There is feasibility of miniaturized design as size of the DRA is inversely proportional to frequency, and unlike other metallic antennas where size is proportional to frequency, DRA offers size reduction at high radio frequencies. Due

---

A. Khan · S. Mohanty (✉)

Department of Electronics and Communication Engineering, SRMS College of Engineering and Technology, Bareilly, UP, India  
e-mail: [mohanty.sovan@gmail.com](mailto:mohanty.sovan@gmail.com)

A. Khan

e-mail: [alinakhan2612@gmail.com](mailto:alinakhan2612@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_52](https://doi.org/10.1007/978-981-15-2329-8_52)

509

to the absence of metallic surface, they do not involve tolerance error as micro-strip antennas do. DRA offers good temperature stability and high radiation efficiency. Only losses occur due to imperfections in dielectric material and variations in electrical parameters [3]. Almost all feeding techniques can be possible with DRA like slot coupling, micro-strip lines feed, dielectric image guide, coplanar waveguide, coplanar strip lines, etc. making them practically available for existing and upcoming technologies [2]. Resonant frequency of a DRA can also be altered by altering its shape, size, dielectric permittivity, modes, and feed line selection. Coplanar waveguide has been used as a feed element since it provides low radiation losses and high directivity to the coupler. Using CPW, there is no need to create slot in the dielectric material. Also surface-mounted elements get easy grounding [4]. DRA is accessible in various shapes as rectangular, cylindrical, triangular, spherical, and other conformal structures. In this design, we are using rectangular shape because it avoids mode degeneracy by opting its three dimensions leading to low cross-polarization levels and easy fabrication. In this paper, dielectric waveguide model (DWM) has been used for RDRA's resonant frequency analysis (isotropic case analysis). The antenna is centered on the ground plane of linear CPW feed line. The feed line couples energy to the antenna. Lowest order mode of RDRA  $TE_{111}^y$  has been excited using CPW [4]. High-frequency structure simulator (Ansys HFSS) software operating on finite element method (FEM) has been used for structural analysis of the proposed antenna design [5].

## 2 Method of Analysis

Dielectric waveguide model (DWM) is used for analyzing RDRA which is based on both Marcatili's and effective dielectric constant (EDC) approximation [2].

**Marcatili's Approximation:** In this, it is assumed that fixed dimensioned resonator is a truncated portion of infinitely long rectangular resonator, both having same characteristic impedance equation. Strongest components have sinusoidal field distribution inside the dielectric and field decays exponentially outside the dielectric following magnetic wall condition [2]. According to Marcatili's approximation, characteristic equation for wave number  $k_x, k_y, k_z$  in respective directions in  $TE_{111}^y$  mode is as [6]:

$$k_x^2 + k_y^2 + k_z^2 = \epsilon_r k_0^2 \quad (1)$$

Here,

$$k_y b = n\pi - 2 \tan^{-1} \frac{k_y}{\epsilon_r k_{y0}} \quad (2a)$$

where

$$k_{y0} = \sqrt{(\varepsilon_r - 1)k_0^2 - k_y^2} \quad (2b)$$

and

$$k_x a = m\pi - 2 \tan^{-1} \frac{k_x}{k_{x0}} \quad (3a)$$

where

$$k_{x0} = \sqrt{(\varepsilon_r - 1)k_0^2 - k_x^2} \quad (3b)$$

In above equations,

- $\varepsilon_r$  relative dielectric permittivity.
- $a, d, b$  length, width, and height of rectangular resonator.
- $k_0$  free space wave number.
- $k_{x0}, k_{y0}$  decay constant of field along  $x$ - and  $y$ -direction.
- $m, n$  number of extremes along  $x$ - and  $y$ -direction of wave number inside waveguide

Using above equations [Eqs. (2a, 2b), (3a, 3b)], we can calculate value of propagation constant  $k_z$ , formulating using Eq. (1). Under magnetic wall model which is the basic of Marcatili's approximation, tangential component of magnetic field tends to zero and only normal components exist. Hence, truncating  $z$  at  $\pm d/2$ , standing wave pattern comes along with  $z$ -direction too. According to DWM method, infinitely long waveguide is truncated to form defined dimensioned waveguide as shown in figure. It is observed that characteristic Eqs. (2a, 2b) of rectangular waveguide is same as characteristic equation of TM mode of rectangular slab guide of thickness ' $b$ ' and dielectric constant  $\varepsilon_r$ . Also it is observed that characteristic Eqs. (3a, 3b) are same as characteristic equation of dielectric sheet guide of TE mode of width ' $d$ ' and dielectric constant  $\varepsilon_r$ .

**EDC Approximation:** According to effective dielectric constant (EDC) method which is a modification of Marcatili's approximation, no tangential or normal field component is assumed to be zero as in Marcatili's approximation where tangential magnetic component is assumed to be zero following magnetic wall condition [2]. Approximating to EDC method, Eqs. (2a, 2b) for wave number in  $y$ -direction remain same while in Eqs. (3a, 3b) for wave number in  $x$ -direction, dielectric constant is replaced by effective dielectric constant given as [2]:

$$\varepsilon_{\text{eff}} = \varepsilon_r - \frac{k_y^2}{k_0^2} \quad (4)$$

Here,

- $\epsilon_{\text{eff}}$  effective dielectric constant,
- $\epsilon_r$  dielectric constant,
- $k_0$  free space wave number.

### 3 Proposed Antenna Design

Ansoft HFSS full analysis software is used using finite element method (FEM) structural analysis method [5]. Design procedure is as follows: Initially, a dielectric material has to be chosen such that there are minimum surface wave losses and waveguide leaky modes. Back conductor like PEC is not used. Rectangular slots for CPW are chosen with  $50 \Omega$  characteristic impedance [12]. CPW slot is not efficient radiators by themselves; hence, they need to couple energy from slots to the dielectric resonator [2]. For this, we need to experimentally alter slot length. Firstly, slot length was chosen half the guided wavelength (guided wavelength is 17 mm), i.e., 8.5 mm, but according to HFSS, optimum measurement was taken as 8.2 mm (Fig. 1).

Slot length is taken less than 0.04 times the guided wavelength such that resonant frequency calculation equation is given by [2]:

$$f_0 = \frac{c}{2\pi\sqrt{\epsilon_r}} \sqrt{k_x^2 + k_y^2 + k_z^2} \tag{5}$$

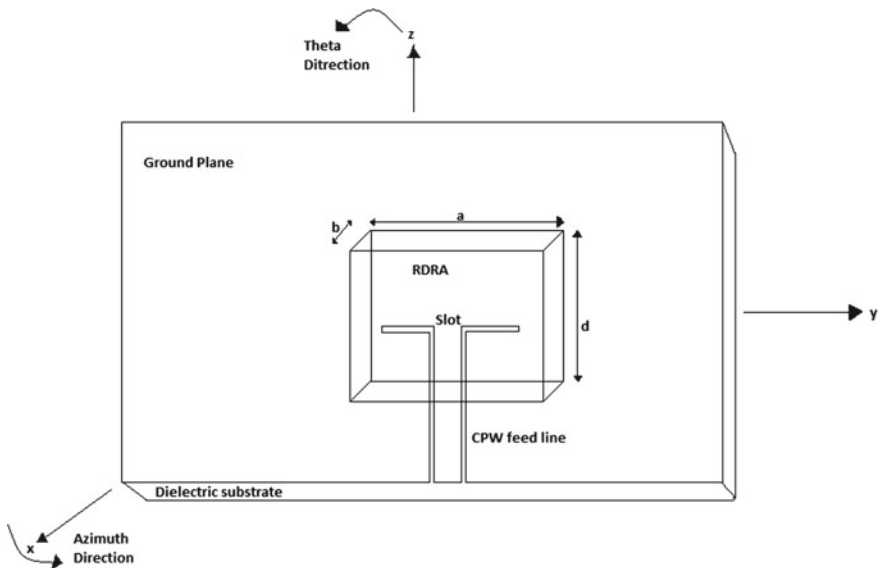


Fig. 1 Top loaded resonator (2D view) onto the ground plane, in the center of CPW feed

Here,

- $f_0$  resonant frequency of the antenna,
- $c$  velocity of light,
- $\epsilon_r$  dielectric constant,
- $k_x, k_y,$  and  $k_z$  wave numbers in  $x$ -,  $y$ -,  $z$ -directions.

Note: Values of dielectric constant and substrate thickness are chosen in order to maintain width of resonator to be more than the slot length [4]. Thus, resonator occupies total slot coverage. Antenna’s coupling and radiation characteristics are also studied.

### 4 Dimensions and Parameters

Design procedure in section III has been followed for obtaining optimum results as per the objectives. Dimension of slots on ground plane of CPW is shown in Fig. (2). Rectangular slab is placed over CPW lines as shown. Coupling between feed lines and rectangular resonator is done via narrow slots over which resonator are placed. Perpendicular to the slot is  $E$  plane ( $xz$ -plane with  $0^\circ$  azimuth angle) and parallel to the slot is  $H$ -plane ( $yz$ -plane with  $90^\circ$  azimuth angle). Here, azimuth angle has been measured from the positive  $x$ -axis. Chosen rectangular slab is Rogers RT/duroid 6010/6010LM (tm) dielectric material with dielectric permittivity of 10.2. The cuboid structure has dimensions of 140 mm  $\times$  140 mm  $\times$  100 mils. Ground plane is square of length 140 mm placed onto the slab. CPW dimensions are: Slot gap is 0.5 mm, and slot dimensions are 8.2 mm length, 70.50 mm height, and 0.26 mm width all

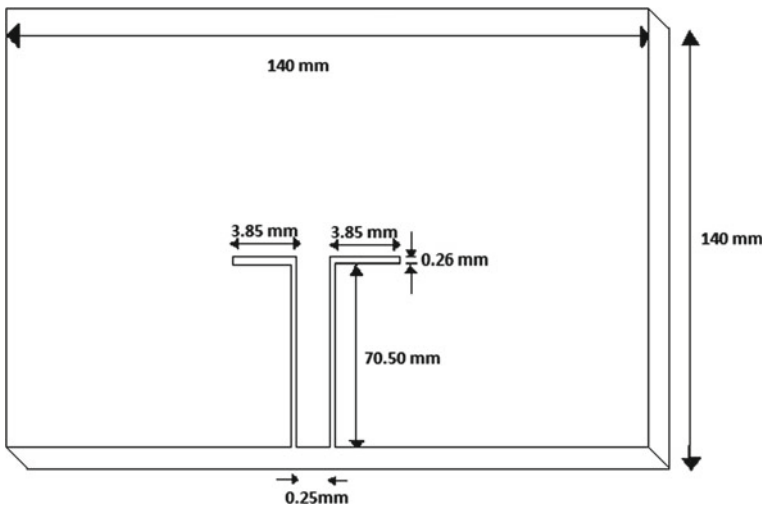
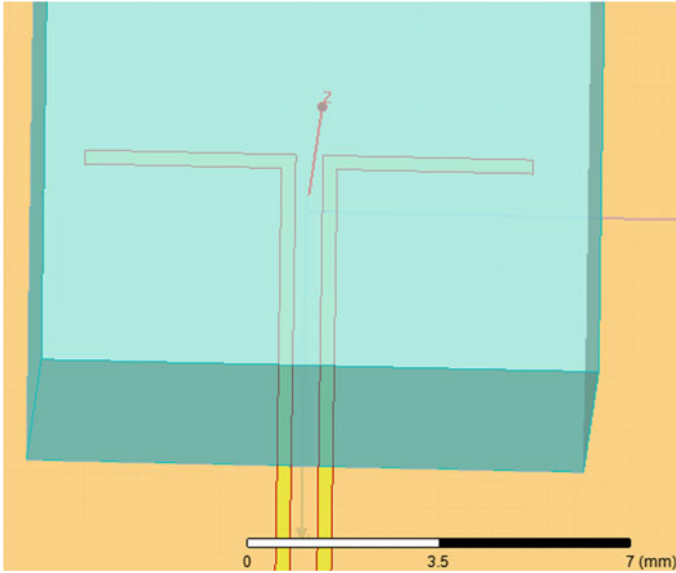


Fig. 2 Top view design dimensions of proposed antenna



**Fig. 3** Top view of slot lines with DRA

over (as shown in Fig. 2) giving  $50\Omega$  characteristics impedance. Dimensions are chosen to maximize coupling between feed and resonator. The resonator is placed onto CPW slot lines as per design with dimension  $10.2\text{ mm} \times 10.2\text{ mm} \times 7.89\text{ mm}$  and dielectric permittivity of 20.

As per the dimensions drawn on HFSS, CPW slot is shown in Fig. (3). Lowest order  $TE_{111}^y$  mode of RDRA [4] is excited using these dimensions. The electrical height of the resonator onto ground plane appears to be twice of physical dimension due to image theory [11], and it radiated like a slot fed magnetic dipole. Linear polarization of electric field with low cross-polarization level is obtained from the results of design consideration.

## 5 Results

### A. Input Characteristics

**Reflection Coefficient:** It indicates fraction of power which is lost due to reflections and not delivered to the antenna [7].

Ideally, it must be zero and is given by:

$$\Gamma = \frac{V_{\text{reflected}}}{V_{\text{incident}}} = |\Gamma|e^{j\theta}$$



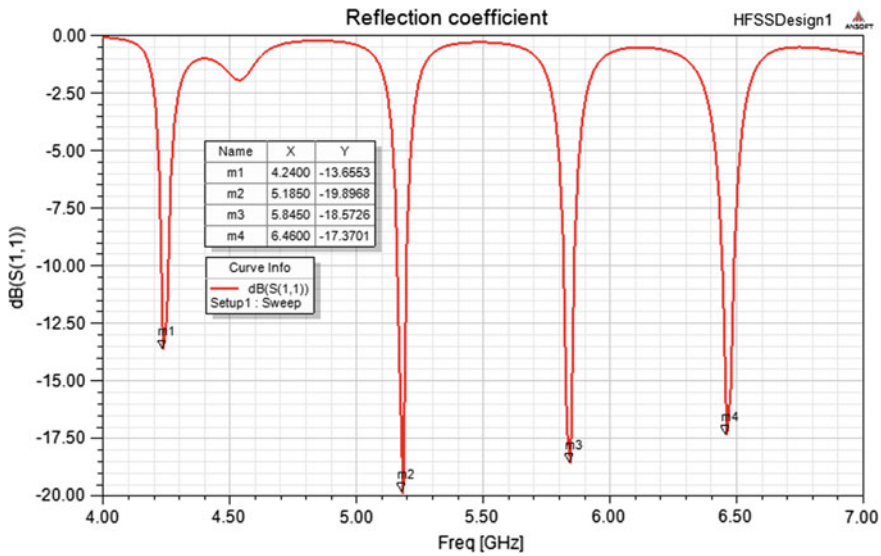


Fig. 4 Reflection coefficient plot of rectangular DRA

Here,  $j$  = complex number, and  $\theta$  = phase (Fig. 4).

Four resonant frequencies below  $-10$  dB obtained are 4.24 GHz, 5.18 GHz, 5.84 GHz, and 6.46 GHz, respectively. Hence, our antenna resonates at four different frequencies, respectively.

**Smith Chart:** Smith chart is a two-dimensional plot used for transmission line parametric analysis [8]. Unit circle is most considerable because here value of reflection coefficient is zero. Plot cuts unit circle and results in capacitive loading (Fig. 5).

## B. Output Characteristics

**Current Density Variation:** The variation of electric current per cross-sectional area as per electric field variation [7] along the feed line is shown. Maximum current density is centered under DRA along the feed line with normal electric field component (Fig. 6).

**Gain Plot in Logarithmic (dB) Scale:** Antenna gain is the ratio of output power to input power and is measured mainly in the far-field zone [7]. It is generally measured in logarithmic scale to deal with high numerical values [9]. Here, we have measured gain in both elevation and azimuth plane in  $0^\circ$  and  $90^\circ$ , respectively [10] (Figs. 7, 8, and 9).

As per the above 3D gain plot, maximum power intensity is in the main lobe direction and avoids back radiation. In our design, even without using back conductor, antenna's radiation is mainly focused in front lobe and hence possesses very gain and thus high directivity [13] (Table 1).

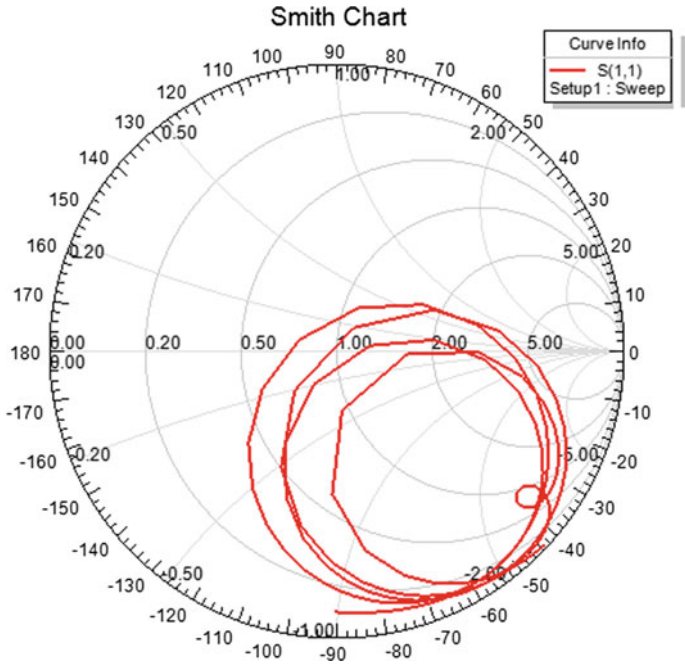


Fig. 5 Smith chart plot with capacitive loading is obtained for proposed antenna design

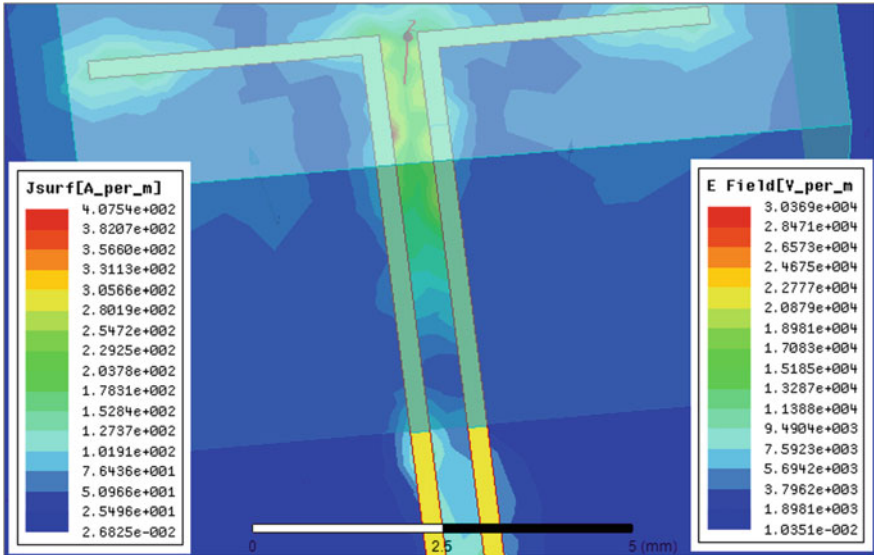


Fig. 6 Current density variation with electric field on ground plane of CPW feed line and rectangular DRA

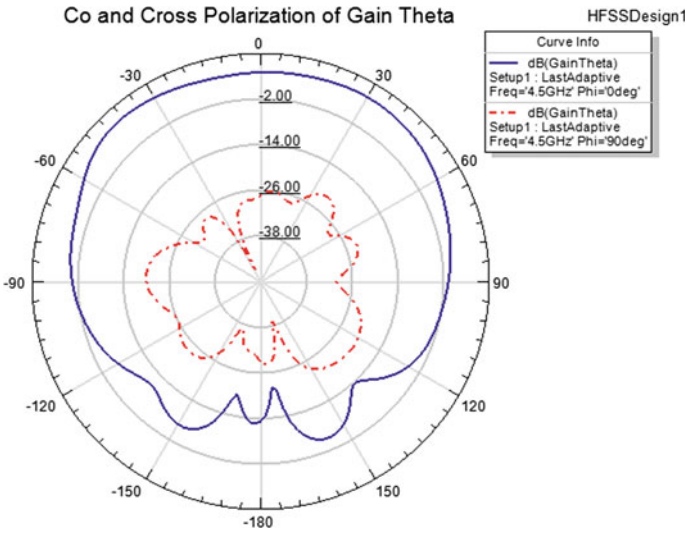


Fig. 7 Gain plot with co- and cross-polarization levels of gain in elevation plane

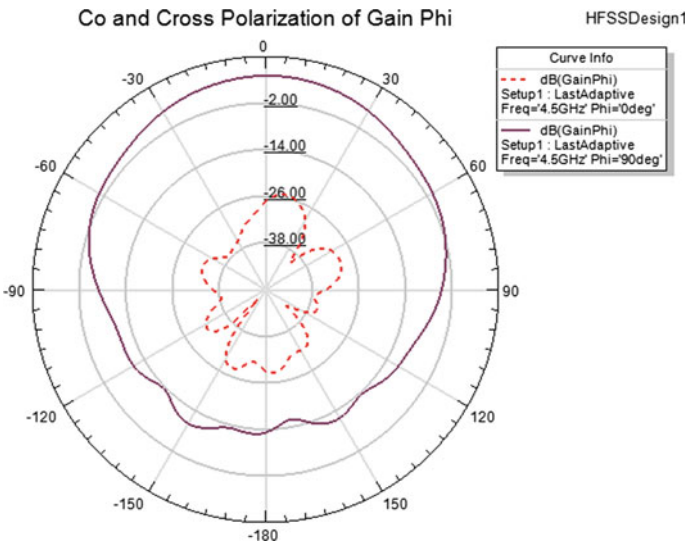
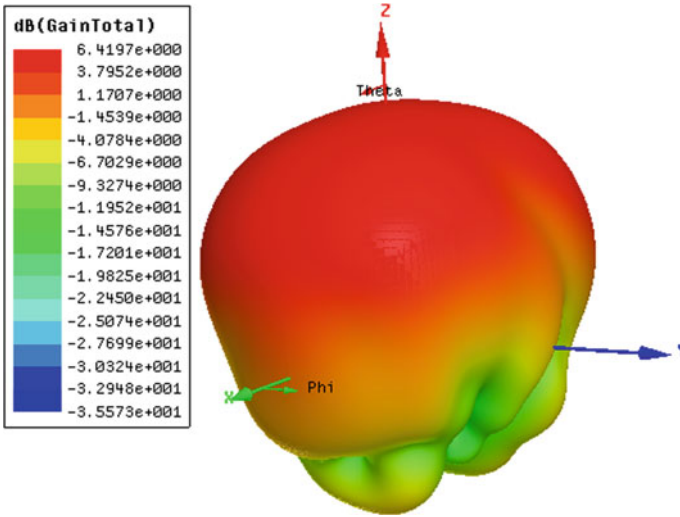


Fig. 8 Gain plot with co- and cross-polarization levels of gain in azimuth plane

## 6 Conclusion

A novel quad-band frequency rectangular DRA antenna with linear polarization has been proposed. By using CPW-based feed line, low cross-polarization levels are being achieved. The requirements for the paper have been successfully achieved with



**Fig. 9** Three-dimensional gain plot of rectangular DRA

**Table 1** Resultant antenna parameters

S. No.	Quantity	Values with units
1	Max radiation intensity	0.11064 W/sr
2	Peak directivity	4.1908 dB
3	Peak gain	4.385 dB
4	Radiated power	0.33176 W
5	Accepted power	0.31707 W
6	Incident power	1 W
7	Radiation efficiency	1.0463 dB
8	Front to back ratio	31.849 dB

impedance bandwidth efficiency of 1.622% at 4.85 MHz and 1.292% at 5.64 MHz. Further front to back ratio and back radiation characteristics can be improved by employing electromagnetic band-gap (EBG) structure. EBG patch underneath the DRA is especially effective in suppressing the *H*-plane back lobe. *E*- and *H*-plane cross-polarization level should not be affected by the presence of EBG structure. Good coupling was achieved between antenna and feed. The resulting antenna finds huge applications in Millimeter-Wave Integrated Circuits (MMIC).

## References

1. Petosa A, Ittipiboon A, Antar YMM, Roscoe D, Cuhaci M Recent advances in dielectric-resonator antenna technology. *Antenn Propag*
2. Ittipiboon A, Mongia R, Antar YMM, Bhartia P, Cuhaci M (1993) Aperture fed rectangular and triangular dielectric resonators for use as magnetic dipole antennas. *Electron Lett* 29(23):2001–2002
3. Hoffmann RK (1987) *Handbook of microwave integrated circuits*. Artech House, Norwood, MA
4. Luk KM, Leung KW *Dielectric resonator antennas*. Electronic and Electrical Engineering Research Studies
5. Ansoft Corporation (2000) Four station square. Pittsburg, PA. (Online). Available <http://www.ansoft.com/>
6. Yaduvanshi RS, Parthasarathy H (2016) *Rectangular dielectric resonator antennas theory and design*. Springer ISBN 978-81-322-2499-0 ISBN 978-81-322-2500-3 (eBook). <https://doi.org/10.1007/978-81-322-2500-3>. Springer, New Delhi, Heidelberg, New York, Dordrecht, London © Springer, India
7. Constantine AB (2005) *Antenna theory analysis and design* (3rd edn)
8. IEEE standard for definitions of terms for antennas. IEEE Std 145™-2013 (Revision of IEEE Std 145-1993)
9. Leung KW, Luk KM, Lai KYA (1993) Theory and experiment of a coaxial probe fed hemispherical dielectric resonator antenna. *IEEE Trans Antenn Propag* 41:1390–1398
10. Long SA, McAllister MW, Shen LC (1983) The resonant cylindrical dielectric cavity antenna. *IEEE Trans Antenn Propag* 31:406–412
11. Harrington RF *Time harmonic electromagnetic fields*. IEEE Press
12. Garg R, Bhartia P, Bahl I, Ittipiboon A (2001) *Microstrip antenna design handbook*. Artech House, Canton, MA
13. Antar YMM, Al Salameh MS, Seguin G (2000) Coplanar waveguide fed dielectric resonator antenna. In: *Proceedings of Progress Electromagnetic Research Symposium on PIERS 2000*. Cambridge, MA, 5–14 Jul 2000, p 695

# Design and Performance Analysis of Circular Microstrip Patch Array (2 × 2) for S-Band Wireless Applications



Saptarshi Gupta, Neeta Awasthy and R. L. Sharma

**Abstract** This paper presents design and performance analysis of circular microstrip patch array for S-band (2–4 GHz) wireless communications. This design exhibits improved gain with minimum return loss. Directivity of the antenna also improved. This design is prepared and simulated using HFSS software.

**Keywords** Wireless communication · Return loss · Patch antenna · Gain

## 1 Introduction

Microstrip patch antennas are become familiar and famous due to several advantages like lightweight, simple in structure, less space consumption, easy to mount on a variety of geometrical structure [1–3], etc. This type of antenna has some limitations like exhibit less gain, less bandwidth, more cross polarization lead to affect in frequency, increment in end-fire radiation [4], etc. The application of this antenna is like in satellite communication, mobile communication, personal electronic gadget, etc., where required space of occupancy of electronic components and the antenna plays a very major role to determine the size of the whole system. Single element antenna structure cannot provide the good signal strength also various antenna parameters like gain, directivity, radiation pattern degrades or not able to satisfy/support the desired application. Antenna array can enhance the overall antenna performance like: boosting up the strength of the signal, gain, directivity, signal to noise ratio

---

S. Gupta (✉) · N. Awasthy

Department of Electronics and Communication Engineering, Noida International University,  
Gautam Budh Nagar, UP, India  
e-mail: [ece.saptarshi@gmail.com](mailto:ece.saptarshi@gmail.com)

N. Awasthy

e-mail: [drneetaa@gmail.com](mailto:drneetaa@gmail.com)

R. L. Sharma

Department of Electronics and Communication Engineering, Greater Noida Institute of  
Technology, Gautam Budh Nagar, UP, India  
e-mail: [rlsharmadigudr@yahoo.co.in](mailto:rlsharmadigudr@yahoo.co.in)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_53](https://doi.org/10.1007/978-981-15-2329-8_53)

also reduction in minor lobes, and power wastage [5], etc. Several researches were done in the previous years [6–10] for designing an efficient antenna structure for various bands like S, C, etc. This paper deals with design and performance analysis of circular microstrip patch array (2 × 2) for S-band wireless applications. In this design, the special type of circular patch geometry introduced to enhance the overall performance, which is having a small hole at the centre of the patch geometry which is presents in all the four patches.

## 2 Model Design and Simulation

Figure 1 shows the proposed antenna array (2 × 2) structure. This antenna is designed on Fr4 substrate ( $\epsilon_r = 4.4$ ) and the design parameters are listed below:

DIELECTRIC SUBSTRATE: Fr4\_epoxy

$$Sl = 6 h + Pl = 110 \text{ mm}$$

$$Sw = 6 h + Pw = 180 \text{ mm}$$

$$\text{Substrate height } (h) = 1.6 \text{ mm}$$

where

Pl = Patch length

Pw = Patch width

fr = Resonant frequency = 2.4 GHz

GROUND:

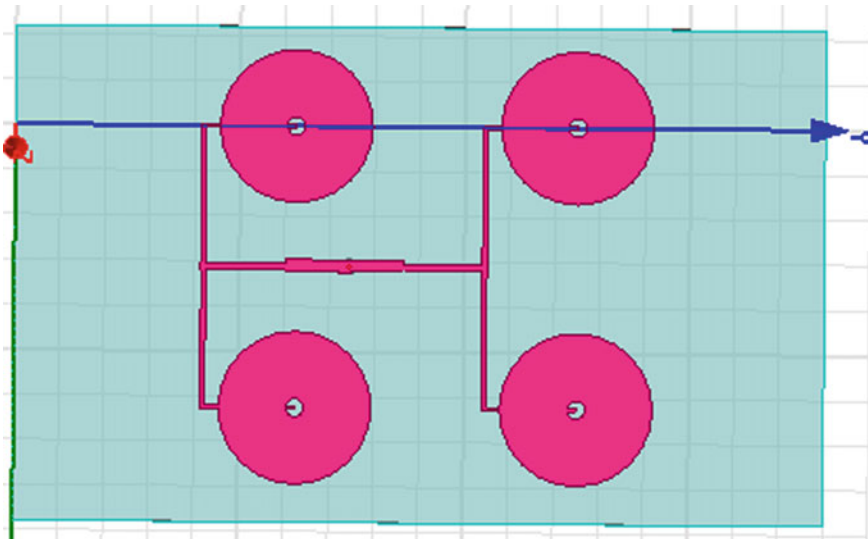


Fig. 1 Geometry of the 2 × 2 circular patch antenna array

**Table 1** Design parameters of designed antenna

Material used for substrate	Fr4_epoxy
	$\epsilon_r = 4.4$
Width of substrate (Sw)	110 mm
Substrate length (Sl)	180 mm
Substrate height (Sh)	1.6 mm
<i>Circular patch (ALL)</i>	
Patch radius (a)	17 mm
Patch mid-cut radius (ar)	2 mm
Radbox material	Vacuum
Radbox width (Rw)	110 mm
Radbox length (Rl)	180 mm
Radbox height (Rh)	32.85 mm
<i>Coaxial feed</i>	
Probe:	
Material (probe)	Pec <sup>a</sup>
Radius (probe)	0.35 mm
Height (probe)	1.6 mm
Pin:	
Material (pin)	Pec <sup>a</sup>
Radius (pin)	0.7 mm
Height (Pin)	5 mm
Coax:	
Material (coax)	Teflon
Radius (coax)	1.6 mm
Height (coax)	5 mm
Source radius	1.6 mm
Ground width	110 mm
Ground length	180 mm

<sup>a</sup>Pec Perfect electrical conductor

$Gl = 6 h + Pl = 110 \text{ mm}$   
 $Gw = 6 h + Pw = 180 \text{ mm}$   
 where  
 $Gl = \text{Ground length}$   
 $Gw = \text{Ground width}$

**RADIATION BOX:**

$\text{Radbox length} = 6 h + Pl = 110 \text{ mm}$   
 $\text{Radbox width} = 6 h + Pw = 180 \text{ mm}$   
 $\text{Radbox height} = \left(\frac{\lambda}{4} + h\right) = 32.85 \text{ mm (Table 1)}.$



In the proposed design, the special tree structure consisting of two branches each holding patches (circular in shape with hole cut in centre) used for boost up the result. In the process of feed design, probe is connected between the patch and ground, pin is connected lower part of the ground of same position of the cut which was previously made on ground. The end side of pin is connected to the coax round surface. Source is connected at last position of coax.

### 3 Result and Discussion

The proposed antenna array has been designed and simulated using HFSS software. The antenna performance parameters such as S11 (return loss), bandwidth, VSWR, gain, and radiation pattern are considered for the performance evaluation. The simulated results are shown below and the obtained values are listed in Table 2.

#### 3.1 S11 Response

S11 from Fig. 2 it is observed that at 2.40 GHz S11 value is  $-22.3$  dB.

#### 3.2 Bandwidth

Data transfer capacity of antenna allude to the scope of frequencies over which the antenna can work effectively. From Fig. 3, the BW value is 59.8 MHz.

#### 3.3 VSWR

Voltage standing wave ratio indicates how efficiently the RF power is transmitted to the load from the source. VSWR value is lie between 1 and 2 of the better receiving antenna outline. From Fig. 4, the VSWR value is 1.1662 for proposed design.

**Table 2** Reading of proposed antenna

Parameters	Values
S11	$-22.3$
BW (MHz)	59.8
VSWR	1.1662
Gain (dB)	6.71

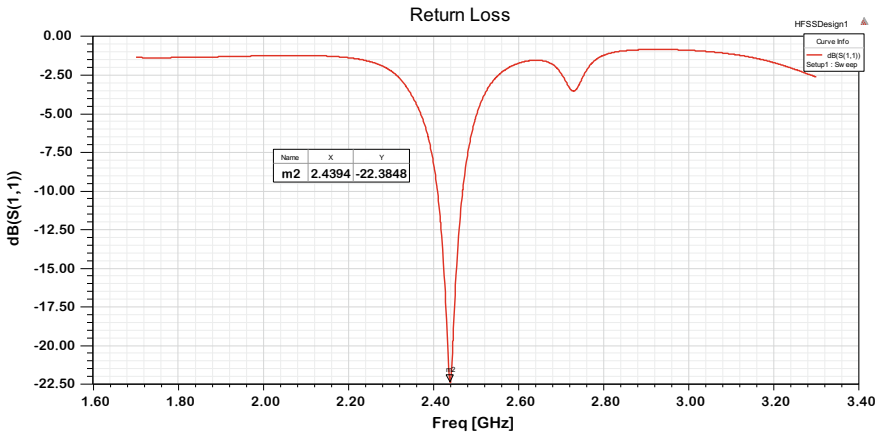


Fig. 2 S11 curve of the proposed antenna

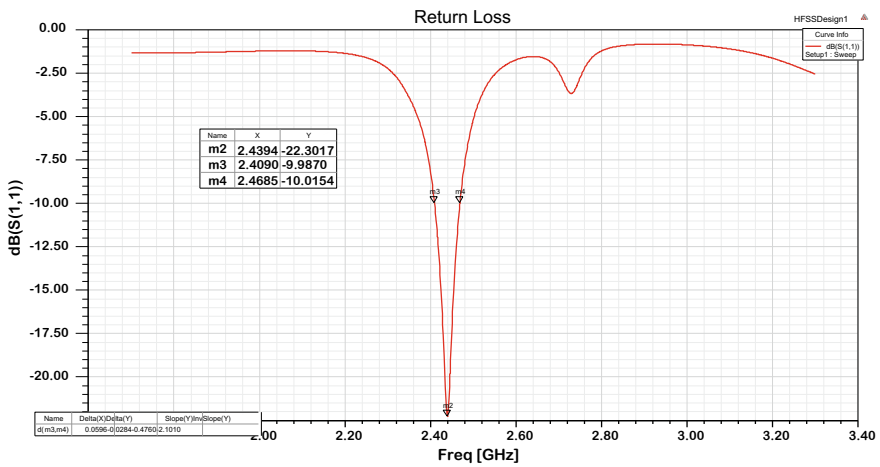


Fig. 3 BW of the proposed antenna

### 3.4 Antenna Gain

Another valuable measure depicting the execution of antenna is the gain. Regardless of the way that the gain of antenna is solidly related to directivity, it is a measure that considers effectiveness of reception apparatus and moreover its directional abilities. Here, for proposed design gain is 6.71 dB (Fig. 5).

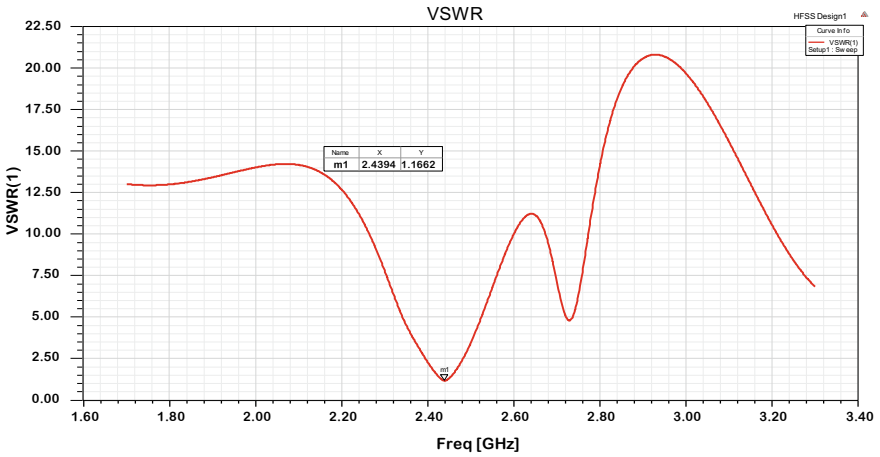
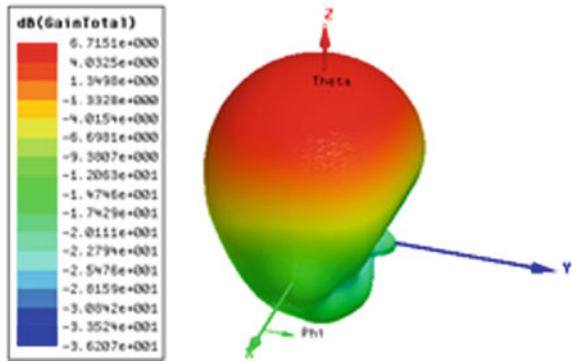


Fig. 4 VSWR of the proposed antenna

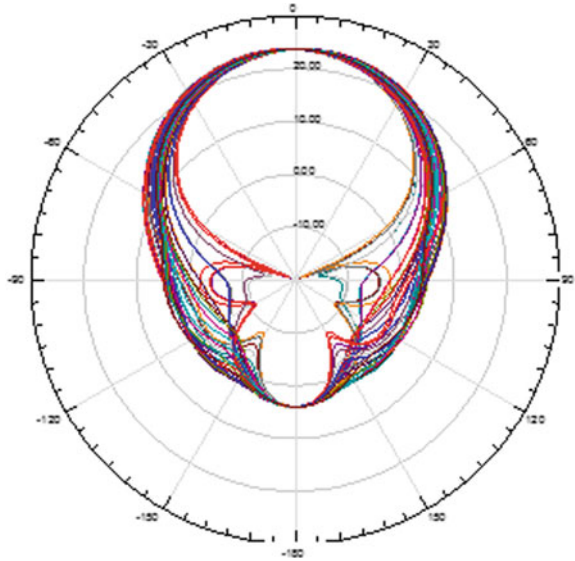
Fig. 5 Gain of the proposed antenna



### 3.5 Radiation Pattern

The radiation figure of antenna is a visual view of the disbursement of power radiated from antenna. From Figs. 6 and 7, it is observed that the reduction in minor lobe and the good amount of RF power radiation is taking place towards main lobe.

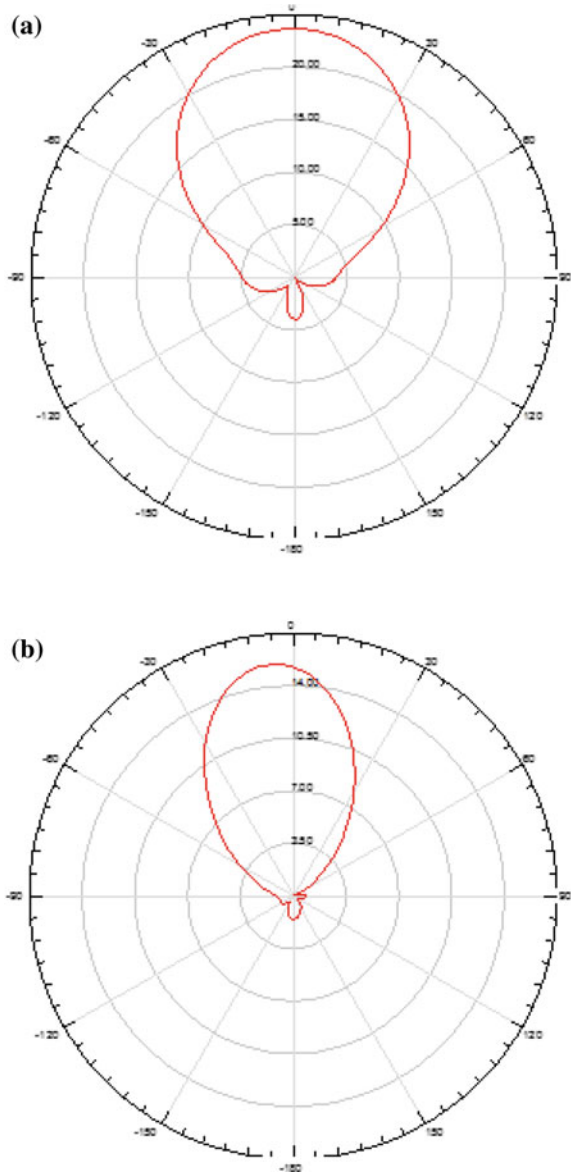
**Fig. 6** Radiation pattern of the proposed antenna



## 4 Conclusion

This paper deals with design and performance analysis of circular microstrip patch array ( $2 \times 2$ ) for S-band wireless applications. In this design, the special type of circular patch geometry introduced to enhance the overall performance, which is having small hole at the centre of the patch geometry which is presents in all the four patches. The return loss for the proposed antenna is  $-22.3$  dB and bandwidth achieved is 59.8 MHz which indicates favourable results. Obtained VSWR value 1.1662 indicates maximum RF power transmitted to the load. The proposed antenna structure uncovers the likelihood to use in WLAN, wireless personal gadget, MIMO applications, etc.

**Fig. 7** Radiation diagram  
**a** E field and **b** H field



## References

1. Kumar G, Ray KP (2003) Broadband microstrip antennas. Artech House, Boston
2. Balanis CA (2005) Antenna theory analysis and design. Wiley, New York
3. Bancroft R (2006) Microstrip and printed antenna design. Prentice Hall, India
4. Bhunia S (2013) Microstrip patch antenna's limitation and some remedies. Int J Electron Commun Technol 4:1

5. TutorialPoint. [https://www.tutorialspoint.com/antenna\\_theory/antenna\\_theory\\_arrays.htm](https://www.tutorialspoint.com/antenna_theory/antenna_theory_arrays.htm). Last accessed 2019/08/27
6. Singh A, Vijay S, Baral RN (2018) Performance analysis of high gain beamforming conformal array for avionic applications. *Int J Syst Contr Commun* 9(3):266–276
7. Gupta S, Sharma RL (2016) Microstrip patch antenna and applications. *VSRD Int J Tech Non-Tech Res* 7:171–174
8. Gupta S, Sharma RL (2017) Design, performance analysis and comparison of nearly square, nearly square corner trimmed and nearly square corner trimmed with slot antenna for 5.2 GHz wireless applications. *VSRD Int J Tech Non Tech Res* 8(8):209–214
9. Gupta S, Sharma RL (2018) Circular split ring resonator loaded circular patch microstrip antenna for 5.2 GHz ISM band. *Int J Intell Eng Syst* 11(5):246–253
10. Gupta S, Sharma RL (2018) Design and performance analysis of conformal antenna for C band wireless applications. In: 2nd international conference on micro-electronics and telecommunication engineering (ICMETE), 20–21 Sept 2018, pp 176–180

# A Secure Key Agreement Protocol for Data Communication in Public Network Based on the Diffie-Hellman Key Agreement Protocol



Chukhu Chunka , Subhasish Banerjee , Soumyajit Nag   
and Rajat Subhra Goswami 

**Abstract** Idea behind the key agreement protocol is to enable the entities to communicate safely over insecure public networks. In this paper, we proposed a secure key agreement protocol using Diffie-Hellman key agreement. In 2005, Lee and Lee [1] proposed a key agreement protocol dependent on Diffie-Hellman and guarantee that their protocol beat the attacks like man-in-the-middle attack. However, we brought up that Lee and Lee are defenseless against man-in-the-middle attack, impersonate attack and replay attack. Further proposed an improved key agreement protocol and demonstrated that the protocol is secure against the attacks. To confirm the security properties, we have done formal verification called ProVerif tools. Finally, to demonstrate the efficient, we compared other related authentication key agreement schemes with the proposed scheme.

**Keywords** Diffie-Hellman · Session key · Secure communication · Authentication · ProVerif

## 1 Introduction

Due to the quick development of web innovation, one of the major concerning issues is security which has been brought up in open network since a decade ago. In this context, the secret key exchange turns into a significant and is a basic fundamental requirement. It is necessary to verify the secret key trade between two parties.

---

C. Chunka (✉) · S. Banerjee · S. Nag · R. S. Goswami  
Department of Computer Science and Engineering, National Institute of Technology, Yupia, AP  
791110, India  
e-mail: [chukhu20@gmail.com](mailto:chukhu20@gmail.com)

S. Banerjee  
e-mail: [subhasishism@gmail.com](mailto:subhasishism@gmail.com)

S. Nag  
e-mail: [nag.soumyajit73@gmail.com](mailto:nag.soumyajit73@gmail.com)

R. S. Goswami  
e-mail: [rajat.nitap@gmail.com](mailto:rajat.nitap@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_54](https://doi.org/10.1007/978-981-15-2329-8_54)

Accordingly, numerous researchers have been working on the authenticated key agreement (AKA) protocols [1–8] for secure key exchange in the uncertain network between two parties. In 1976, a protocol dependent on the key agreement method was initially proposed by the Diffie and Hellman [2]. However, it is defenseless against man-in-the-middle attack and it does not give the genuineness. Seo and Sweeney [3] proposed a simple protocol for authentication key agreement (SAKA) to give a pre-shared secret password to confirm each other and to acquire basic session key. An additional message is attached to the session key for approval. Simultaneously, Tseng [4] raised that the SAKA is powerless against vulnerable to forgery, password guessing attacks, replay attacks, perfect forward secrecy, backward replay attacks and off-line password guessing attacks.

Ku and Wang [5] proposed a protocol to anticipate backward replay attack without change. Hsu et al. [6] extended once again that the Ku-Wang scheme remains vulnerable to an attack on modifications. Later, Lee and Lee [7] find out that Hus et al. [6] are vulnerable to the modification attacks and he proposed a scheme which repairs the security flaw. As the research continuation, Diffie-Hellman key agreement protocol [2] session key controlled by exchanging Diffie-Hellman public key of two correspondence parties. In public key cryptography itself does not give any verification and digital signature. As of later, many researchers also work on robust group key agreement protocol which resists against malicious insiders and active attacks in public networks [2, 9–13]. Harn and Lin [10] proposed a scheme for group Diffie-Hellman key agreement protocols which oppose unknown key-share attack and key compromise impersonation attack. It likewise shows that the protocol can provide key secrecy and perfect forward secrecy. In this paper, we have discovered that Lee and Lee [1] scheme is shaky from a man-in-the-middle attack, impersonate attack and replay attack. The targets of the proposed scheme is to conquer the security defect in Lee and Lee. The rest of the article arranged as follows. Section 2 describes the review of the Lee and Lee scheme; Sect. 3 consists of the security review of the Lee and Lee scheme; the proposed scheme discussed in Sect. 4; the security investigation of the proposed scheme is described in Sect. 5; the performance evolution has done in Sect. 6; and finally, the work has been concluded.

## 2 Review on Lee and Lee Scheme

In this section, we briefly describe the Lee and Lee scheme, assume  $Q$  is a piece of secret information,  $g^X$  and  $g^Y$  are public, where  $X = aQ$  and  $Y = bQ$ , the value of  $g^a$  and  $g^b$  cannot be derived from  $g^X$  and  $g^Y$  without knowing  $Q$ . Table 1 shows the list of notations and their meaning that define for proposed work.



**Table 1** List of notations and their meaning that define for proposed work

Symbol	Description
$S/R$	Sender/receiver
$E$	Encryption
$S_{pk}, R_{pk}$	Sender/receiver private key
$g$	Integer number
$p$	Prime number
$x, y$	Discrete logarithm value
$\alpha$	Set bit of dataset
$Q$	Secret information
$X, Y, a, b, c$	Secret value
$S_K, R_K$	Sender/receiver secret key
$SR_K, SA_K, RA_K, A_K$	Sender/receiver/attacker shared secret key
$C_1, C_2, C_3$	Current counter
$ID_s, ID_r$	Sender/Receiver identity

## 2.1 Key Establishment Phase

1. Sender computes  $X = aQ$ ,  $S = g^X \bmod p$  and sends  $S$  to receiver.
2. Receiver computes  $Y = bQ$ ,  $R = g^Y \bmod p$  and sends  $R$  to sender.
3. Sender computes the session key  $S_k = R^{Q-1a} \bmod p$ .
4. Receiver computes the session key  $R_k = S^{Q-1b} \bmod p$ .

Sender and Receiver get a secret session key  $S_k = R_k = g^{ab} \bmod p$ .

## 2.2 Key Validation Phase

1. Sender checks whether  $S_k \neq 1$  holds or not; if it holds, sender computes  $h(ID_s, S, S_k)$  and sends it to receiver.
2. Receiver validates  $R_k$  by checking if  $h(ID_s, S, S_k)$  sent by the sender is equal to his/her own  $h(ID_s, S, R_k)$ .
3. Receiver checks whether  $R_k \neq 1$  holds or not; if it holds, receiver computes  $h(ID_R, R, R_k)$  and sends it to sender.
4. Sender validates  $S_k$ , checking if  $h(ID_R, R, R_k)$  sent by receiver is equal to her/his own  $h(ID_R, R, S_k)$ .

Now, sender and receiver are ready to share the common secret session key  $S_k = R_k = g^{ab} \bmod p$ .

### 3 Security Weakness in Lee and Lee Protocol

#### 3.1 Man-in-the-Middle Attack

Lee and Lee [1] claimed that their scheme resists against a man-in-the-middle attack. However, we claim incorrect. Let the attacker intercepts the communication between sender and receiver. The attacker can perform the man-in-the-middle attack which is illustrated in Fig. 1.

1. Attacker replaces  $S = g^x \text{ mod } p$  with attacker key  $A = g^z \text{ mod } p$  and computes  $A_{k1} = S^{Q-1c} \text{ mod } p$ . The session key is shared between sender and attacker  $SA_k = g^{ac} \text{ mod } p$ .
2. Similarly, the attacker replaces  $R = g^y \text{ mod } p$  with attacker key  $A = g^z \text{ mod } p$  and computes  $A_{k2} = R^{Q-1c} \text{ mod } p$ . The session key shared between attacker and receiver  $RA_k = g^{bc} \text{ mod } p$ .

where,  $SA_k = g^{ac} \text{ mod } p$  is a secret session key shared between sender and attacker and  $RA_k = g^{bc} \text{ mod } p$  is a secret session key shared between attacker and receiver. In this scenario, the sender thinks that it is a key shared between the receiver and himself. Actually, attacker fools both the parties in believing session secret key is shared between the sender and the receiver. Even the attacker can intervene the key validation phase as shown below.

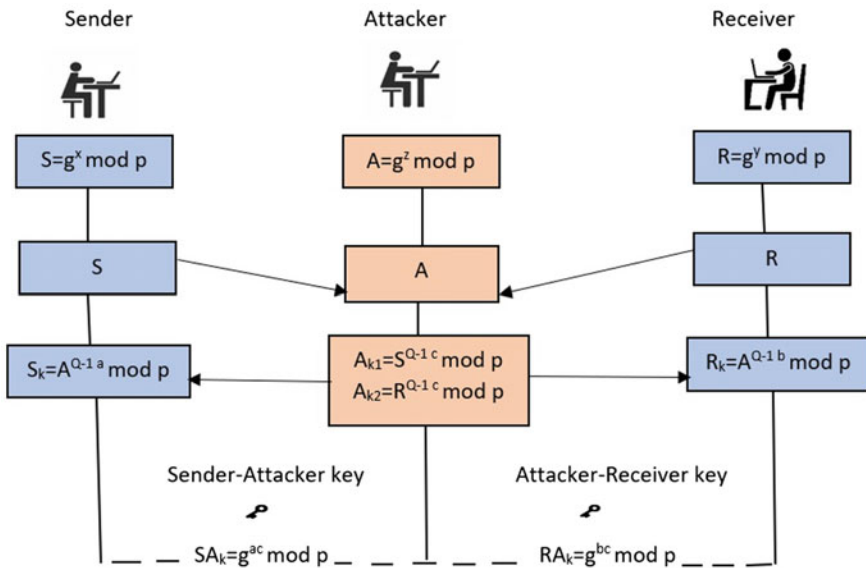


Fig. 1 Man-in-the-middle attack

1. Attacker replaces sender  $h(\text{ID}_s, S, S_k)$  with  $h(\text{ID}_A, A, A_k)$  and sends to receiver since, the attacker knows all the parameter used and computes  $h(\text{ID}_s, S, S_k)$  using message digest.
2. Attacker replaces receiver  $h(\text{ID}_R, R, R_k)$  with  $h(\text{ID}_A, A, A_k)$  and sends to sender since the attacker knows all the parameter used and computes  $h(\text{ID}_R, R, R_k)$  using message digest.

### 3.2 Reply Attack

The attacker captures the ongoing communication between sender and receiver  $S = g^X \text{ mod } p$ ,  $R = g^Y \text{ mod } p$ . Attacker observes the value of  $X$ ,  $Y$  and hash( $h$ ) value of both sender and receiver. Attacker acts as the sender to connect the receiver. When the receiver asks a proof identity, attacker sends  $h(\text{ID}_s, S, S_k)$  read from last session and receiver accepts. Now, the receiver gives all information to attacker by thinking he is communicating with the original sender. In this way, the attacker can grant access to the network.

### 3.3 Impersonate Attack

An attacker can modify the shared secret key of both the sender  $S_k = R^{Q-1a} \text{ mod } p$  and receiver  $R_k = S^{Q-1b} \text{ mod } p$  by his secret key  $A_{k1} = S^{Q-1c} \text{ mod } p$ . Now, attacker can send his secret key to both parties  $SA_k = g^{ac} \text{ mod } p$  and  $RA_k = g^{bc} \text{ mod } p$ . In this case, the attacker knows the parameter of both sender and receiver. Hence, the attacker gains access to the network.

## 4 Proposed Work

The proposed work, i.e., high secrecy Diffie-Hellman key agreement (HSDHKA) protocol focuses on how to set up a secure communication link between sender and receiver. Here, the value of  $g$  stands for integer number,  $p$  is the prime number which is freely accessible. While  $x$  and  $y$  are discrete logarithms, the secret value,  $\alpha$  is the set bit of the dataset which is secretly shared between sender and receiver. For attacker to know the secret value of  $x$  and  $y$  is troublesome in light of discrete logarithm problem. The scheme resists against man-in-the-middle attack, key offset attack/key replicating attack, impersonate attack, reply attack, perfect forward attack and guessing attack. The detailed description of the scheme has been shown in Fig. 2.

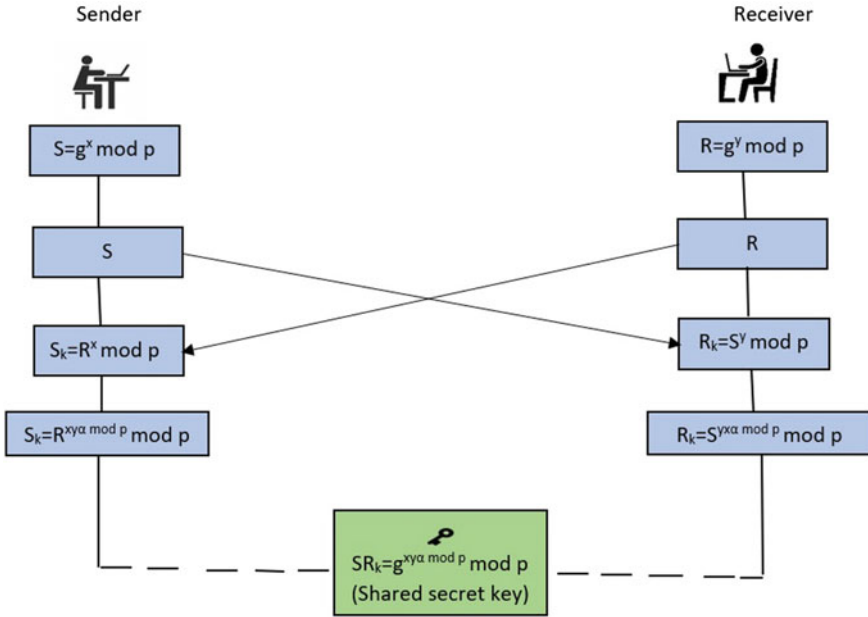


Fig. 2 High secrecy Diffie-Hellman key agreement protocol

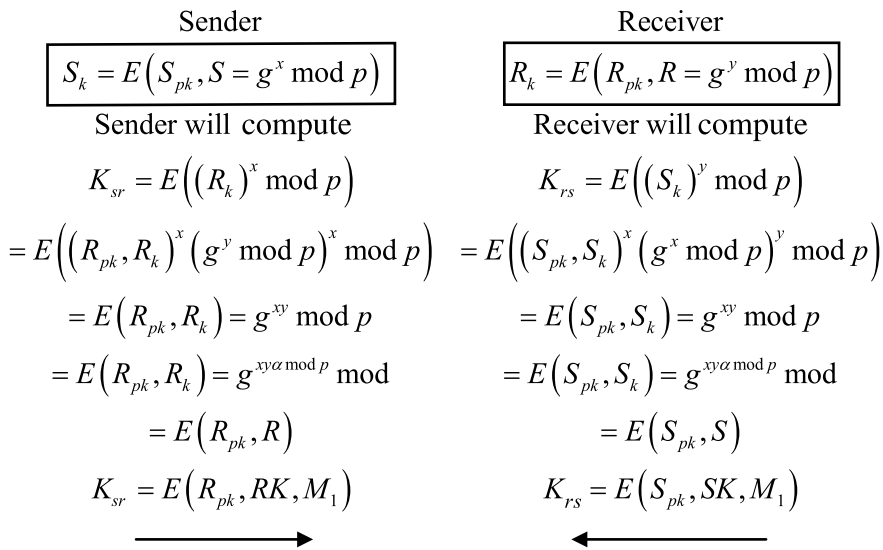
### 4.1 Key Establishment Phase

1. The sender has discrete logarithm value  $x$ , such that  $0 < x < p$ ,  $S = g^x \text{ mod } p$ .
2. The receiver has discrete logarithm value  $y$ , such that  $0 < y < p$ ,  $R = g^y \text{ mod } p$ .
3. The sender sends  $S$  to receiver and receiver sends  $R$  to sender. Note: Sender and receiver do not send the value of  $x$  and  $y$ .
4. Sender after receiving  $R$ , sender calculates,  $S_k = (R)^x \text{ mod } p$  or  $S_k = (g^y \text{ mod } p)^x \text{ mod } p = g^{xy} \text{ mod } p$ .
5. To achieve a secure symmetric session key between sender and receiver calculates by adding  $\alpha \text{ mod } p$  where  $\alpha$  is the set dataset bit shared between sender and receiver.

The sender secret session key  $SK = g^{xy\alpha \text{ mod } p} \text{ mod } p$  and receiver  $RK = g^{xy\alpha \text{ mod } p} \text{ mod } p$  session key shared between the sender and receiver.

6. Finally, both sender and receiver shared the common session secret key  $SK = RK = SR_k = g^{xy\alpha \text{ mod } p} \text{ mod } p$ .

To overcome the man-in-the-middle attack and key validation, we include the point-to-point protocol/digital signature. The steps are explained below.



1. The Sender encrypts the message sent to the receiver along with the private key of the sender and the publicly known key  $S_K = E(S_{pk}, S = g^x \text{ mod } p)$ .
2. Receiver encrypts the message along with receiver private key and publicly known key  $R_K = E(R_{pk}, R = g^y \text{ mod } p)$  sends to sender.
3. Sender figures  $K_{sr} = (R_k)^x \text{ mod } p$ , where,  $K_{sr} = E(R_{pk}, R_k = g^{xy \text{ mod } p} \text{ mod } p)$  secret key, public key is shared.
4. Receiver computes  $K_{rs} = E(S_{pk}, S = g^x \text{ mod } p)$  where,  $K_{rs} = E(S_{pk}, S_k = g^{xy \text{ mod } p} \text{ mod } p)$ , secret key, public key is shared.

Here, the attacker cannot intercept the message of sender and receiver because the sender sends  $S_{pk}$  and  $S = g^x \text{ mod } p$  to receiver in encrypted form.

The receiver decrypts the message through the sender public key since in Diffie-Hellman the attacker fooled both sender and receiver by sending his secret key and public key. The attacker creates the environment in which the sender believes that the message comes from the receiver. The sender uses the attacker's public key to decrypt the message. However, in this environment man-in-the-middle attack cannot be occurred because point-to-point connection protocol is adopted, that is,  $R_{pk}, S_{pk}$  is encrypted with message. Nevertheless, if attacker knows the public key of receiver/sender the receiver cannot decrypt the message using public key send by the attacker because there is mismatch in public key, one from the attacker and other from sender.

### 4.2 Key Validation Phase

Let sender identity as  $ID_s$ , receiver identity as  $ID_r$ ,  $h$  be one-way hash function,  $C_1$  is used as the current timestamp of sender and  $C_2$  is the current timestamp of receiver.

1. The sender computes  $h(\text{ID}_s, S_{pk}, \text{SK}, C_1)$  and send to receiver.
2. Receiver validates RK, current timestamp  $C_1$  by checking whether  $h(\text{ID}_s, S_{pk}, \text{SK}, C_1)$  is equal to his own  $h(\text{ID}_s, S_{pk}, \text{RK}, C_1)$ .  
 $h(\text{ID}_s, S_{pk}, \text{SK}, C_1) = h(\text{ID}_s, S_{pk}, \text{RK}, C_1)$ .
3. The receiver computes  $h(\text{ID}_r, R_{pk}, \text{RK}, C_2)$  and sends to sender.
4. Sender validates SK, current timestamp  $C_2$  by checking whether  $h(\text{ID}_r, R_{pk}, \text{RK}, C_2)$  is equal to own  $h(\text{ID}_r, R_{pk}, \text{SK}, C_2)$ .  
 $h(\text{ID}_r, R_{pk}, \text{RK}, C_2) = h(\text{ID}_r, R_{pk}, \text{SK}, C_2)$ .

In key validation session, if both  $\text{SK} = \text{RK} = g^{xy\alpha \bmod p} \bmod p$  and if both sender and receiver timestamp  $C_1 = C_1$ , the common session key has the validation.

In the proposed scheme, the attacker cannot modify or masquerade the message shared between the two parties.  $\text{SK} = \text{RK} = g^{xy\alpha \bmod p} \bmod p$  is public but the value of  $x$  and  $y$  are unknown values of adversary. Therefore, even if adversary knows the  $g^{xy\alpha \bmod p} \bmod p$  it would be difficult for attacker to derive the value of  $x$ ,  $y$  and  $\alpha$ , as per discrete logarithm problem. However, for the attacker it would be infeasible to forge a valid signature without knowing the sender private keys, i.e.,  $S_{pk}$  and receiver private key  $R_{pk}$ . The security totally depends upon the digital signature, current counter and hash function. If the network latency and synchronize counter  $C_1 \neq C_2$  of session key is not equal, then we can conclude that, our proposed scheme resists against man-in-the-middle attack and even attacker cannot manipulate  $S_{pk}$  and  $R_{pk}$  during the key establishment phase.

## 5 Security Analysis of High Secrecy Diffie-Hellman Key Agreement Protocol

In this section, we confirm the strength of our proposed scheme using informal and formal security analysis.

### 5.1 Informal Verification

In this subsection, we revealed the proposed scheme's informal security verification that overcomes the weaknesses of the Lee and Lee scheme. Possible attacks are as follows.

#### 5.1.1 Man-in-the-Middle Attack

In the scenario of man-in-the-middle attack, the attacker tricks both the sender and receiver by sending his two keys: one between himself and sender and another between himself and receiver. Both trust that they communicate with each other

actually, attacker at the middle. To overcome those shortcomings, we have used point-to-point communication protocol. In proposed work, the secret value of  $x$  and  $y$  is very large and for attacker, it would be difficult to know the value of ' $x$ ' and ' $y$ ' because of discrete logarithm. Even if attacker knows the value of ' $p$ ' and ' $g$ ' is not enough to compute the secret key value.

### 5.1.2 Key Offset Attack/Key Replicating Attack

In session key agreement phase, sender forwards the message  $h(\text{ID}_S, S_{pk}, \text{SK}, C_1)$  to receiver. Similarly, the receiver will also forward to the sender  $h(\text{ID}_R, R_{pk}, \text{RK}, C_2)$ . In this phase, the adversary does not know the private key of the sender and receiver. Suppose if the attacker captures and modifies it to  $h(\text{ID}_A, A_{pk}, A_K, C_3)$ . Even then the attacker will never forge the signature until he knows the private key of the sender and current counter. If the receiver receives  $(\text{ID}_A, A_{pk}, A_K, C_3)$  which was modified by the attacker, even then the attacker not able do replicating attack because of the message validity, i.e.,  $(\text{ID}_S, S_{pk}, \text{SK}, C_1) \neq (\text{ID}_A, A_{pk}, A_K, C_3) \neq (\text{ID}_R, R_{pk}, \text{RK}, C_2)$ . Hence, the verification fails and receiver terminates the session agreement. Therefore, the key offset attacker will not be applicable to proposed work.

### 5.1.3 Perfect Forward Secrecy (PFS)

Disclosure of user private keys must not allow to the compromise of any past session keys. In the proposed scheme, the attacker cannot recover any past session keys unless he/she knows the current session  $S_{pk}$  and  $R_{pk}$  shared between the entities since for every new session different key is generated and even every time new session contains new value of  $x$ ,  $y$ ,  $\alpha$  and modulo. Moreover, the Attacker will have no clue about new session and past session key. Therefore, the attacker cannot compromise for future encryption/decryption. Hence, proposed scheme resists against perfect forward secrecy.

### 5.1.4 Guessing Attack

The attacker attempts each conceivable mix to get secret key value of  $x$  and  $y$ . If the attacker is able to guess  $g^x$  and  $g^y$  even then it would be hard for him or her to decrypt/decode the value of  $\alpha \bmod p$ ,  $x$ ,  $y$  and  $\alpha$ . Therefore, proposed scheme resists against guessing attack.

### 5.1.5 Reply Attack

Here, attacker secretly listens to ongoing communication between sender and receiver. Meanwhile, the attacker stores the  $p$  and  $g$  value for performing the attack in future. In this attack, the attacker will not be able to get success, since the attacker does not get authentication, and only  $p$  and  $g$  value are insufficient for an attacker to connect with either sender or receiver. Therefore, proposed scheme resists against reply attack.

### 5.1.6 Impersonate Attack

The attacker plays the role of legal user between sender and receiver to gain access to a system/network, i.e., secret key shared between both sender and receiver. If the attacker knows the shared secret key of both parties  $SR_k = g^{xy\alpha \bmod p} \bmod p$  even then attacker will not be able to get the secret value of  $x$ ,  $y$  and  $\alpha \bmod p$ . Secondly, if the attacker somehow captures  $S_{pk}$ ,  $R_{pk}$ ,  $S_k$ ,  $R_k$  values and obtained the differ counter value  $h(\text{ID}_S, S_{pk}, SK, C_1) \neq h(\text{ID}_R, S_{pk}, RK, C_2) \neq h(\text{ID}_A, A_{pk}, A_K, C_3)$  which illustrate that, proposed scheme resists against impersonate attack.

## 5.2 Formal Verification Using ProVerif

ProVerif [14, 15] is used to verify the secrecy and authentication properties. Here, we have exhibited a proposed scheme that overcomes the man-in-the-middle attack by adding the point-to-point protocol. For verification purpose, we have characterized one channel  $ch1$  which is known to the public, two communication party sender  $S$ , receiver  $R$  and private key through which they encrypt the message.



<pre> free ch1:channel. free Rpk:bitstring[private]. free Spk:bitstring[private]. query attacker(Rpk). query attacker(Spk). (*type key[private]*) type private. type public. fun senc(private,bitstring):bitstring. fun renc(private,bitstring):bitstring. Reducforall m:bitstring,k:private;sdec(senc(k,m),k) = m. </pre>	<pre> LetersenderS(Rpk:private,Rk:public, m:bitstring)= let C1 = senc(Rpk,m)in out(ch1,C1). let recieverR (Sk:public,m: bitstring,Spk:private)= let C2 = renc(Spk,m)in out(ch1,C2); 0. process new m:bitstring; new Sk:public; new Spr:private; new Rpr:private; new Rk:public; ((!senderA(Rpr,Rk,m))   (!recieverB(Sk,m,Spr))) </pre>
--	--

**Table 2** Security comparison among authentication key agreement protocol

Security features	Lee and Lee [1]	Seo et al. [3]	HSDHKA
Man-in-the-middle attack	No	No	Yes
Key offset attack/key replicating attack	Yes	Yes	Yes
Perfect forward secrecy (PFS)	No	No	Yes
Guessing attack	Yes	Yes	Yes
Reply attack	No	No	Yes
Impersonate attack	No	Yes	Yes

The output code is executed using ProVerif 2.00. The results show that the attacker will not be able to decrypt or know the private key of both sender and receiver. Hence, it proved mutual authentication and secrecy in our protocol.

RESULT not attacker(Rpk[]) is true.

RESULT not attacker(Spk[]) is true.

## 6 Performance Evolution

In this section, in order to show the advantages of our scheme we compare security among relevant authentication key agreement schemes concerning Lee et al. [1] and Seo and Sweeney [3] schemes in Table 2.

From Table 2, it shows that our proposed scheme is resisting against man-in-the-middle attack, key offset attack/key replicating attack, perfect forward secrecy (PFS), guessing attack, reply attack and impersonate attack.

## 7 Conclusion

In this article, we have proved that Lee and Lee scheme suffers from man-in-the-middle attack, impersonate attack and reply attack. To overcome the security flaw, we have proposed an efficient scheme which is secure against man-in-the-middle attack, key offset attack/key replicating attack, perfect forward secrecy (PFS), guessing attack, reply attack and impersonate attack. Numerical derivation and comparison show that our proposed scheme resists against some known attacks.

## References

1. Lee KJ, Lee BJ (2005) Cryptanalysis of the modified authenticated key agreement scheme. *Appl Math Comput* 170(1):280–284

2. Diffie W, Hellman M (1976) New directions in cryptography. *IEEE Trans Inf Theory* 22(6):644–654
3. Seo DH, Sweeney P (1999) Simple authenticated key agreement algorithm. *Electron Lett* 35(13):1073–1074
4. Tseng YM (2000) Weakness in simple authenticated key agreement protocol. *Electron Lett* 36(1):48–49
5. Ku WC, Wang SD (2000) Cryptanalysis of modified authenticated key agreement protocol. *Electron Lett* 36(21):1770–1771
6. Hsu CL, Wu TS, Wu TC, Mitchell C (2003) Improvement of modified authenticated key agreement protocol. *Appl Math Comput* 142(2–3):305–308
7. Lee NY, Lee MF (2004) Further improvement on the modified authenticated key agreement scheme. *Appl Math Comput* 157(3):729–733
8. Blake-Wilson S, Menezes A (1999) Authenticated Diffie-Hellman key agreement protocols. In: *International workshop on selected areas in cryptography*. Springer, Berlin, Heidelberg, pp 339–361
9. Steiner M, Tsudik G, Waidner M (1996) Diffie-Hellman key distribution extended to group communication
10. Harn L, Lin C (2014) Efficient group Diffie-Hellman key agreement protocols. *Comput Electr Eng* 40(6):1972–1980
11. Brecher T, Bresson E, Manulis M (2009) Fully robust tree-Diffie-Hellman group key exchange. In: *International conference on cryptology and network security*. Springer, Berlin, Heidelberg, pp 478–497
12. Jarecki S, Kim J, Tsudik G (2010) Flexible robust group key agreement. *IEEE Trans Parallel Distrib Syst* 22(5):879–886
13. Yoon EJ, Jeon IS (2011) An efficient and secure Diffie-Hellman key agreement protocol based on Chebyshev chaotic map. *Commun Nonlinear Sci Numer Simul* 16(6):2383–2389
14. Blanchet B (2001) An efficient cryptographic protocol verifier based on prolog rules. *CSFW* 1:82–96
15. Abadi M, Blanchet B, Comon-Lundh H (2009) Models and proofs of protocol security: a progress report. In: *International conference on computer aided verification*. Springer, Berlin, Heidelberg, pp 35–49

# A Secure Steganographic Scheme Based on Chaotic Map and DNA Computing



Bhaskar Mondal 

**Abstract** This paper proposes a secure steganographic scheme based on DNA computation and chaos based on random bit generation. This scheme uses a cross-coupled chaotic map to generate random DNA sequences. DNA computing is used for encoding the secret message, and the encoded secret is embedded onto a cover image using well-known LSB substitution technique. This scheme is equally effective for hiding image as well as text. The secret message is converted to a DNA sequence, and another random DNA sequence is generated using a cross-coupled chaotic system. Now, the two DNA sequences are added using DNA addition. Then, the LSBs of the cover image pixels are substituted by the secret DNA sequence. The results of the experiment and security analysis show that the algorithm devised by us is effective and also can resist statistical and differential attack.

**Keywords** Chaotic map · DNA computation · Steganography · PRNG

## 1 Introduction

The communication network has become wider and larger in terms of size and users with time. People use the network for all aspects of daily life for rapid and massive transmission of multimedia data. With the increment of the data volume as well as users in the public network, the security of confidential data has become a significant issue. As a result, the study on information security always becomes an on-demand topic of research. Everyone has a concern to prohibit undesirable disclosure of information, and so it needs to be concerned about securing our information before transmitting it through public network [1, 2].

Techniques like cryptography [3–6], steganography [7, 8], or secret sharing [9] are used for protecting information from unauthorized access. The literal meaning of cryptography is the art of writing or solving codes. These schemes make information

---

B. Mondal (✉)

Xavier School of Computer Science and Engineering, Xavier University Bhubaneswar, Orissa  
752050, India  
e-mail: [bmondal@xub.edu.in](mailto:bmondal@xub.edu.in)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_55](https://doi.org/10.1007/978-981-15-2329-8_55)

545

(image) meaningless to any unauthorized user by conversion of data into a secret code based on some key [10].

Steganography is the practice of concealing messages or information within other non-erect text or media. It has been proposed as a technique of transmitting data (image) by embedding into a media (image, audio or video, etc.) which act as a cover. So, the steganography techniques conceal the existence of secret data (image) from the undesired users. Earlier, several steganography schemes [11] have been proposed in the literature for transmission of the secret image. Recently, a trend is notable that secret data is encrypted before embedding onto the cover image [12] which adds extra computational overhead to the steganography schemes. One of the most commonly used techniques is LSB substitution [8] as it has lower computational complexity and higher hiding capacity. An effort of improving the distortion function for steganography schemes is made using steganalytic knowledge [13]. In [14], a Steganography scheme using G.729 bitstream and matrix coding with interleaving is presented. Further, a servery on hiding text message is furnished in [15]. As every image has its own characteristics, they may be optimized to hide maximum amount of data [16]. Darbani et al. proposed a scheme to hide text in JPEG cover images based on LSB techniques [17]. Recently, researchers tried to embed secrets onto a cover image using artificial neural networks (ANN) [18]. Content-adaptive steganography is also practiced improving the methods against steganalysis and features responsible for it for JPEG cover images [19]. In [20], deep convolutional generative adversarial network (DCGAN) is used to embed secret content to the cover images to resist steganolysis feature detections [21].

The way this scheme is different from the usual techniques is, in this, the secret message is directly embedded using DNA computation into a transformed cover image without employing conventional cryptography approaches like DES, RSA, etc., on the secret image. Substitution technique uses simple LSB for embedding the DNA-encoded secret message into the cover image and results greater security with less computational complexity. So, meeting the demand for successful and secure transmission, this paper proposes a more intellectual, fast, and effective lightweight scheme.

In Sect. 2, the preliminaries of the generation of PRNS and LSB substitution are discussed in short. Section 3 intensifies the proposed steganography scheme followed by the experimental results and analysis in Sect. 4. Finally, in Sect. 5, the conclusion and future scope are discussed.

## 2 Preliminaries

### 2.1 *LSB Substitution*

The LSB substitution technique can be simply depicted in Fig. 1. As the LSBs of an image carry significantly low information about the image pixels, it can be changed

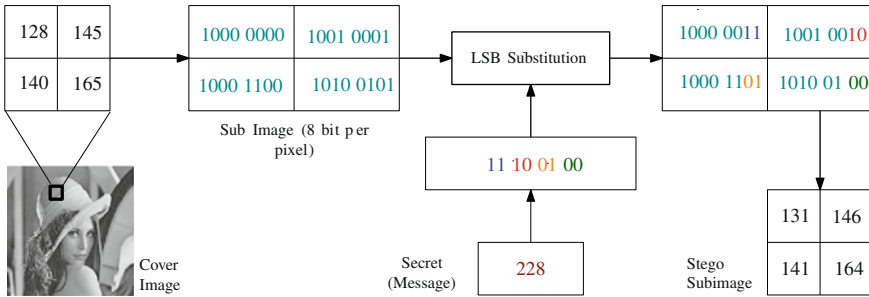


Fig. 1 LSB substitution-based steganography technique

to hide some secret. Therefore, in this method, the secret bits are been hidden in the LSBs of the cover media which is also called as stego-image in case the medium is an image [8].

### 2.2 The Chaotic Logistic Map

The widely chaotic logistic map, which is expressed by Eq. 1:

$$x_{k+1} = f(x) = \mu x_k(1 - x_k) \tag{1}$$

when  $\mu \in (3.5699456; 4)$ , the system demonstrates in chaotic nature. Some of the important characteristics that we can obtain from the logistic map are its simple structure, highly sensitive to initial condition, and two random sequences generated by different initial parameters are totally uncorrelated. It has statistically similar properties with the white noise, and therefore, chaotic signals can be used in communication [22].

### 2.3 Encryption Using DNA Sequencing

German biochemist named Friedrich Miescher was the first to observe deoxyribonucleic acid (DNA) in 1869. In this paper, the bases are represented in binary, as presented in Fig. 2. Further, the arithmetic using DNA encoding is presented in Fig. 2 [4, 23 and 24].

+	1	2	3	4	5	6	7	8
A	00	00	01	01	10	10	11	11
T	11	11	10	10	01	01	00	00
G	01	10	00	11	00	11	01	10
C	10	01	11	00	11	00	10	01

(a)

Addition					Subtraction				
+	A	G	C	T	-	A	G	C	T
A	A	G	C	T	A	A	T	C	G
G	G	C	T	A	G	G	A	T	C
C	C	T	A	G	C	C	G	A	T
T	T	A	G	C	T	T	C	G	A

(b)

Fig. 2 a DNA sequence encoding table, b addition and subtraction operations for DNA sequence

### 3 The Proposed Scheme

For retrieving the secret, the authenticated user will need the initial parameters for CCRBG which is the secret key of the scheme. For this purpose, reverse process of embedding scheme will be followed as shown in Fig. 3.

#### 3.1 The PRNS Generator

The PRNS generator is designed by combining two chaotic logistic maps in crisscross manner, in which the output of one map is used as input feedback to the other map for the next iteration at the same time. If the output of first map  $x_{i+1}$  is greater than the output of the second map  $y_{i+1}$ , it generates a bit 1 else generates 0 as presented in Eq. 2. This technique decreases the chance of predictability and generates more randomized PRNS.

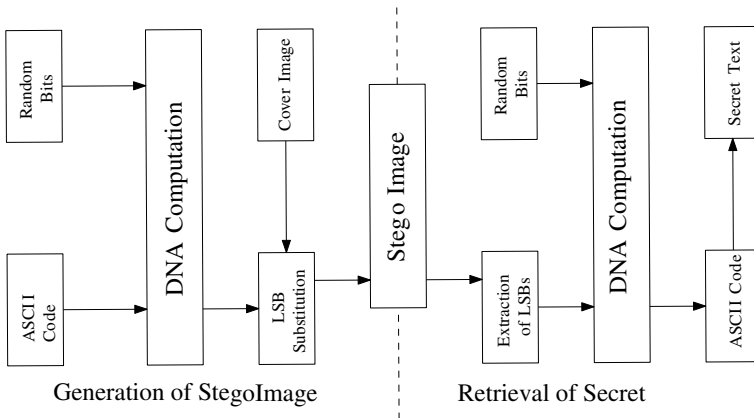


Fig. 3 Proposed scheme

$$f(x_{i+1}, y_{i+1}) = \begin{cases} 1 : x_{i+1} > y_{i+1} \\ 0 : \text{Otherwise} \end{cases} \quad (2)$$

The step-by-step process of the proposed steganography scheme:

- Step 1 Select a cover image  $I$
- Step 2 The plain text  $T$  is converted to ASCII code  $T_0$  and the ASCII code to DNA sequence  $A$
- Step 3 Generate pseudo-random bit sequence  $R$  using the cross-coupled binary bit generator
- Step 4 Convert the PRB sequence to random DNA sequence  $B$  of same length.  
 $R \rightarrow B$
- Step 5 Add DNA sequences  $A$  and  $B$ , say the result is a new DNA sequences  $C$ .  
 $C = A + B$
- Step 6 Embed the DNA sequences  $C$  to the cover image  $I$  using LSB substitution, which will give output image  $I'$ .

## 4 Experimental Results

For experimental tests, a text “Well come to the world of cryptography” is used as secret message and Lena image in Fig. 4a is used as cover image. The experimental results are presented below.

### 4.1 Histogram

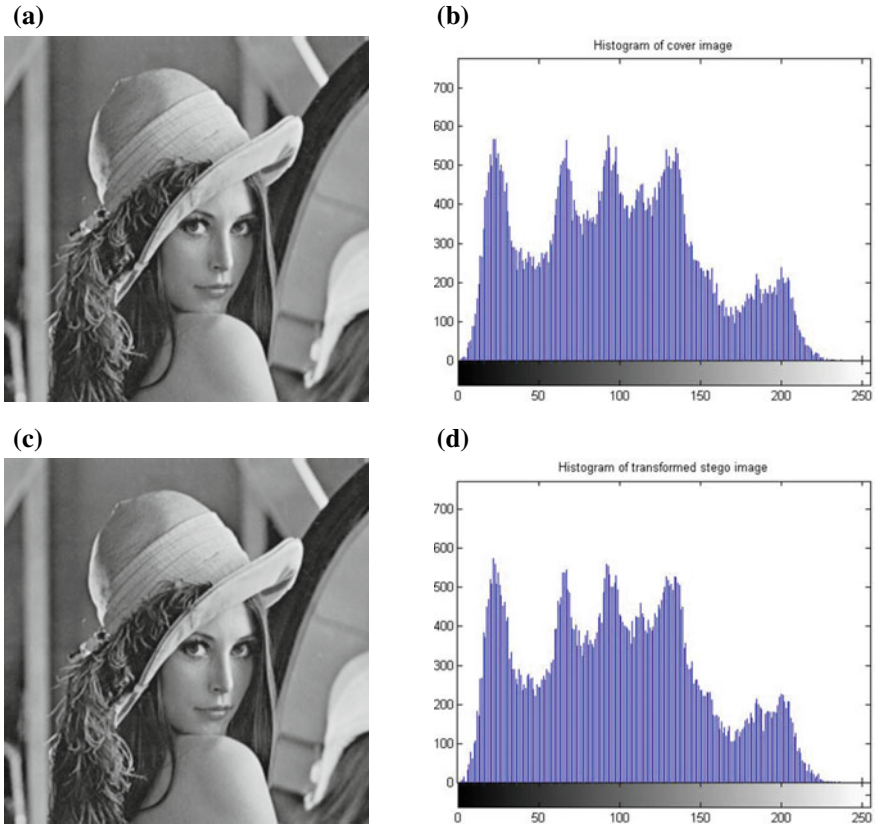
In Fig. 4a, the cover medium or plain image, and in Fig. 4b, the histogram of the cover image are presented. The stego-image is presented in Fig. 4c whether the histogram of the stego-image is in presented in Fig. 4d. Comparing the cover and stego images and corresponding histograms, it is hard to differentiate the cover image from the stego-image, which demonstrates the good quality of the proposed scheme.

### 4.2 Mean Squared Error (MSE)

For a gray scale image,  $A$  of dimension  $h \times w$  and its noisy approximation  $A'$ .

MSE ( $m$ ) is calculated by Eq. 3. Where  $w$  is the width,  $h$  is the height,  $A(i; j)$  is value of pixel in the plain cover image at  $(i; j)$ , and  $A'(i; j)$  is the value of pixel in the stego-image at grid  $(i; j)$ ;  $m$  is calculated for Fig. 4a,  $c$  is 0.





**Fig. 4** Test the scheme on Lena image to compare histogram. **a** Cover image, **b** histogram of cover image, **c** stego-image after 1 LSB substitution, and **d** histogram of stego-image

$$m = \frac{1}{hw} \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} [A(i; j) - A'(i; j)]^2 \tag{3}$$

**4.3 Peak Signal-to-Noise Ratio (PSNR)**

PSNR can be used to find the quality of a steganography scheme. PSNR gives the ratio of the original cover image and the stego-image. It can be calculated using Eq. 4. The lower value of PSNR indicates a better image quality. Calculated value of PSNR of original image is 1. Pnsnr calculated for the above experiment, Fig. 4a the cover image and 4c the stego-image, is 99 dB.

$$\text{PSNR} = 10 \log_{10} \frac{\max^2 A}{m} \quad (4)$$

#### 4.4 Key Space Analysis

The proposed scheme uses the skew tent map which had initial parameter and variable  $\mu$  and  $x_0$ . The values for  $\mu$  and  $x_0$  are used up to  $10^{-20}$  precession places. Therefore, the key space becomes  $2^{260}$ .

#### 4.5 Coefficient Correlation Test

It is used to compute the similarity between two images. The coefficient correlation is expressed by Eq. 5:

$$\text{CC} = \frac{\frac{1}{N} \sum_{k=1}^L [(P_{ck} - E(P_c))(P_{sk} - E(P_s))]}{(\sigma_x \times \sigma_y)} \quad (5)$$

Here,  $P_c$  and  $P_s$  are the gray scale values of two pixels in the same place in the cover image and stego-images. The correlation between the original cover image and stego-image is 0:0096 which shows high similarity between the two.

#### 4.6 Information Entropy Analysis

Entropy is the measure of uncertainty of a random source. It is also the amount of information a random variable carries about itself, and in Eq. 6,  $p(m_i)$  represents the probability of occurrence of the symbol. Ideally, for a source generating 256 symbols randomly with uniform probability distribution, the entropy will be 8 bits. Hence, if the value of entropy is more closure to 8, the less the chance of predictability or certainty and the system is more resistible to entropy attacks. For the test image, entropy was calculated and found 7.5733 for cover image and 7.5869 for stego-image.

$$H(m) = \sum_{i=0}^{2^N-1} p(m_i) \times \log_2 \frac{1}{p(m_i)} \quad (6)$$

#### 4.7 Maximum Deviation

The high amount of irregularity and deviation of pixels among the cover and stego-image provides good quality of encryption. First, the histogram for cover image and the stego-image is taken and their differences are calculated. Let  $d_i$  be the absolute difference between the two histograms for intensity,  $I$ , and then maximum deviation,  $D$ , is calculated as shown in Eq. 7. The test results of the maximum deviation for the cover image was 0 and for the stego-image was  $6.1799 \times 10 + 04$

$$D = \frac{d_0 + d_{255}}{2} + \sum_{i=0}^{255} d_i \quad (7)$$

#### 4.8 Irregular Deviation

This is another quantitative measure of detecting the difference between the cover image and the stego-image. A low difference assures a good quality of steganography. It first calculates the average of histogram as Eq. 8  $h_i$  is the frequency in the histogram  $H$ . Then, the irregular deviation  $D_i$  is computed as Eq. 8. For the cover image, the irregular deviation was calculated 97,920 and for stego-image 72,061.

$$D_i = \sum_{i=0}^{255} p_i - \frac{1}{255} \sum_{i=0}^{255} p_i \quad (8)$$

### 5 Conclusion

In this paper, the authors proposed a simple, effective, and secure steganography scheme. The results show that the scheme produces high-quality stego-image without using high-computational overhead of encryption. The scheme uses lightweight process like DNA computation and chaos-based random number generators. Therefore, the scheme remains a lightweight and first process suitable for any limited resource devices. The secret message is secured with secret keys. The scheme can be further modified for color images and compressed images.

## References

1. Mondal B, Kumar D, Mandal T (2015) Security challenges in internet of things. *Int J Softw Web Sci* 8–12
2. Mondal B, Singh S, Kumar P (2019) A secure image encryption scheme based on cellular automata and chaotic skew tent map. *J Inf Secur Appl* 45:117–130. <https://doi.org/10.1016/j.jisa.2019.01.010>
3. Mondal B, Mandal T (2016) A nobel chaos based secure image encryption algorithm. *Int J Appl Eng Res* 11(5):3120–3127
4. Mondal B, Mandal T (2017) A light weight secure image encryption scheme based on chaos & DNA computing. *J King Saud Univ Comput Inf Sci* 29(4):499–504. <https://doi.org/10.1016/j.jksuci.2016.02.003>
5. Mondal B, Priyadarshi A, Hariharan D (2013) An improved cryptography scheme for secure image communication. *Int J Comput Appl* 67(18):23–27. <https://doi.org/10.5120/11496-7206>
6. Mondal B, Sinha N, Mandal T (2015) A secure image encryption algorithm using lfsr and rc4 key stream generator. In: *Proceedings of 3rd international conference on advanced computing, networking and informatics*. Springer India, pp 227–237
7. Mondal B, Mandal T (2014) A secret shearing algorithm based on lsb substitution. *Int J Comput Appl* 92(4)
8. Mondal B, Singh SK (2013) A highly secure steganography scheme for secure communication. In: *Proceedings international conference of computation and communication advancement (IC3A)-2013*, vol 3, pp 88–92, Jan 2013
9. Mondal B, Mandal T, Singh SK, Acharjee KM (2013) A novel (k, n) secret key sharing scheme based on linear equations. *Int J Eng Res Technol (IJERT)* 2(10):1679–1682
10. Mondal B, Bhowmick A, Choudhury T, Mandal T (2016) A key agreement scheme for smart cards using biometrics. In: *2016 international conference on computing, communication and automation (ICCCA)*, pp 1011–1015, Apr 2016. <https://doi.org/10.1109/CCAA.2016.7813863>
11. Bilal M, Imtiaz S, Abdul W, Ghouzali S, Asif S (2013) Chaos based zero-steganography algorithm. In: *Multimedia Tools and Applications*, Mar 2013
12. Zhou X, Gong W, Fu W, Jin L (2016) An improved method for lsb based color image steganography combined with cryptography. In: *2016 IEEE/ACIS 15th international conference on computer and information science (ICIS)*, pp 1–4, June 2016
13. Wang Z, Qian Z, Zhang X, Yang M, Ye D (2018) On improving distortion functions for jpeg steganography. *IEEE Access* 6:74917–74930
14. Wu Z, Cao H, Li D (2015) An approach of steganography in g.729 bitstream based on matrix coding and interleaving. *Chin J Electron* 24(1):157–165
15. Sharma S, Gupta A, Trivedi MC, Yadav VK (2016) Analysis of different text steganography techniques: a survey. In: *2016 second international conference on computational intelligence communication technology (CICT)*, pp 130–133, Feb 2016
16. Wu X, Tan S (2018) An optimization strategy for improving security in steganography. In: *2018 IEEE SmartWorld, ubiquitous intelligence computing, advanced trusted computing, scalable computing communications, cloud big data computing, internet of people and smart city innovation*, pp 1461–1466, Oct 2018
17. Darbani A, AlyanNezhadi MM, Forghani M (2019) A new steganography method for embedding message in jpeg images. In: *2019 5th conference on knowledge based engineering and innovation (KBEI)*, pp 617–621, Feb 2019
18. Santos Brandao A, Calhau Jorge D (2016) Artificial neural networks applied to image steganography. *IEEE Latin Am Trans* 14(3):1361–1366
19. Denmark TD, Boroumand M, Fridrich J (2016) Steganalysis features for content-adaptive jpeg steganography. *IEEE Trans Inf Forensics Secure* 11(8):1736–1746
20. Hu D, Wang L, Jiang W, Zheng S, Li B (2018) A novel image steganography method via deep convolutional generative adversarial networks. *IEEE Access* 6:38303–38314

21. Mondal B, Kumar P, Singh S (2018) A chaotic permutation and diffusion based image encryption algorithm for secure communications. *Multimedia Tools Appl* 77(23):31177–31198. <https://doi.org/10.1007/s11042-018-6214-z>
22. Wang R, Li Q, Yan D (2006) A high robust audio watermarking algorithm. In: *Intelligent control and automation. The sixth world congress on WCICA 2006*, vol 1, pp 4171–4174
23. ur Rehman A, Liao X, Kulsoom A, Abbas SA (2015) Selective encryption for gray images based on chaos and dna complementary rules. *Multimedia Tools Appl* 74(13):4655–4677
24. Wang XY, Zhang HL, Bao XM (2016) Color image encryption scheme using CML and DNA sequence operations. *Biosystems* 144:18–26

# Rice Disease Detection and Classification Using Deep Neural Network Algorithm



S. Ramesh and D. Vydeki

**Abstract** In this paper, deep neural networks were proposed to find the crop disease for the normal image, brown spot, blast, sheath rot and bacterial blight. Dataset consists of 209 images. In the image preprocessing, RGB images are converted into HSV to remove the background portion using hue and saturation part. The image segmentation by k-means clustering, various colour and texture features are extracted. The classification is done with existing KNN algorithm. The accuracy obtained is 88% bacterial blight, 82% blast, 88% brown spot, 87% sheath rot and 86% normal images. To improve the accuracy our proposed DNN is implemented. The accuracy obtained for DNN is 93% bacterial blight, 89% blast, 93% brown spot, 92% sheath rot and 96% normal images.

**Keywords** Rice · Disease and normal images · Preprocessing · Segmentation · Feature extraction · KNN and DNN algorithm

## 1 Introduction

Rice is the main staple food crop in India. It is also the world's largest important food crops for Indians besides wheat and corns [1]. It is the major carbohydrate source for people in India. At the same time 90% people including rice in their daily needs. The problem in the agriculture field is climatic change, pests and diseases, and it will reduce the rice maximum production.

However, rice is quite susceptible by diseases. The disease will affect the crop from initial to seedling stage. It is threatened by various diseases when growing, resulting in tremendous economic loss in agriculture worldwide. It results in huge economic loss for farmers. The size of the affect depends upon the level of attack in the plants. The diseases will affect the crop at any stage and at any time.

---

S. Ramesh (✉) · D. Vydeki  
VIT Chennai, Chennai, India  
e-mail: [s.ramesh2015@vit.ac.in](mailto:s.ramesh2015@vit.ac.in)

D. Vydeki  
e-mail: [vydeki.d@vit.ac.in](mailto:vydeki.d@vit.ac.in)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_56](https://doi.org/10.1007/978-981-15-2329-8_56)

Timely detection of diseases will reduce the yield losses. In most cases, the detection is performed manually either by human visual assessment or by laboratory analyses [2].

According to the report of 2016–17 (ICAR), India is the second-largest rice production in the globe, contributing around 275.11 t of world production [3, 4].

India economic development is based on the agriculture farming and the major of the farmland cultivating rice crop in the country. The farmers face losses due to majority of diseases. The main sources of diseases are bacteria, fungi and viruses [5, 6].

To overcome these issues, it needs a suitable recognition system for disease detection. This paper proposed the plant disease recognition using the neural network algorithm.

Rest of the paper is organized as follows: Sect. 2 explains the brief literature survey, Sect. 3 comprises a proposed method, Sect. 4 shows the performance measures of confusion matrix, and Sect. 5 represents the conclusion of work.

## 2 Literature Survey

Some of the latest research work belonging to rice plant diseases is given below.

The author Pujari et al. [7] proposed the concept of image processing techniques to identify and classify the fungi diseases symptoms on horticulture and agriculture crops. The proposed architecture will remotely monitor the crops using GSM to reduce the yield loss and its results in its intelligent farming.

The author Yeh et al. [8] proposed the concept of machine learning methods for comparison of hyperspectral disease assessments. The proposed method is used to detect the anthracnose disease in strawberries fruits. These diseases will affect the plants up to 100%. So the above methods will reduce the yield loss with the detection as earliest.

The author Singh and Misra [9] proposed the concept of plant disease leaf detection using image segmentation and soft computing techniques.

## 3 Proposed Method

### 3.1 Acquisition of Images

The images are collected for bacterial blight, brown spot, blast, sheath rot and normal images with high resolution camera. The dataset totally consists of 209 images, which includes bacterial blight 41 brown spot 37, blast 59, sheath rot 39 and normal 33. Figure 1 shows the sample images.

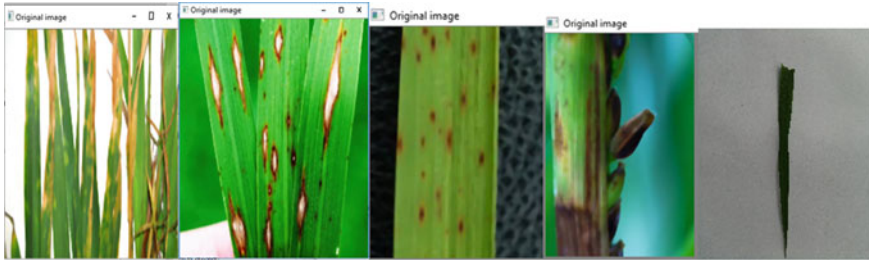


Fig. 1 Various disease affected and healthy image

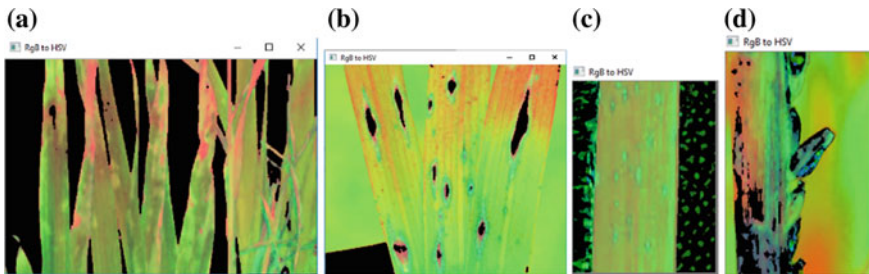


Fig. 2 a Bacterial blight, b blast, c brown spot, d sheath rot

### 3.2 Preprocessing

In the preprocessing, the images are cropped and resized into the dimension of 250 × 300 pixels. The background part of the collected images is removed using hue and saturation part.

The RGB are converted to HSV for various disease images are shown in Fig. 2.

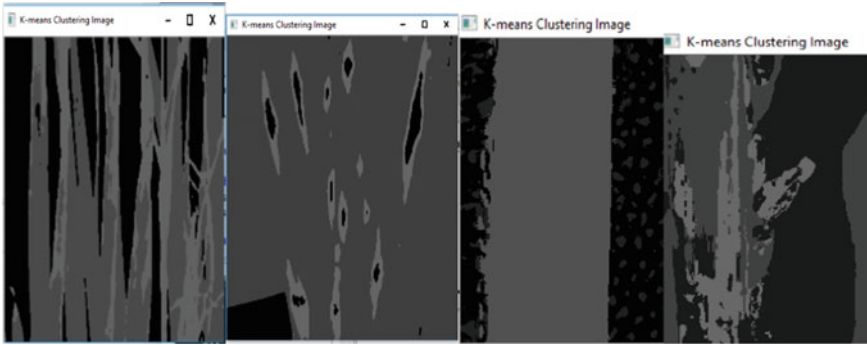
### 3.3 K-Means Clustering for Segmentation

Image segmentation is used to extract the part of diseases images. For segmenting, the portion k-means is used Fig. 3.

### 3.4 Extracting Features

The extracted features are colour and texture. In the colour feature, standard deviation and mean are extracted. For the texture following features were extracted as contrast, energy, homogeneity and correlation [10].





**Fig. 3** Various disease images segmentation using k-means clustering

### Mean Value

The colour value is extracted with the formula [1, 2]

$$\text{Mean} = \sum_{j=1}^N Pij. \quad (1)$$

### Standard deviation

$$S.D = \sqrt{1/N \sum_{j=1}^N (Pij - Ei)^2}. \quad (2)$$

The texture feature is extracted with the formula [3, 5, 7, 8]

$$\text{Energy} = \sum_{i,j=0}^{N-1} (Pij)^2 \quad (3)$$

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{Pij}{1 + (i - j)^2} \quad (4)$$

$$\text{Contrast} = \sum_{i,j=0}^{N-1} Pij(i - j)^2 \quad (5)$$

$$\text{Correlation} = \sum_{i,j=0}^{N-1} \frac{Pij(1 - \mu)(1 - \sigma)}{\sigma^2}. \quad (6)$$

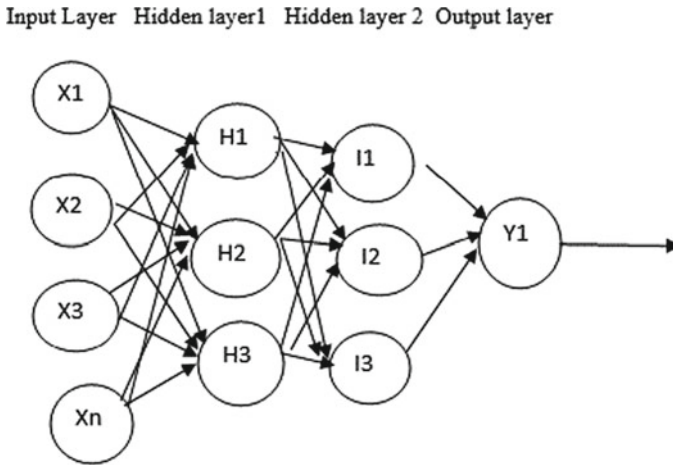


Fig. 4 Deep neural network architecture

### 3.5 Optimized DNN-Based Classification

Deep learning is a subset of machine learning in artificial intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabelled. The architecture of DNN consists of i/p layer, hidden layer and o/p layer. Figure 4 shows the architecture of DNN.

## 4 Confusion Matrix

Figure 5a, b shows the KNN and DNN confusion matrix. From this confusion matrix, true positive, false negative, true negative and false positive are evaluated.

### 4.1 KNN

Figure 5a represents the KNN confusion matrix. The sheath rot values obtained for TP, FN, TN and FP are 20, 8, 163, 18; the brown spot values of TP, FN, TN and FP are 18, 14, 166, 11; the blast values of TP, FN, TN and FP are 42, 22, 127, 18; the bacterial blight values of TP, FN, TN and FP are 29, 11, 153, 16; the normal values of TP, FN, TN and FP are 26, 16, 156, 11.

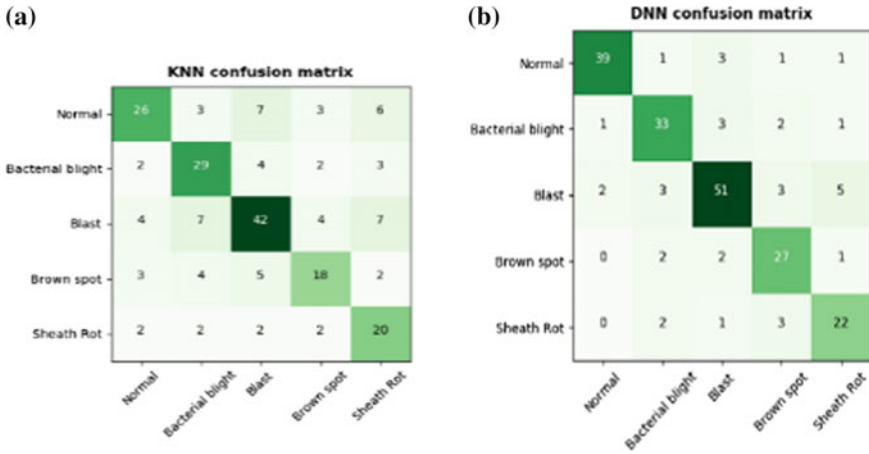


Fig. 5 a KNN confusion matrix, b DNN confusion matrix

### 4.2 DNN Confusion Matrix

Figure 5b represents the DNN confusion matrix. The sheath rot values obtained for TP, FN, TN and FP are 22, 6, 173, 8; the brown spot values of TP, FN, TN and FP are 27, 5, 168, 9; the blast values of TP, FN, TN and FP are 51, 13, 136, 9; the bacterial blight values of TP, FN, TN and FP are 33, 7, 161, 8; the normal values of TP, FN, TN and FP are 39, 6, 161, 3.

### 4.3 F1 Score

Figure 6 shows the graph accuracy for the comparison of both classifiers with different classes. The values obtained in DNN algorithm are 81% for normal, 69% bacterial leaf blight, 71% blast, 66% brown spot and 61% sheath rot. The values obtained for KNN algorithm are 45% for normal, 52% bacterial leaf blight, 51% blast, 43% brown spot and 42% sheath rot.

### 4.4 Precision

Figure 7 shows the graph accuracy for the comparison of both classifiers with different classes. The values obtained in DNN algorithm are 92% for normal, 79% bacterial leaf blight, 85% blast, 75% brown spot and 74% sheath rot. The values obtained for KNN algorithm are 69% for normal, 63% bacterial leaf blight, 68% blast, 62% brown spot and 52% sheath rot.

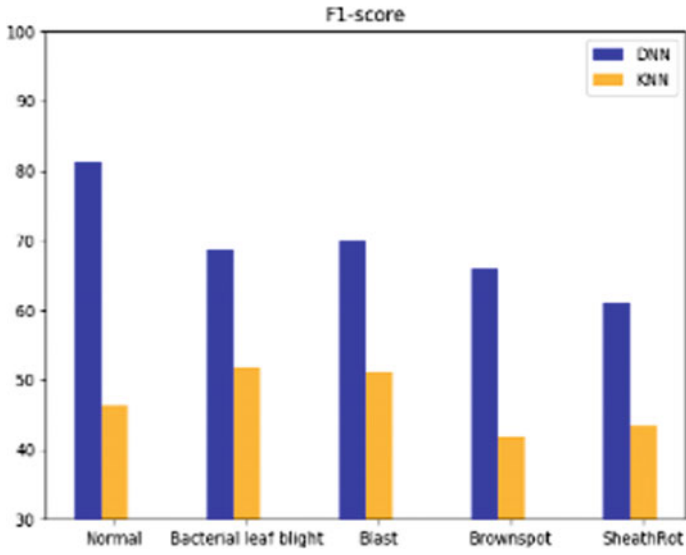


Fig. 6 F1 score performance measure

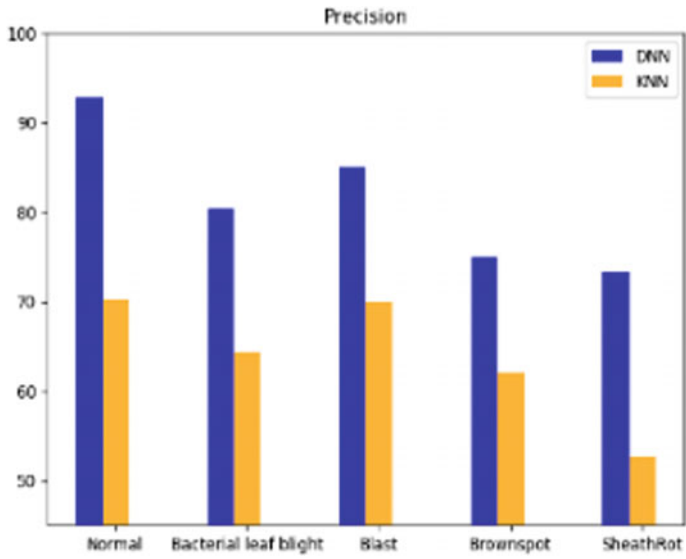


Fig. 7 Precision

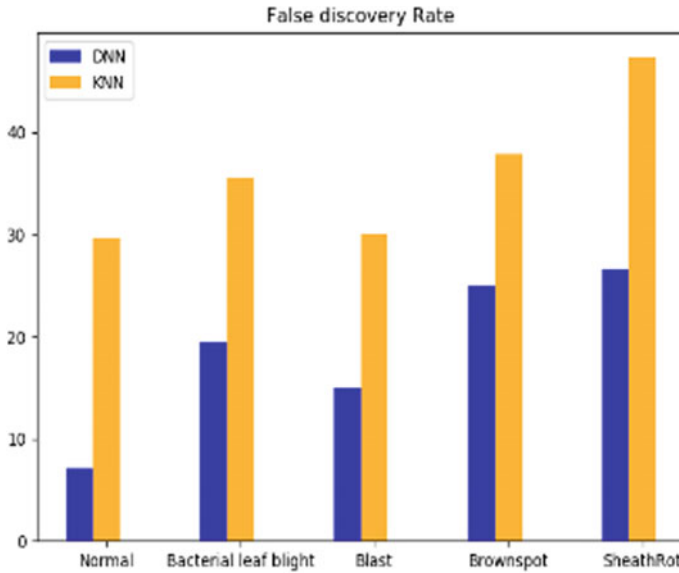


Fig. 8 False discovery rate

### 4.5 False Discovery Rate

Figure 8 shows the graph accuracy for the comparison of both classifiers with different classes. The values obtained in DNN algorithm are 7% for normal, 19% bacterial leaf blight, 13% blast, 25% brown spot and 27% sheath rot. The values obtained for KNN algorithm are 29% for normal, 34% bacterial leaf blight, 30% blast, 35% brown spot and 48% sheath rot.

### 4.6 False positive Rate

Figure 9 shows the graph accuracy for the comparison of both classifiers with different classes. The values obtained in DNN algorithm are 1.9% for normal, 4.4% bacterial leaf blight, 6.1% blast, 4.5% brown spot and 4.1% sheath rot. The values obtained for KNN algorithm are 6.3% for normal, 9.8% bacterial leaf blight, 12.2% blast, 5.9% brown spot and 9.8% sheath rot.

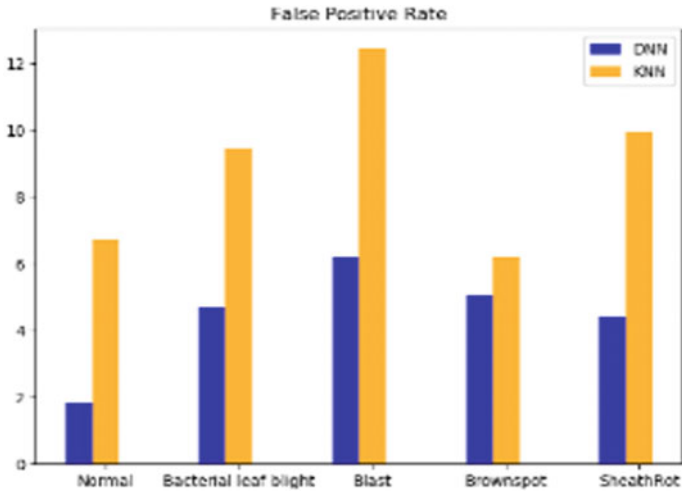


Fig. 9 False positive rate

### 4.7 Negative Predict Value

Figure 10 shows the graph accuracy for the comparison of both classifiers with different classes. The values obtained in DNN algorithm are 96% for normal, 95% bacterial leaf blight, 93% blast, 97% brown spot and 97% sheath rot. The values

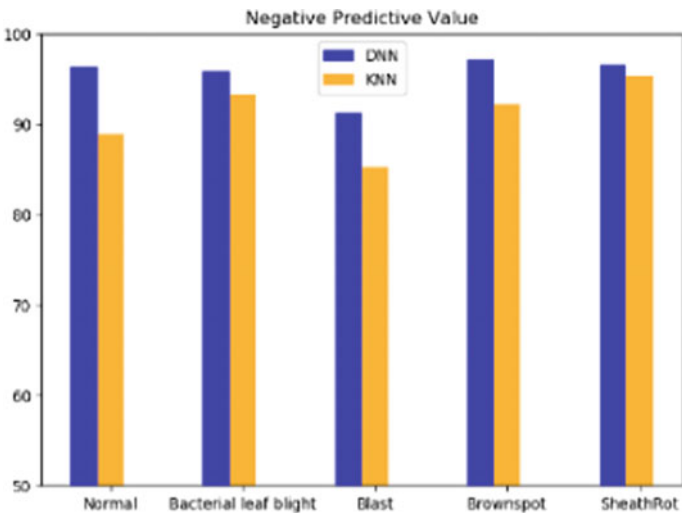
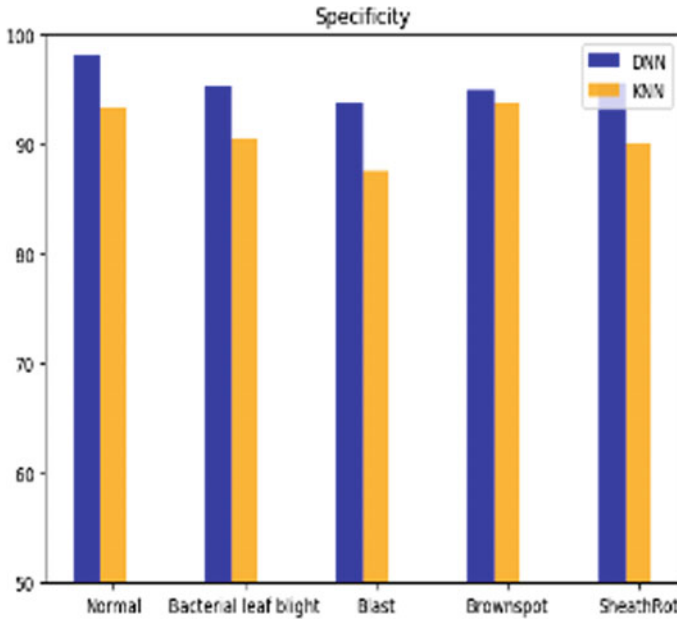


Fig. 10 Negative predict value



**Fig. 11** Specificity

obtained for KNN algorithm are 88% for normal, 92% bacterial leaf blight, 85% blast, 91% brown spot and 93% sheath rot.

#### 4.8 Specificity

Figure 11 shows the graph accuracy for the comparison of both classifiers with different classes. The values obtained in DNN algorithm are 98% for normal, 96% bacterial leaf blight, 95% blast, 96% brown spot and 97% sheath rot. The values obtained for KNN algorithm are 93% for normal, 91% bacterial leaf blight, 88% blast, 94% brown spot and 90% sheath rot.

#### 4.9 True Positive Rate

Figure 12 shows the graph accuracy for the comparison of both classifiers with different classes. The values obtained in DNN algorithm are 86% for normal, 84% bacterial leaf blight, 78% blast, 86% brown spot and 76% sheath rot. The values obtained for KNN algorithm are 58% for normal, 72% bacterial leaf blight, 75% blast, 53% brown spot and 72% sheath rot.

### 4.10 Accuracy

Figure 13 shows the graph accuracy for the comparison of both classifiers with different classes. The values obtained in DNN algorithm are 96% for normal, 93% bacterial leaf blight, 89% blast, 93% brown spot and 92% sheath rot. The values obtained for KNN algorithm are 86% for normal, 88% bacterial leaf blight, 82% blast, 88% brown spot and 87% sheath rot.

## 5 Conclusion

In this research work, the different leaf images are classified by using the proposed algorithms. Images considered for this work are normal, blast, brown spot, bacterial leaf blight and sheath rot. In the image preprocessing, the RGB converted into HSV images to remove the background part. In the segmentation, the diseases and non-diseases leaves are spitted. The features are extracted for colour and texture. DNN algorithm is obtained well in accuracy when compared with the existing KNN algorithm.

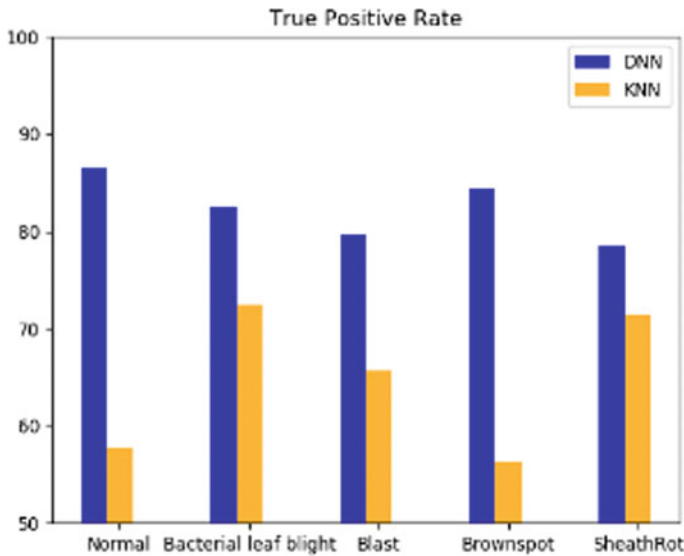


Fig. 12 True positive rate



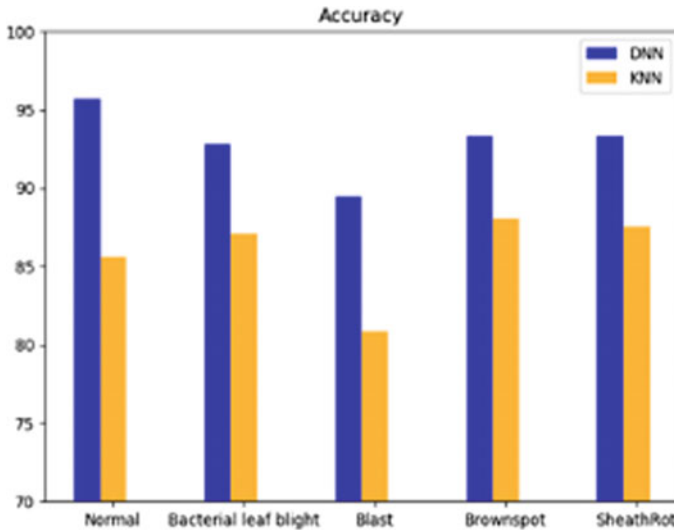


Fig. 13 Overall accuracy

## References

1. Kusumo BS, Heryana A, Mahendra O, Pardede H (2018) Machine learning-based for automatic detection of corn-plant diseases using image processing, pp 93–97
2. Wu N, Li M, Chen L, Yuan Y, Song S (2017) A LDA-based segmentation model for classifying pixels in crop diseased images. In: 2017 36th Chinese control conference (CCC). Dalian, pp 11499–11505
3. Sandika B, Avil S, Sanat S, Srinivasu P (2016) Random forest based classification of diseases in grapes from images captured in uncontrolled environments. In: 2016 IEEE 13th international conference on signal processing (ICSP). Chengdu, pp 1775–1780
4. <https://economictimes.indiatimes.com/news/economy/agriculture/india-set-for-record-foodgrain-output-in-2017-18-agricultureministry/articleshow/63098420.cms>
5. Joshi AA, Jadhav BD (2016) Monitoring and controlling rice diseases using Image processing techniques. In: 2016 international conference on computing, analytics and security trends (CAST). Pune, pp 471–476
6. [https://icar.org.in/files/DAREAnnual%20Report-2017-18\\_\(English\).pdf](https://icar.org.in/files/DAREAnnual%20Report-2017-18_(English).pdf)
7. Pujari JD, Yakkundimath R, Byadgi AS (2015) Image processing based detection of fungal diseases in plants. *Procedia Comput Sci* 46:1802–1808. ISSN 1877-050
8. Yeh YHF, Chung WC, Liao JY, Chung CL, Kuo YF, Lin TT (2013) A comparison of machine learning methods on hyper spectral plant disease assessments. *IFAC Proc* 46(4):361–365. ISSN 1474-6670, ISBN 9783902823304
9. Singh V, Misra AK (2017) Detection of plant leaf diseases using image segmentation and soft computing techniques. *Inf Process Agric* 4(1):41–49
10. [https://support.echoview.com/WebHelp/Windows\\_and\\_Dialog\\_Boxes/Dialog\\_Boxes/Variable\\_properties\\_dialog\\_box/Operator\\_pages/GLCM\\_Texture\\_Features.htm#Energy](https://support.echoview.com/WebHelp/Windows_and_Dialog_Boxes/Dialog_Boxes/Variable_properties_dialog_box/Operator_pages/GLCM_Texture_Features.htm#Energy) (futureequation reference)

# Relative Investigation of Methods to Generate Millimeter Wave in Radio-Over-Fiber Communication



M. Vinoth Kumar and Vinod Kumar

**Abstract** Integration of fiber and wireless (Fi-Wi) technology in latest communication has brought massive improvements in terms of high data rate, low interference and attenuation. The frequency range of 30–300 GHz (millimeter wave) is capable of generating orders of magnitude greater bandwidth. Merging two concepts fiber communication has large bandwidth capacity and wireless coverage are combining both features also became an emerging platform for Internet providers for ever-increasing bandwidth hungry. Bandwidth and channel capacity can be improved by increasing frequency range and implementing wavelength division multiplexing (WDM). Direct and indirect modulated waves imposed with signal generator at transmitter side to generate millimeter (mm) wave. Phase modulator (PM) and Mach–Zehnder modulator (MZM) based techniques of generating mm-wave are investigated and results are shown for various input sine wave generators and different fiber wavelengths.

**Keywords** Radio-over-fiber · Mm-wave · Mach–Zehnder modulator

## 1 Introduction

Before the fifth generation, cellular Internet system with low bandwidth offered minimum data rate connectivity. Later 5G technology started using high bands such as millimeter wave leading to high data rate. Bands in millimeter wave are tremendously high but the propagation is not so decent. Closer cell sites are to be used to the user. Several schemes that enable integrating fiber and wireless that can make system closer to the user. The Fi-Wi network system can create complete usage of the photonics-assisted millimeter wave technology with a modest configuration, low cost and numerous types of multiplexing technologies such as spatial and frequency-band multiplexing [1]. These network systems overcome the bandwidth blockage

---

M. Vinoth Kumar (✉) · V. Kumar  
SRMIST, Ghaziabad, India  
e-mail: [vinothkm@srmist.edu.in](mailto:vinothkm@srmist.edu.in)

V. Kumar  
e-mail: [vinodkur1@srmist.edu.in](mailto:vinodkur1@srmist.edu.in)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_57](https://doi.org/10.1007/978-981-15-2329-8_57)

of electronic devices. The multiplexing methods to drop signal baud rate and bandwidth necessities for optoelectronic devices to upsurge wireless transmission volume impressively. Technology like digital signal processing is being applied to Fi-Wi systems to compensate various linear and nonlinear impairments due to integration fiber and wireless and to realize high spectrum efficiency also high receiver sensitivity.

**High capacity:** High capacity communication system is essential because of extensive usage of Internet, a consistent communication tool in today's society increases high-speed information sharing requirements worldwide every year. Bit rates in GHz range are expected to be in present communication technology to expand the capacity for huge data transfer. Due to enormous systems transmitting data in applications, spectrum in lower gigahertz range is jammed and the need for higher range is observed [2]. For flexible arrangement and easy installation, a radio transmission link is considered to aggregate huge network traffics which have prior characteristics in system deployment. Radio-over-fiber (RoF) technology integrates both optical and wireless network that can provide sufficient bandwidth to individual users [3]. **Limits:** Mobile communication interference is a noticeable factor to be avoided. Fiber chromatic dispersion which is transmission impairments occurs in RoF communication. Optical single sideband (OSSB) is a modulation method that will compensate such dispersion effects. Also, this modulation supports longer distance RoF communication promising.

### *1.1 Millimeter Wave Communication in RoF Technology*

For long distance communication and great capacity millimeter wave delivery, the optical fiber as transmission medium is realized well. Compared to copper cable optical fiber has less weight and low transmission loss that is 0.25 dB/km at 1550 nm (193.41 THz) wavelength range. So by employing optical local oscillator (at 1550 nm range), the mm-carrier wave (33–100 GHz) can be considered with transmission loss of 0.25 dB/km. Insertion loss optical fiber is less (<0.5 dB) than copper because of less weight, i.e., weight of optical fiber is 1.7 kg/km and copper cable is 567 kg/km [4]. Furthermore, the transmission and insertion loss are compensated by optical fiber amplifier. So the transmission can be realized for several kilometers distance.

Deliver of mm-wave through air medium is usually very short distance due to atmospheric and free space path loss. In the finest atmospheric situations, W-band frequency (75–110 GHz) at 80 GHz has atmospheric loss about 0.35 dB/km which limits transmission distance of the terrestrial transmission link, i.e., with in 20 km [5]. There are greater diffraction and penetration losses which create reflected and scattered signals are more significant. Typical penetration losses for building supplies differ from a few decibels to more than 40 dBs [6]. As frequency increases, there is absorption in atmosphere also and there are certain approving bands where atmospheric losses are quite small, i.e., <1 dB/km [5]. Table 1 shows the international telecommunication union proposed wave bands. As stated above, characteristics of

**Table 1** ITU proposed mm-wave bands (in GHz)

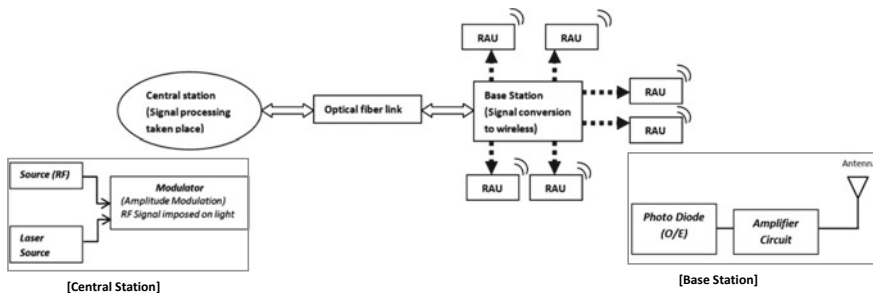
24.25–27.5	31.8–33.4	37–40.5	40.5–42.5	45.5–50.2	50.4–52.6	66–76	81–86
------------	-----------	---------	-----------	-----------	-----------	-------	-------

mm-wave signal in the optical fiber and in the air are different and so enabling methods are required to understand the integration of the optical fiber and the wireless air medium links. For practical implementation of Fi-Wi network and system, simple and nominal cost methods of photonic mm-wave generation techniques is key factors. Numerous techniques have been proposed to realize mm-wave RoF structure. Few techniques to generate mm-wave are based on optical heterodyning, i.e., the heterodyne beating of two different wavelength signals from two independent laser sources, based on optical multi-carrier source, based on external intensity modulators, etc., each technique has its own advantages and limitations. This paper investigates two techniques of generating mm-wave in RoF communication. Firstly, the overview of RoF and MZ modulators is done with diagrammatic representations. Second, the comparison of previously researched works [4, 7–9] has been done. Third simulation of phase modulator and MZM-based mm-wave generation are performed. Results of generated waves are analyzed for different oscillator frequency ranges and for different fiber lengths.

## 2 Radio-Over-Fiber (ROF)

An impressive consideration has been committed to converging radio frequency and optical fiber which advances expectation to build the limit and portability of the access network. Since, it was first exhibited for cordless or cell phone administration in 1990, a great deal of research has been completed to examine its constraints and grow new and superior RoF [1] advancements (Fig. 1).

Radio-over-fiber method eases the growing request for using wireless devices with high data rate transmission [10]. Radio-over-fiber technology comprises a central station and a base station connecting through optical fiber network. Switching, routing



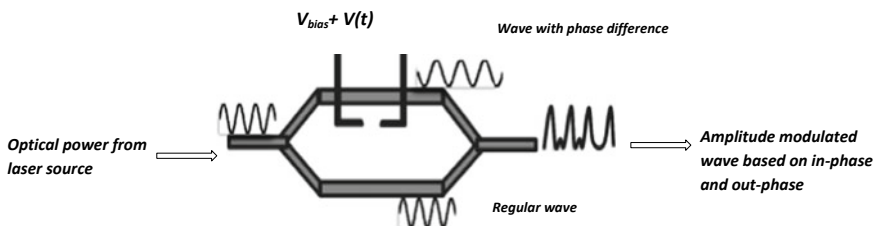
**Fig. 1** Radio-over-fiber link

and MAC, frequency organization functionalities are carried out in central station, whereas base station is for wireless signal distribution and there is no signal processing. Base station converts optical to wireless and vice versa. In radio-over-fiber (RoF) technology, the analog transmission imposed with light is carried. Light is amplitude modulated which carrying information through optical fiber cable to receiver (base station) and wireless access is facilitated at receiver section. Depending on the range of frequency to be transmitted RoF is classified as RF-over-fiber ( $>10$  GHz) and IF-over-fiber ( $<10$  GHz). Wireless signals in RF-over-fiber are optically spread to base stations at high frequencies which effects no requirement of frequency up and down conversion, whereas signal up conversion is performed to RF range at base station after O/E conversion. Based on laser and modulation method RF signals are transmitted over fiber up to many GHz range. Benefits of RoF technology are low attenuation which reduces using more repeaters, less complexity (remote station consisting only O/E conversion and amplification also frequency up/down conversion if require), supporting high data rate and many.

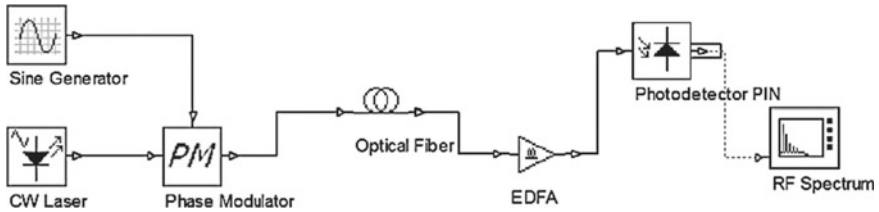
### 3 Mach–Zehnder Modulator

Optical carriers in RoF system is modulated by electrical signal and transmitted over optical path. There is degradation of modulated optical signal due to distortion while traveling which reasons reduction in dynamic range of system. Mach–Zehnder modulator (MZM)—Lithium niobate ( $\text{LiNbO}_3$ ) type is the most preferred modulator also the performance of RoF communication rest on various parameters of MZM such as RF bias, bias voltage, extinction ratio and insertion loss. MZM is used as electrical-optical (EO) converter to achieve linearization or for the compensation of dispersion in signal. MZM single drive modulator is shown in Fig. 2.

The light wave (optical power) entering at input side of modulator is split into two equal waves by two arms of modulator. Phase delay is caused in waves passing through these arms by applied bias across the arms and waves are combined at another end of MZM. In and out phase between two waves make changes in high and low amplitude of light wave at output side of MZM, respectively. The amplitude



**Fig. 2** Mach–Zehnder modulator



**Fig. 3** mm-wave generation using phase modulation

variation in light wave at output of MZM depends on bias applied on arms. Flexibility of using MZM in peripheral location is a great advantage which is not likely with other interferometers.

## 4 Review of Modulation Techniques

Millimeter wave generation by phase modulation technique is shown in the figure. Laser source having 1550 nm wavelength is externally modulated. Sine wave generators used in this technique are set as 20, 30 and 40 GHz. Modulated optical signal is propagated through single-mode fiber (SMF) of 20, 50 and 100 km. Erbium-doped fiber amplifier (EDFA) is used on a fiber link having 5 m length. Light wave is detected by PIN photodetector at receiver side where the generated mm-wave range is recognized [7] (Fig. 3).

Mach-Zehnder modulator based mm-wave generation system shown in the figure. The optical source with 1550 nm wavelength and sine wave generators of 20, 30 and 40 GHz. Light entering MZM is split into two half waves with equal amplitudes. DC bias applied to one of the arms of MZM causes phase delay with propagating wave. Two half waves combining at the output end of MZM are having difference in phase.

So in-phase signals give high power output and out-phase signals give low power output leading to amplitude modulation. Modulated wave is propagating through 20, 50 and 100 km length of SMF and the light wave is converted into electrical at receiver by photodetector.

## 5 Results and Discussion

The setup shown in Figs. 4 and 5 are simulated using optisystem simulation tool. RF spectrum analyzer shows (Figs. 4, 5 are of phase modulation method and Figs. 7, 8 are of MZM method) mm-wave generated from both techniques. Spectrums (power versus frequency) are viewed up to 110 GHz in which it is clear that the peak of power is increasing as input sine wave signal range increases. Also, the signal gets losses in power as the communication channel length increases [7] (Fig. 6).

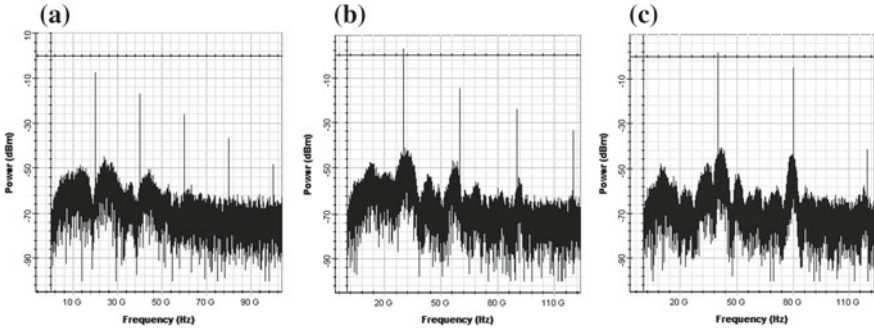


Fig. 4 mm-wave generation with simulation results shown for different input sine wave of a 20 GHz, b 30 GHz, c 40 GHz

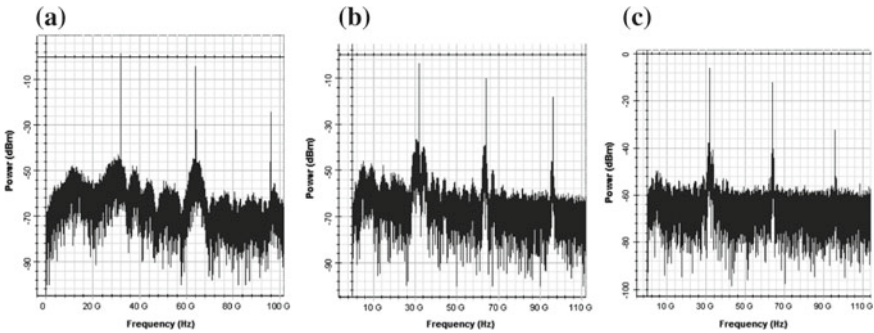


Fig. 5 Wave propagation with simulation results shown for different fiber lengths a 20 km, b 50 km, c 100 km

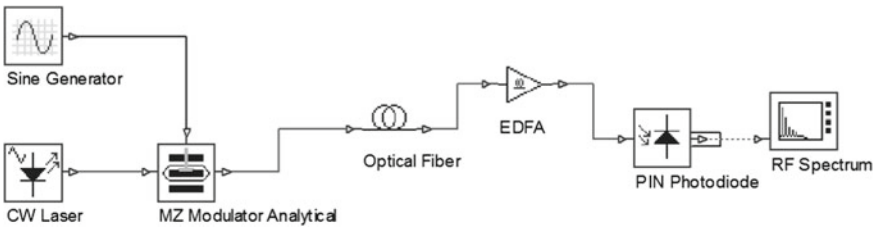
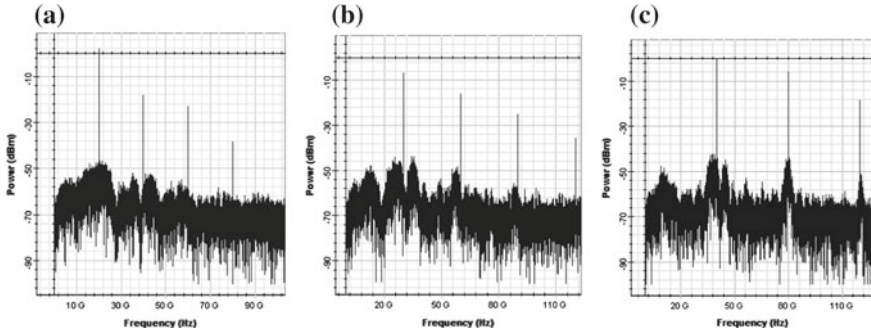


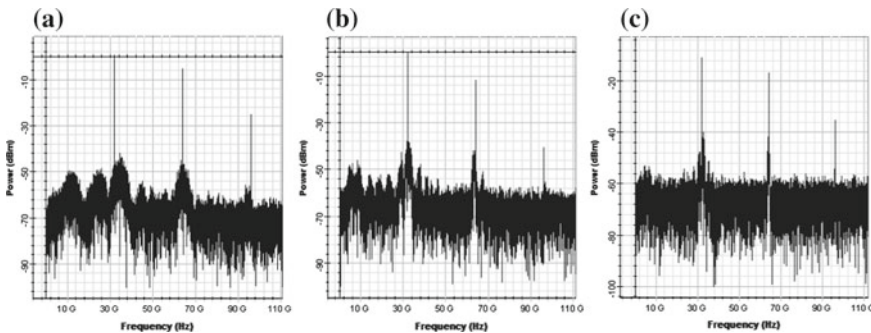
Fig. 6 mm-wave generation using MZM

## 6 Conclusion

High data rate, greater bandwidth for developing Internet applications and mobile communication are achieved by improving operating frequency to millimeter wave range (30–300 GHz). Methods of generating mm-wave by using phase modulator and Mach–Zehnder modulator are investigated for different input oscillator frequency



**Fig. 7** mm-wave generation with simulation results shown for different input sine wave of **a** 20 GHz, **b** 30 GHz, **c** 40 GHz



**Fig. 8** Wave propagation with simulation results shown for different fiber lengths **a** 20 km, **b** 50 km, **c** 100 km

ranges at transmitter section. Powers in dB with respect to frequency up to 110 GHz are analyzed using RF spectrum analyzer. Results are shown satisfactory improvements in output power for higher sine wave range applied to modulator and reduction in output power as communication channel (SMF length) increases.

## References

1. Eghbal MK, Aminian F, Shadaram M (2018) A method to increase the capacity of a millimeter wave radio-over-fiber system. In: 2018 20th international conference on transparent optical networks (ICTON). Bucharest, pp 1–4
2. Yu S, Gu W, Yang A, Jiang T, Wang C (2013) A frequency quadrupling optical mm-wave generation for hybrid fiber-wireless systems. *IEEE J Sel Areas Commun* 31(12):797–803
3. Karthikeyan R, Prakasam S (2013) A survey on radio over fiber for wireless broadband access technologies. *Int J Comput Appl* 64:975–8887
4. Husaini MH, Beson MRC, Anuar MS, Endut R, Aljunid SA (2017) Optical generation of 80 GHz downstream data in radio over fiber system based on stimulated Brillouin scattering. In: MATEC web conference 140 01020



5. Path loss at millimeter wave frequencies. <http://www.raymaps.com/index.php/path-loss-at-millimeter-wave-frequencies/>
6. Rappaport TS, Xing Y, MacCartney GR, Molisch AF, Mellios E, Zhang J (2017) Overview of millimeter wave communications for fifth-generation (5G) wireless networks-with a focus on propagation models. In: IEEE transactions on antennas and propagation, special issue on 5G, Nov 2017
7. Al-Dabbagh RK, Al-Raweshidy HS (2017) 64-GHz millimeter-wave photonic generation with a feasible radio over fiber system. Opt Eng 56(2):026117. <https://doi.org/10.1117/1.OE.56.2.026117>
8. Baskaran M, Prabakaran R (2018) Optical millimeter wave signal generation with frequency 16-tupling using cascaded MZMs and no optical filtering for radio over fiber system. J Eur Opt Society-Rapid Publ 14(13)
9. Yu J, Li X, Zhou W (2018) Tutorial: broadband fiber-wireless integration for 5G + communication. APL Photonics 3:111101. <https://doi.org/10.1063/1.5042364>
10. Li Y, Yang Q, Hemadeh I, El-Hajjar M, Chan C, Hanzo L (2018) Experimental characterization of the radio over fiber aided twin-antenna spatial modulation downlink. Opt Express 26:12432–12440

# A Compact Microstrip Patch Antenna for Mobile Communication Applications



A. Sanega and P. Kumar

**Abstract** In this paper, the design and fabrication of a compact rectangular microstrip antenna (RMSA) are presented. The defected ground structure (DGS) along with *T*-shaped slots in patch is utilized to reduce the size of the antenna. The designed patch antenna with optimized dimensions is fabricated using FR4 substrate. Using the aforementioned methods, the antenna patch size is reduced by 25% when compared to a conventional rectangular microstrip patch antenna. The designed antenna is operating at mobile communication frequency band. The designed antenna is simulated and the simulation is performed using computer simulation technology microwave studio software. The presented antenna is compact in size and suitable for mobile devices.

**Keywords** Microstrip antenna (MSA) · Size reduction · Defected ground structure (DGS) · Reflection coefficient (RC)

## 1 Introduction

The size reduction is a key requirement in modern wireless systems to accommodate more applications in a compact device. The compact size of MSA makes these antennas suitable for modern wireless communications. However, the size of these antennas becomes large at low frequencies [1, 2]. The size of the MSA can be decreased by size reduction techniques such as shorting post loading [3, 4], using slots [5], using metamaterials [6], and DGS [7].

In the shorting post-loading technique, the patch is shorted with the ground. The shorting post behaves like an inductor between the patch and the ground plane. Due to the inductance behavior of shorting post, the resonant frequency of the MSA is shifted and the size of the MSA can be reduced [3]. Mazumdar et al. [8] have proposed slits loaded compact MSA. The presented compact MSA resonates at four different frequencies. In [9], a MSA is designed and analyzed for array application. The

---

A. Sanega · P. Kumar (✉)

EECE, School of Engineering, University of KwaZulu-Natal, Durban 4041, South Africa  
e-mail: [pkumar\\_123@yahoo.com](mailto:pkumar_123@yahoo.com)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_58](https://doi.org/10.1007/978-981-15-2329-8_58)

575

numerical models for calculating resonant frequency, input impedance, and radiation patterns of the MSA are given. In [10], a slot loading technique is used to reduce the size of the MSA. Slots along with conducting strips give multi-frequency operation. In [11], a compact MEMS switches-based reconfigurable MSA is designed. The MSA can radiate in different twelve directions in different states. In [12], slots-based compact MSA is designed. The slots loaded MSA operates at multi-bands. Thabet and Hassan [13] have proposed the design of a compact MSA by changing the substrate material. A nanocomposite material having RT Duriod 5880 with nanofillers is used in the design. Bharti et al. [14] have proposed a compact MSA for triple-band applications. Fractal elements are used in the MSA structure which decreases the resonant frequency of the MSA. Monti et al. [15] have used fractal technique and shorting post to reduce the size of the MSA structure. Using fractal technique along with shorting posts, a compact MSA for RFID applications is designed.

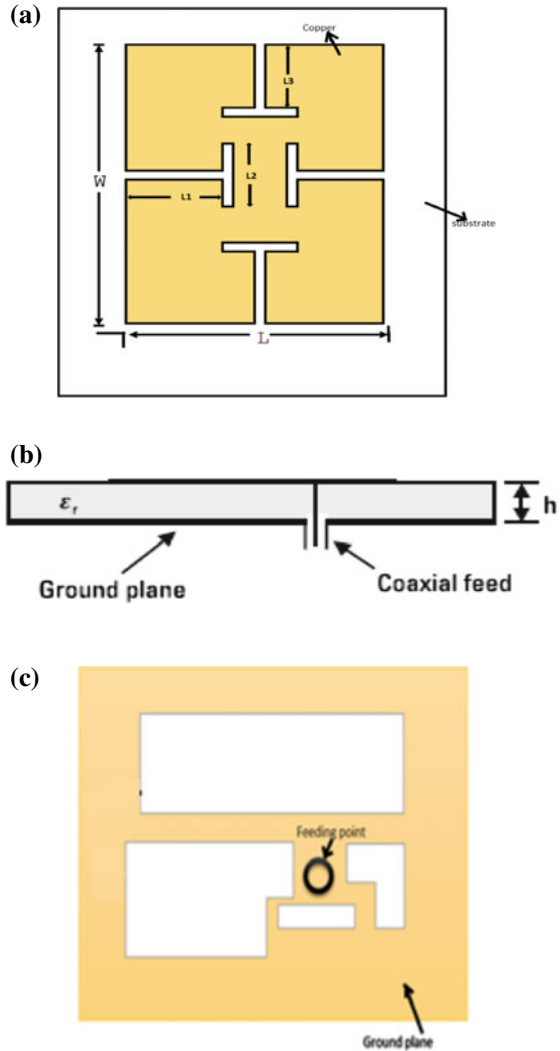
Kumar [16] has designed various shaped compact MSAs for 2.4 GHz applications. The MSAs are designed on a flexible substrate and suitable for wearable applications. Barad and Behara [17] have proposed the design of a compact MSA. The MSA resonates at 2.8 GHz have the beam steering capability. Xiong et al. [18] have presented a compact metamaterial-based MSA for wideband applications. The minimum gain and maximum gain of the designed MSA are 2 dB and 6 dB, respectively. Chatterjee et al. [19] have proposed the design of a compact MSA which is suitable for mobile communication applications. The unequal rectangular slots have been used to reduce the size of the proposed MSA. Wei et al. [20] have proposed a MSA with enhanced bandwidth which is compact in size. The fishnet metasurface is used in the ground plane of the MSA. Gunesar and Seker [21] have proposed the design of a compact MSA which operates in triple bands. *E* shaped and cylindrical slots are used to achieve triple-band operation. The MSA operates at resonant frequencies 1.8, 2.1, and 3.5 GHz, which makes MSA suitable for GSM, UMTS, and WiMAX applications. Anitha et al. [22] have designed a compact MSA for Ku band applications. *U* shaped patch and a slot at center are used in the structure. The MSA is fabricated on FR4 substrate. Alam et al. [23] have designed a compact fractal MSA for 2.4 GHz applications. Roy and Thomas [24] have designed a compact MSA for high-performance wireless network applications. The structure utilizes proximity coupled rectangular slot loaded patch and *V*-shaped slot loaded patch. Kumar and Singh [3] have reviewed the shorting post loaded MSAs and suggested that the size of the MSAs can be reduced by shorting the patch. Singh and Kumawat [25] have proposed a compact *E* shaped wideband MSA for wireless local area networks applications.

In this paper, the concept of DGS along with *T*-shaped slots in the patch [5] is used to decrease the size of the MSA. The design is simulated and optimized using CST microwave studio. The presented MSA is compact and suitable for mobile applications. The rest of the paper is organized in three sections as follows. Section 2 discusses the MSA design and configuration. Simulated results, fabricated MSA, and measured results are given in Sect. 3. Section 4 presents the conclusion of the work.

## 2 Antenna Design and Geometry

Figure 1 shows the geometry of the proposed compact rectangular MSA with DGS and T-shaped slots in the patch. The top and bottom views of the proposed MSA depict the combination of two methods which include the methods of using T-shaped slots [5] in the patch and the use of slots in the ground plane. The size of the MSA is reduced by incorporating T-shaped slots in the microstrip patch as shown in Fig. 1a and incorporating slots in the ground plane as shown in Fig. 1c. Figure 1b shows the side view of the MSA configuration. The MSA is fed by a coaxial connector as

**Fig. 1** Structure of the proposed MSA **a** top view of the MSA, **b** side view of the MSA, **c** bottom view of the MSA



shown in Fig. 1. The MSA is designed using FR4 substrate with dielectric constant of 4.3 and thickness of 1.5 mm. The MSA dimensions are computed by utilizing the transmission line method [1, 2]. The width ( $W$ ) of the patch is calculated using the following expression [1, 2]:

$$W = \frac{c}{2f_0\sqrt{\frac{\epsilon_r+1}{2}}} \quad (1)$$

where ' $c$ ,' ' $f_0$ ,' and ' $\epsilon_r$ ' are velocity of light, resonant frequency of the MSA, and dielectric constant of the substrate, respectively.

The effective dielectric constant ( $\epsilon_{\text{reff}}$ ) of the MSA is computed by [1, 2]:

$$\epsilon_{\text{reff}} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \frac{1}{\sqrt{1 + 12h/w}} \quad (2)$$

where ' $h$ ' is substrate thickness.

Following equation gives the effective length ( $L_{\text{reff}}$ ) [1, 2]:

$$L_{\text{reff}} = \frac{c}{2f_0\sqrt{\epsilon_{\text{reff}}}} \quad (3)$$

The length extension ( $\Delta L$ ) due to fringing field is given by [1, 2]:

$$\Delta L = 0.412h \frac{(\epsilon_{\text{reff}} + 0.3)\left(\frac{W}{h} + 0.264\right)}{(\epsilon_{\text{reff}} - 0.258)\left(\frac{W}{h} + 0.8\right)} \quad (4)$$

The actual length ( $L$ ) of the patch is obtained as [1, 2]:

$$L = L_{\text{reff}} - 2\Delta L \quad (5)$$

The dimensions of the ground plane are given by [1, 2]:

$$L_g = 6h + L \quad (6)$$

$$W_g = 6h + W \quad (7)$$

The coaxial connector location ( $x_f, y_f$ ) is calculated using the following equations [1, 2]:

$$x_f = \frac{L}{2\sqrt{\epsilon_{\text{reff}}}} \quad (8)$$

$$y_f = \frac{W}{2} \quad (9)$$

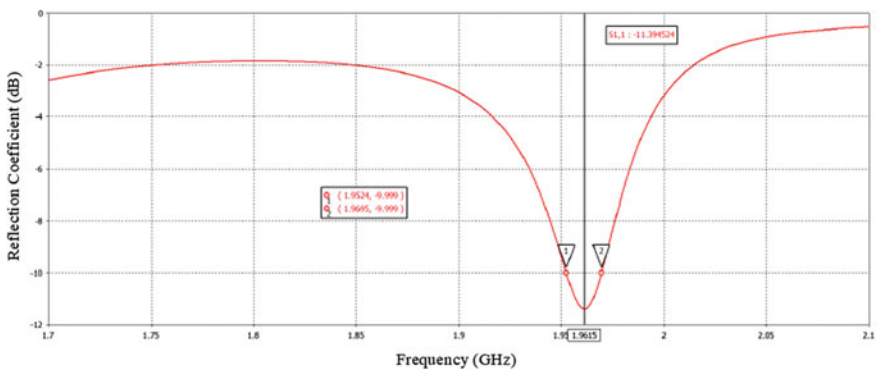
**Table 1** Comparison of size of the designed MSA with conventional MSA

S. No.	Parameter	Dimensions of the conventional MSA (mm)	Dimensions of the proposed compact MSA (mm)
1	Length of patch	36.6	27.5
2	Width of patch	46.7	35
3	Length of ground plane	45.6	34.2
4	Width of ground plane	55.6	41.8

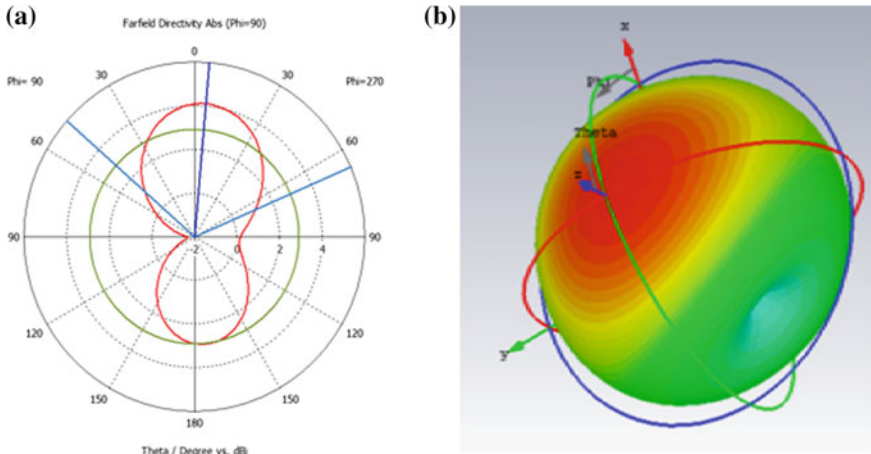
The comparison between the dimensions of the conventional MSA and the MSA with *T*-shaped slots and DGS is shown in Table 1. From Table 1, it can be observed that the size of the proposed MSA is reduced significantly. The size of the patch is reduced by approximately 25%.

### 3 Results and Discussion

For simulations and optimization of the MSA structure, CST microwave studio is used. After carrying the simulations in CST microwave studio, the RC of the proposed MSA is shown in Fig. 2. Figure 2 shows that the RC of the MSA at resonant frequency of the MSA is approximately -11.4 dB. The bandwidth of the MSA is approximately from 1.95 to 1.97 GHz that is approximately 40 MHz. The two-dimensional and three-dimensional radiation patterns of the MSA are depicted in Fig. 3a, b, respectively. The radiation pattern of the MSA in  $\theta = 90^\circ$  plane shows that the shape of the pattern is approximately figure of eight and the main lobe direction is toward  $5^\circ$  with 3 dB angular beam width of  $114.8^\circ$ . From the three-dimensional radiation pattern of the MSA, it is observed that the major lobe of the proposed MSA is toward positive *z*-axis. The maximum gain, maximum directivity, and efficiencies of the designed MSA are



**Fig. 2**  $S_{11}$  of MSA with inserted *T*-slots and DGS



**Fig. 3** Radiation pattern of reduced MSA **a** 2D pattern, **b** 3D pattern

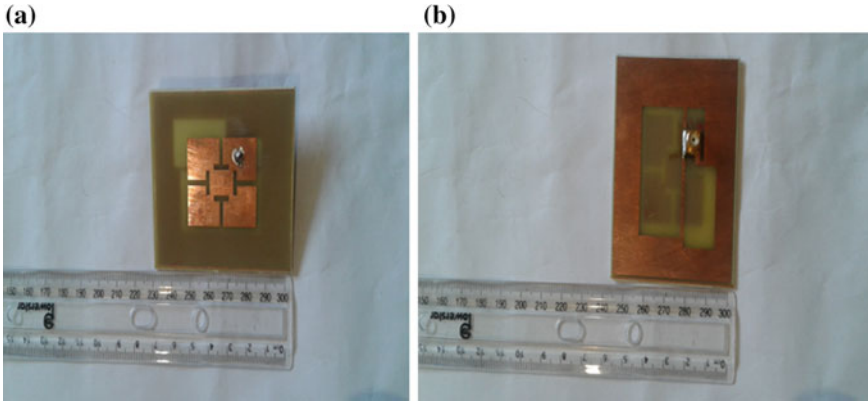
**Table 2** Directivity, gain, radiation efficiency, and total efficiency of the MSA

S. No.	Parameter	Value
1	Max. directivity	4.101 dBi
2	Max. gain	3.675 dB
3	Radiation efficiency	-1.835 dB
4	Total efficiency	-5.662 dB

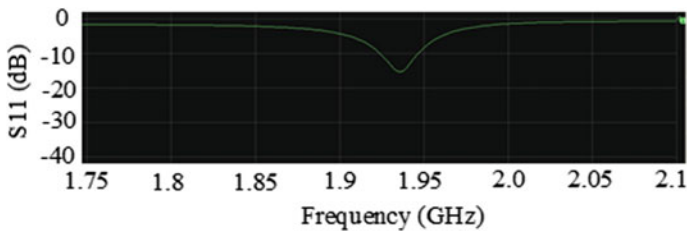
presented in Table 2. The maximum directivity, maximum gain, total efficiency, and radiation efficiency are 4.101 dBi, 3.675 dB, -5.662 dB, and -1.835 dB, respectively. For further validation of the proposed MSA structure, the MSA is fabricated. The fabricated MSA is shown in Fig. 4. The RC of the fabricated MSA is measured using network analyzer and it is shown in Fig. 5. Little deviation between simulated and measured RC can be observed and it may be due to connector losses, fabrication errors, soldering errors, etc. The maximum directivity and maximum gain of the MSA are reasonable good and the RC of the MSA is less than -10 dB. So, the designed and presented MSA is suitable for mobile devices.

### 4 Conclusion

A compact MSA utilizing *T*-slots and DGS has been presented. The design is simulated using CST microwave studio and for further validation of the design, MSA is fabricated and measured. The size of the MSA is reduced by incorporating *T*-shaped slots in the patch and making ground defected. The size of the MSA is reduced and the size of the patch is reduced approximately by 25%. The maximum directivity



**Fig. 4** Fabricated compact MSA



**Fig. 5** Measured  $S_{11}$  for the compact MSA

and maximum gain of the proposed MSA are 4.101 dBi and 3.675 dB, respectively. The proposed MSA is suitable for compact mobile devices.

## References

1. Balanis CA (2005) Antenna theory-analysis and design. Wiley
2. Garg R, Bhartia P, Bahl I, Ittipiboon A (2001) Microstrip antenna design handbook. Artech House Publishers, Boston, London
3. Kumar P, Singh G (2009) Microstrip antennas loaded with shorting post. Engineering 1(1):41–45
4. Kumar P, Singh G (2011) Theoretical investigation of the input impedance of gap-coupled circular microstrip patch antennas loaded with shorting post. J Comput Electron 10(1):195–200
5. Chakraborty M, Rana B, Sarkar PP, Das A (2012) Size reduction of a rectangular microstrip patch antenna with slots and defected ground structure. Int J Electron Eng 4(1):61–64
6. Kulkarni N, Lohiya GB (2016) A compact microstrip patch antenna using metamaterial. Int J Eng Trends Technol 42(7):365–369
7. Pandhare RA, Zade L, Abegaonkar P (2015) Compact microstrip patch antenna array with defected ground structure for wimax and UAV application. Int J Electr Electron Data Commun 3(11):51–54



8. Mazumdar B, Chakraborty U, Bhowmik A, Chowdhury SK, Bhattacharjee AK (2012) A compact microstrip patch antenna for wireless communication. *Global J Res Eng* 12(5):13–16
9. George J, Aanandan CK, Mohanan P, Nair KG (1998) Analysis of a new compact microstrip antenna. *IEEE Trans Antennas Propag* 46(II):1712–1717
10. Roy A, Bhunia S, Sarkar DC, Sarkar PP (2017) Slot loaded compact microstrip patch antenna for dual band operation. *Prog Electromagnet Res C* 73:145–156
11. Ma WD, Wang GM, Wang YW, Zong BF (2017) Compact microstrip antenna with pattern-reconfigurable characteristic. *Radio Engineering* 26(3):662–667
12. Pradeep AS, Laxman K, Arpitha GP, Mahamed AM (2018) Design of compact microstrip patch antenna for multiband operations. *Int J Adv Res Electr Electron Instrum Eng* 7(5):2740–2744
13. Thabet A, Hassan A (2011) Design of compact microstrip patch antenna with and without ground plane slot using new nano-composite materials. *J Eng Sci* 39(6):1375–1385
14. Bhartia G, Bhatia S, Siviak JS (2016) Analysis and design of triple band compact microstrip patch antenna with fractal elements for wireless applications. *Procedia Comput Sci* 85:380–385
15. Monti G, Catarinucci L, Tarricone L (2009) Compact microstrip antenna for RFID applications. *Prog Electromagn Res Lett* 8:191–199
16. Kumar V (2017) Compact patch antenna for 2.4 GHz. In: 2017 4th international conference on electronics and communication systems, pp 1–5
17. Barad D, Behara S (2017) Beam diversity analysis of compact microstrip antenna with suspended superstrate: an experimental study. *Adv Electromagn* 6(3):5–12
18. Xiong H, Hong JS, Tan MT, Li B (2013) Compact microstrip antenna with metamaterial for wideband applications. *Turkish J Electr Eng Comput Sci* 21:2233–2238
19. Chatterjee S, Chakraborty U, Sarkar I, Sarkar PP, Chowdhury SK (2010) A compact microstrip antenna for mobile communication. In: 2010 annual IEEE India conference (INDICON), pp 1–3
20. Wei J, Deng X, Xu X (2018) A compact patch antenna with enhanced bandwidth and efficiency on a fishnet metasurface ground plane in extremely low profile. *Microwave Opt Technol Lett* 60:2748–2753
21. Guneser A, Seker C (2017) Tri-band compact microstrip antenna with multi slots for GSM/UMTS/WIMAX applications. *Int J Adv Comput Eng Netw* 5:1–3
22. Anitha P, Santhosh HV, Reddy AS, Giri Prasad MN (2016) A CPW-fed compact Ku-band microstrip antenna. *Int J Innovative Res Sci Eng Technol* 5:14559–14565
23. Alam S, Surjati I, Ferawan A, Firmansyah T (2018) Design and realization of compact microstrip antenna using fractal sierpenski carpet for wireless fidelity application. *Indonesian J Electr Eng Inf* 6:70–78
24. Roy JS, Thomas M (2007) Compact and broad band microstrip antennas for next generation high-speed wireless communication using HIPERLAN/2. *Int J Microw Sci Technol* 2007:1–4
25. Singh R, Kumawat H (2014) Wideband single patch E-shaped compact microstrip antenna for WLAN. *Int J Adv Eng Res Sci* 1:19–22

# Microstrip Patch Antenna with Enhanced Gain for 2.4 GHz Wireless Local Area Network Applications



N. L. Nhlengethwa and P. Kumar

**Abstract** Wireless communication systems require high gain antennas for many reasons such as for better quality of receiving signals and for low transmitting power. In this paper, a rectangular microstrip antenna (MSA) with enhanced gain for 2.4 GHz wireless local area network applications is presented. The gain of the MSA is increased by using defected ground structure (DGS) technique and a reflected ground plane. The simulation of the designed MSA is performed in CST microwave studio. Using CST microwave studio, the dimensions of the MSA are optimized. The designed MSA is fabricated and measured. Simulated MSA parameters along with and measured results are presented and discussed. The comparison between simulated and measured results shows reasonable agreement. The MSA provides a high directivity of 7.282 dBi and gain of 3.675 dB which is much higher than the MSA with normal ground plane and without reflected ground plane. The designed MSA is suitable for 2.4 GHz wireless local area networks (WLAN) applications.

**Keywords** Microstrip antenna (MSA) · Gain enhancement · Defected ground structure (DGS) · Reflection coefficient (RC)

## 1 Introduction

Long-distance wireless communication systems require high gain transmitting and receiving antennas to compensate the losses due to long distance between transmitter and receiver. MSAs are compact in size and light in weight; hence, these antennas are suitable for compact and lightweight applications. However, these antennas suffer with narrow bandwidth and poor gain [1–4]. The gain of the MSAs can be increased by stacked coupling [5, 6], arrays [1, 2], etc.

In [7], the MSA is loaded with a dielectric resonator for enhancing the impedance bandwidth and efficiency. The gain of 1 dB is achieved by the MSA. In [8], the MSA is

---

N. L. Nhlengethwa · P. Kumar (✉)

Discipline of Electrical, Electronic and Computer Engineering, School of Engineering,  
University of KwaZulu-Natal, Durban 4041, South Africa  
e-mail: [pkumar\\_123@yahoo.com](mailto:pkumar_123@yahoo.com)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_59](https://doi.org/10.1007/978-981-15-2329-8_59)

583

loaded with an engineered magnetic superstrate. The engineered magnetic superstrate enhances the gain and efficiency of the MSA. In [9], the concept of metamaterial is used to enhance the gain and bandwidth of the MSA. The proposed MSA utilizes crisscross metamaterial substrate for enhancing the gain and bandwidth of the MSA. In [10], a ring slot-coupled MSA is presented. The structure utilizes a reflector and dielectric to improve the MSA performance. A S-shaped MSA with wide bandwidth and enhanced gain is proposed in [11]. The photonic band gap (PBG) structure is used to improve the MSA parameters. In [12], a four-element array with high gain is designed for 5 GHz WLAN applications. The slots and reflected surface are used for enhancing the gain of the array.

Kurniawan et al. [13] have designed a high gain MSA. The gain of the MSA is increased by using sub-array technique. The MSA utilizes eight sub-arrays with total 64 elements. The MSA was fabricated on FR4 substrate. The MSA was designed for the frequency range 2.52–2.67 GHz of S-band. Kaya [14] has proposed the design of a high gain rectangular broadband MSA. The chip resistor and negative capacitor are embedded with the designed rectangular broadband MSA. AlyAboul-Dahab et al. [15] have designed a compact MSA with high gain for X-band applications. The frequency selective surface technology is used to enhance the performance of the MSA. Jeyakumar et al. [16] have proposed the design of high gain antenna array for K-band 5G cellular communication. The designed array consists of eight elements. The antenna array is designed on Rogers/RT Duroid substrate. Ou et al. [17] have proposed a MSA with high gain for microwave power transmission. To enhance the efficiency of the MSA, air substrate is utilized. The bandwidth of the MSA is 350 MHz, ranging from 5.67 to 6.02 GHz. Bhide et al. [18] have proposed the design of a dual-band high gain MSA. The design utilizes the union-shaped patch. The designed MSA operates at two resonances: 2.19 and 3.98 GHz. Malakar et al. [19] have proposed the design of a rectangular MSA with high. The design utilizes the concept of air cavity for improving the MSA parameters. Rozana and Hariyadi [20] have proposed the design of a circularly polarized MSA with high gain. The bandwidth of the MSA is 1.366–1.655 GHz, and the MSA is suitable for GPS receiver. In [21], parasitic mushroom-type structure is utilized to improve the parameters of MSA. The designed MSA provides high gain and broadband operation. The frequency selective surface (FSS) along with DGS is used to design a broadband MSA with enhanced gain for wireless local area network applications; industrial, scientific, and medical band applications; worldwide interoperability for microwave access applications in [22].

In this paper, a rectangular MSA with enhanced gain is presented. The proposed structure utilizes the DGS along with a reflected ground structure [12] to enhance the gain. Simulated and measured results are presented. The proposed MSA is suitable for 2.4 GHz wireless local area network applications. The organization of the paper is given as follows. Section 2 presents the design, dimensions, and geometry of the MSA. The simulated and measured MSA parameters are given in Sect. 3. Section 4 gives the conclusion.

## 2 Proposed Antenna Design

The geometrical configuration of the designed rectangular patch antenna with high gain is depicted in Fig. 1. The perspective view of the designed antenna is depicted in Fig. 1a, and the ground plane with DGS is shown in Fig. 1b. The patch is designed using transmission line model analysis. The design procedure follows the various dimension calculations for MSA. The width ( $W$ ) of the metallic patch is calculated as [1]:

$$W = \frac{c}{2f_r} \sqrt{\frac{2}{\epsilon_{rs} + 1}} \tag{1}$$

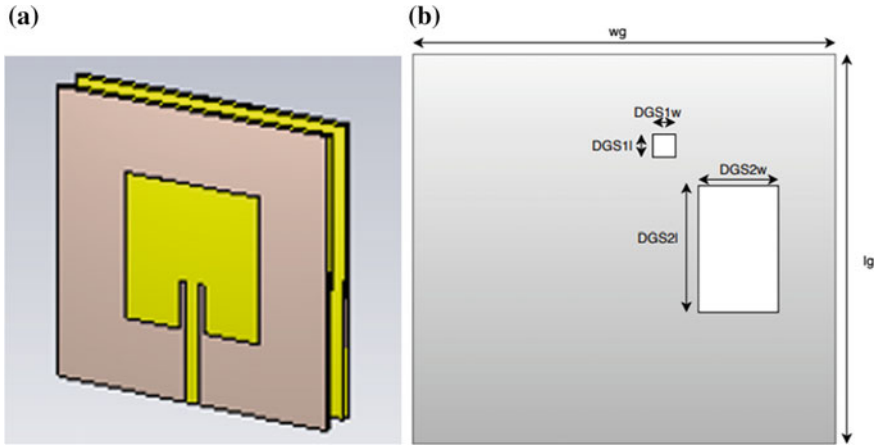
where  $c$ ,  $\epsilon_{rs}$ , and  $f_r$  are velocity of light, relative permittivity (dielectric constant) of the substrate and resonant frequency, respectively.

Considering the fringing field, the effective relative permittivity (effective dielectric constant) ( $\epsilon_{res}$ ) of the substrate is given by [1]:

$$\epsilon_{res} = \frac{\epsilon_{rs} + 1}{2} + \frac{\epsilon_{rs} - 1}{2} \left[ 1 + 12 \frac{h_s}{w} \right]^{-\frac{1}{2}} \tag{2}$$

where  $h_s$  is the substrate thickness.

The fringing effect increases the length of the patch. The increased length ( $\Delta L$ ) of the patch is given by [1]:



**Fig. 1** Structure of the proposed MSA, **a** perspective view, **b** ground plane with DGS

**Table 1** Optimized dimensions of the MSA

S. No.	Parameter	Value (mm)
1	L	28
2	W	28
3	DGS1l	3.33
4	DGS1w	3.33
5	DGS2l	13.33
6	DGS2w	10
7	Thickness of the substrate	1.5
8	Relative permittivity of the substrate	4.3

$$\Delta L = 0.412 \times h_s \times \frac{(\epsilon_{\text{res}} + 0.3) \left( \frac{W}{h_s} + 0.264 \right)}{(\epsilon_{\text{res}} - 0.258) \left( \frac{W}{h_s} + 0.8 \right)} \quad (3)$$

The total length ( $L_e$ ) of the patch is given by [1]:

$$L_e = \frac{c}{2f_r \sqrt{\epsilon_{\text{res}}}} \quad (4)$$

The length of the patch is given by [1]:

$$L = L_e - 2\Delta L \quad (5)$$

The MSA is fed by using a recessed microstrip line as shown in Fig. 1a, and the ground is made defected as shown in Fig. 1b. The defected ground plane along with reflected ground plane [12] is utilized to enhance the gain of the MSA. The thickness and dielectric constant of each substrate layer are  $h$  and  $\epsilon_{\text{rs}}$ , respectively. The proposed structure is optimized using CST microwave studio. The MSA dimensions are given in Table 1. The length/width of the patch is 28 mm. The length/width of smaller slot in ground is 3.33 mm. The length of the bigger slot in ground is 13.33 mm, and width of the bigger slot in ground is 10 mm. The substrate height is 1.5 mm, and the relative permittivity (dielectric constant) of the substrate is 4.3.

### 3 Results and Discussion

This section presents the parameters of the designed MSA obtained from the simulation and measurement. The model is simulated to find the RC and the far-field patterns at a certain frequency (2.4 GHz) using the CST tools. From the far-field values, the gain and directivity are considered. The results of the designed MSA are shown in Figs. 2, 3, 4, 6 and Table 2. The RC of the MSA for the frequency range of

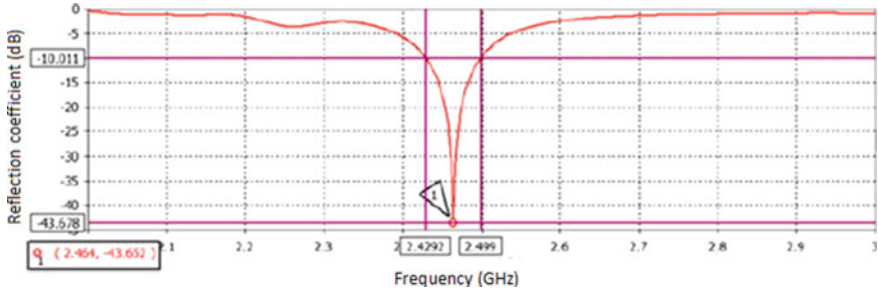


Fig. 2 RC versus frequency

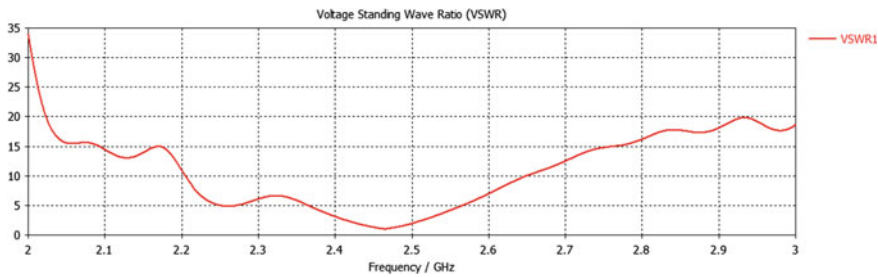
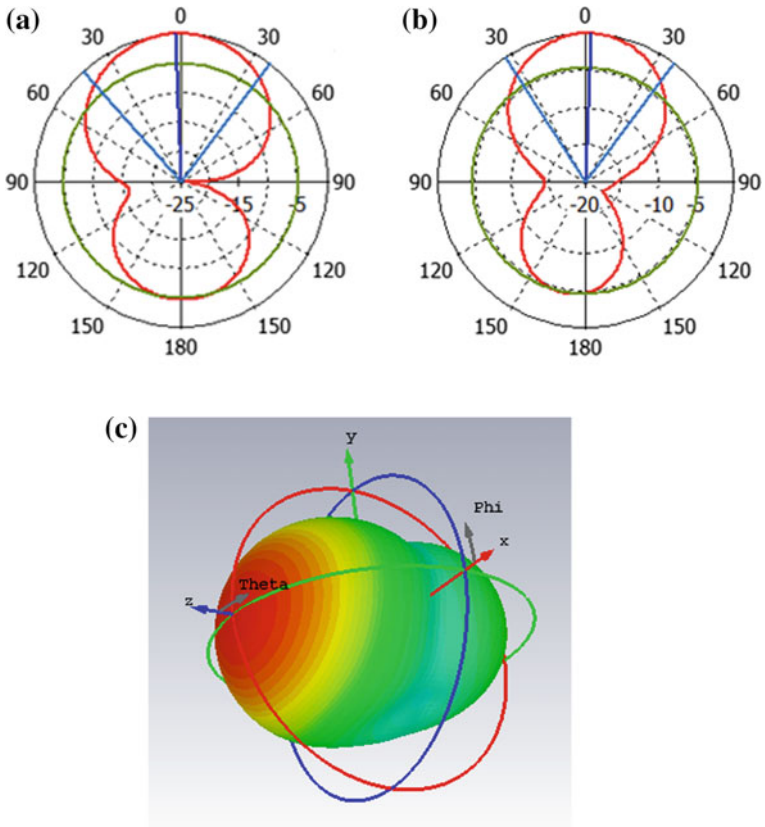


Fig. 3 VSWR of the designed MSA

2–3 GHz is depicted in Fig. 2. The RC of the MSA versus frequency graph depicts that the MSA is operational in the frequency range of 2.417–2.488 GHz. This is obtained from the point where the RC ( $S_{11}$  parameter) is below 10 dB. The MSA resonating frequency is approximately 2.452 GHz. The variation of voltage standing wave ratio with frequency is shown in Fig. 3. From Fig. 3, it can be observed that the voltage standing wave ratio (VSWR) is less than 2 for the bandwidth of the designed MSA.

The radiation patterns of the MSA are depicted in Fig. 4. The normalized pattern in  $\varphi = 0^\circ$  plane is shown in Fig. 4a; the normalized pattern in  $\varphi = 90^\circ$  plane is shown in Fig. 4b; and the absolute three-dimensional radiation pattern is shown in Fig. 4c. From radiation patterns of the MSA, it can be observed that the shape of the patterns is figure of eight and the behavior of the patterns is similar in  $\varphi = 0^\circ$  plane and  $\varphi = 90^\circ$  plane. The parameters of the patterns of the MSA are given in Table 2. The maximum directivity of the MSA is high and is 7.282 dBi. The gain of the MSA with DGS and reflected ground plane is 3.675 dB, which is much higher as compared to the MSA without reflected ground plane and DGS. The gain of the MSA without reflected ground plane and DGS is 1.115 dB only. Hence, the gain of the MSA is improved significantly. The total efficiency and radiation efficiency of the MSA at 2.4 GHz are  $-4.951$  dB and  $-3.607$  dB, respectively. The direction of the main lobe in  $\varphi = 0^\circ$  plane and  $\varphi = 90^\circ$  plane is same and is at two degree. The 3 dB beamwidth in  $\varphi = 0^\circ$  plane and  $\varphi = 90^\circ$  plane is  $78.8^\circ$  and  $69.2^\circ$ , respectively. The designed



**Fig. 4** Radiation patterns of the designed MSA at 2.4 GHz, **a**  $\phi = 0^\circ$  plane, **b**  $\phi = 90^\circ$  plane, **c** 3-dB

**Table 2** MSA radiation parameters at 2.4 GHz

S. No.	Parameter	Value
1	Directivity	7.282 dBi
2	Gain	3.675 dB
3	Total efficiency	-4.951 dB
4	Radiation efficiency	-3.607 dB
5	Main lobe direction ( $\phi = 0^\circ$ plane)	$2.0^\circ$
6	3 dB beamwidth ( $\phi = 0^\circ$ plane)	$78.8^\circ$
7	Main lobe direction ( $\phi = 90^\circ$ plane)	$2.0^\circ$
8	3 dB beamwidth ( $\phi = 90^\circ$ plane)	$69.2^\circ$

MSA is fabricated and measured. The photographs of fabricated MSA and measured RC parameter of the fabricated MSA are shown in Figs. 5 and 6, respectively. The measured RC of the MSA is similar to the simulated RC except little for frequency shift. It may be due to imperfect fabrication and losses. The MSA provides high directivity and suitable for 2.4 GHz wireless local area network application.

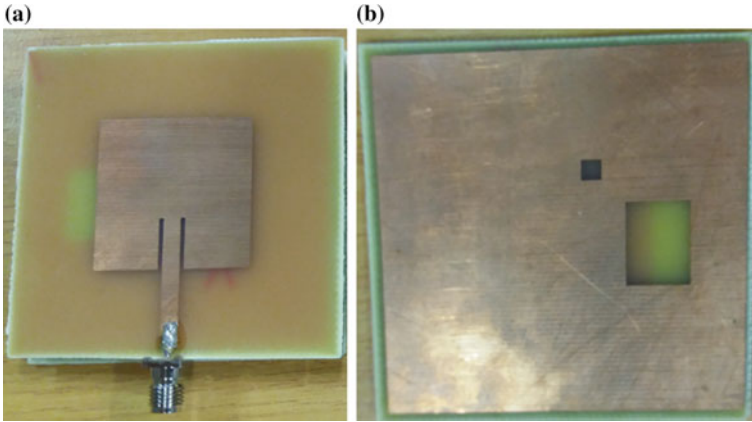


Fig. 5 Fabricated MSA, a patch view, b ground plane with DGS

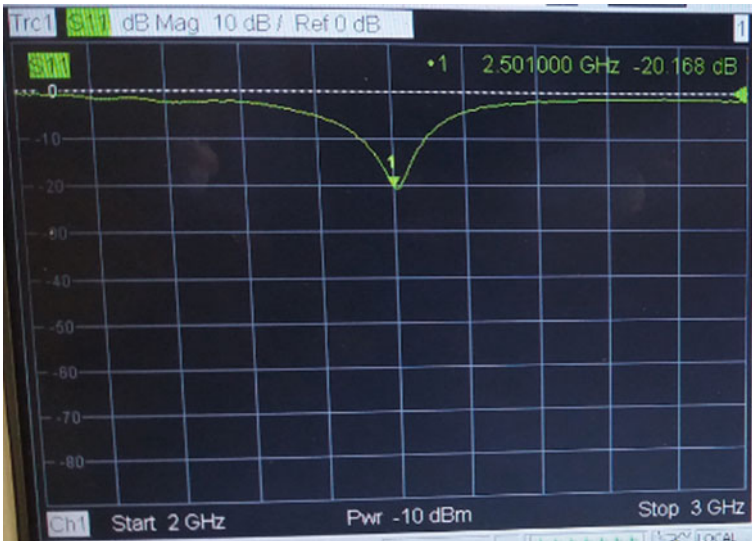


Fig. 6 Measured RC of the designed MSA



## 4 Conclusion

A rectangular MSA with enhanced gain has been designed, simulated, and fabricated. The DGS and a reflected ground plane are used to enhance the gain of the designed MSA. Simulated as well as measured MSA parameters have been presented. The gain of the MSA with the reflected ground plane and the DGS is much higher than the MSA without the reflected ground plane and the DGS. The designed MSA with enhanced gain is suitable for 2.4 GHz wireless local area network applications.

## References

1. Balanis CA (2005) *Antenna theory-analysis and design*. Wiley (2005)
2. Garg R, Bhartia P, Bahl I, Ittipiboon A (2001) *Microstrip antenna design handbook*. Artech House Publishers, Boston, London
3. Kumar P (2017) Design of low cross-polarized patch antenna for ultra-wideband applications. *Int J Comm Antenna Prop* 7:265–270
4. Mabaso M, Kumar P (2018) A dual band patch antenna for bluetooth and wireless local area networks applications. *Int J Microw Opt Technol* 13:393–400
5. Ghassemi N, Rashed-Mohassel J, Neshati MH, Tavakoli S, Ghassemi M (2008) A high gain dual stacked aperture coupled microstrip antenna for wideband applications. *Prog Electromagn Res B* 9:127–135
6. Sethi WT, Vettikalladi H, Alkanhal MA (2014) High gain stacked antenna array for 60 GHz communication systems. In: *Proceedings of IEEE antennas and propagation society international symposium*, pp 1734–1735
7. George J, Aanandan CK, Mohanan P, Nair KG, Sreemoolanathan H, Sebastian MT (1998) Dielectric-resonator loaded microstrip antenna for enhanced impedance bandwidth and efficiency. *Microw Opt Technol Lett* 17:205–207
8. Attia H, Yousefi L, Bait-Suwailam MM, Boybay MS, Ramahi OM (2009) Enhanced-gain microstrip antenna using engineered magnetic superstrates. *IEEE Antennas Wirel Propag Lett* 8:1198–1201
9. Inamdar K, Kosta YP, Patnaik S (2015) Criss-cross metamaterial-substrate microstrip antenna with enhanced gain and bandwidth. *Radioelectronics Commun Syst* 58:69–74
10. Wang NZ, Talbi L, Zeng QS, Xu JD (2013) Wideband ring-slot coupled patch antenna with enhanced gain. *J Electromagn Waves Appl* 27:572–581
11. Niboriya BS, Choudhary C, Prabhakar G (2013) S-shape wideband microstrip patch antenna with enhanced gain and bandwidth for wireless communication. *Int J Comput Appl* 73:17–20
12. Ngobese BW, Kumar P (2018) A high gain microstrip patch array for 5 GHz WLAN applications. *Adv Electromagn* 7:93–98
13. Kurniawan A, Kurniawan DWH, Atqiya A (2018) High-gain microstrip-antenna design using sub-array method. *Proc World Congr Eng* 1:1–4
14. Kaya A (2008) High gain rectangular broad band microstrip antenna with embedded negative capacitor and chip resistor. *Prog Electromagn Res PIER* 78:421–436
15. AlyAboul-Dahab M, Ghouz HHM, Zaki AZA (2016) High gain compact microstrip patch antenna for X-band applications. *Int J Antennas JANT* 2:47–58
16. Jeyakumar P, Chitra P, Christina G (2018) Design and simulation of directive high gain microstrip array antenna for 5G cellular communication. *Asian J Appl Sci Technol* 2:301–313
17. Ou J, Andrenko AS, Fu C, Xie Z, Tan H (2016) High-gain microstrip antenna for microwave power transmission. In: *Proceedings of ISAP 2016*, pp 740–741

18. Bhide G, Nandgaonkar A, Nalbalwar S (2017) Dual band high gain union shaped microstrip patch antenna. *Adv Intell Syst Res* 137:786–791
19. Malakar K, Nandi J, Mitra S, Gorai PK, Chattopadhyay S, Banerjee S (2011) Rectangular microstrip antenna with air cavity for high gain and improved front to back ratio. *J Electromagn Anal Appl* 3:368–372
20. Rozana A, Hariyadi T (2018) Design of a high gain circularly polarized microstrip antenna for GPS receiver. *IOP Conf Ser Mater Sci Eng* 384:1–8
21. Cao Y, Cai Y, Cao W, Xi B, Qian Z, Wu T, Zhu L (2019) Broadband and high-gain microstrip patch antenna loaded with parasitic mushroom-type structure. *IEEE Antennas Wirel Propag Lett* 18:1405–1409
22. Mondal K, Sarkar DC, Sarkar PP (2019)  $5 \times 5$  matrix patch type frequency selective surface based miniaturized enhanced gain broadband microstrip antenna for WLAN/WiMAX/ISM band applications. *Prog Electromagn Res C* 89:207–219

# An Investigation on Drain Current of Junction and Junctionless Surrounding Gate MOSFET



Aditya Agarwal, R. L. Sharma and Prashant Mani

**Abstract** This paper presents investigation about the drain current parameters of Surrounding Gate MOSFET (SG MOSFET) with junction and junctionless transistor. The junctionless SG MOSFET (JLSG MOSFET) exhibits more current available at low voltage but junction based SG MOSFET exhibits less current at same voltage but depends on parameters. The junction based devices are less costly. Device length also improved in JLSG MOSFET. General variable issues like oxide thickness, channel length and doping concentration are also discussed.

**Keywords** Surrounding gate MOSFET · Drain current · Junctionless

## 1 Introduction

Semiconductor is the foundation of today's communication systems like cellular telephone, navigation, wireless systems and radio broadcast etc. Surrounding Gate MOSFETs are become familiar and famous due to several advantages like excellent control on the channel surface potential, less space consumption, suppressed short channel effects etc. In junction based SG MOSFET consist of sharp drain and source junctions because of this the doping variation concentration over with some nanometers closer to the source and drain junctures. This is preserving limitations on doping methodologies. Due to presence of junctions, there is leakage current which run within the device [1]. To reduces Short channel effects, carrier's effective mobility and increase gate control we take the advantage of SG MOSFET. In SGMOSFET channel is surrounded by Gate keep all sides which provide better manage over with the channel. It requires very less area, and less area consumption means less expense and to a greater extent number of devices per chip [1].

---

A. Agarwal (✉) · R. L. Sharma  
Department of Electronics and Communication Engineering, Noida International University,  
Gautam Budh Nagar, Noida, UP, India

P. Mani  
Department of Electronics and Communication Engineering, SRM Institute of Science and  
Technology, NCR Campus, Modinagar, UP, India

JLSG MOSFET exhibits superior scaling capability and reduced Short channel effects (SCEs) than SG MOSFET. Now a days transistor requires small in size and very less SCEs. When the transistor dimension decreases up to nano scale depth the formation of precise junction become complicated that's why it demands ultimate exact doping and temperature conditions. Now for time to come, VLSI practical applications need basic transistor level changes required. JLSG MOSFET is the substitute of junction based SG MOSFET [2]. Thus a novel structure investigated which have not need the formation of p n junction, known as junctionless device [3]. The junctionless transistors are highly doped through the drain, source and channel area. So there is no generation of junctions and these will remove the issue of diffusion of impurities, provides better short channel immunity and reduced SCEs [2]. Junctionless transistors initially in accumulation mode which will less responsive for negative bias thermal effect. So the drain current is less instable with doping concentration. In junctionless transistor, the capacitances can also be neglected which is present in junction based transistor. Therefore the capacitance goes down due to the operating frequency and speed of the transistor can be increased [1, 4-9].

IC projects design operation with JLSG MOSFET device technology will be faster, high energy efficient. The Surrounding gate device has high packing density, immune to SCEs and high current drive capability [3].

## 2 Device Structure

The illustrative 2D view and 3D view of SG MOSFET shown in Figs. 1 and 2 respectively. The gate oxide thickness, gate length and radius of channel are 3 nm, 15 nm,

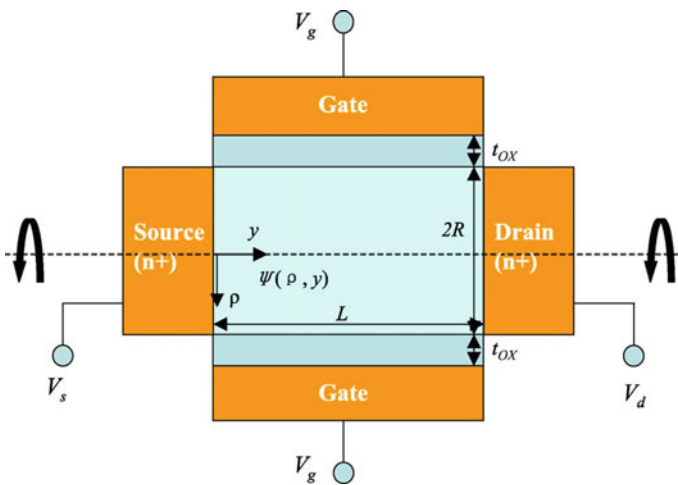
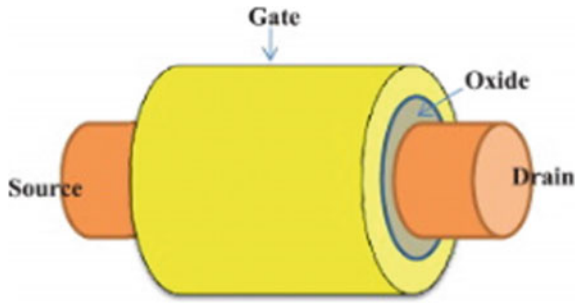


Fig. 1 The 2D view of SG MOSFET



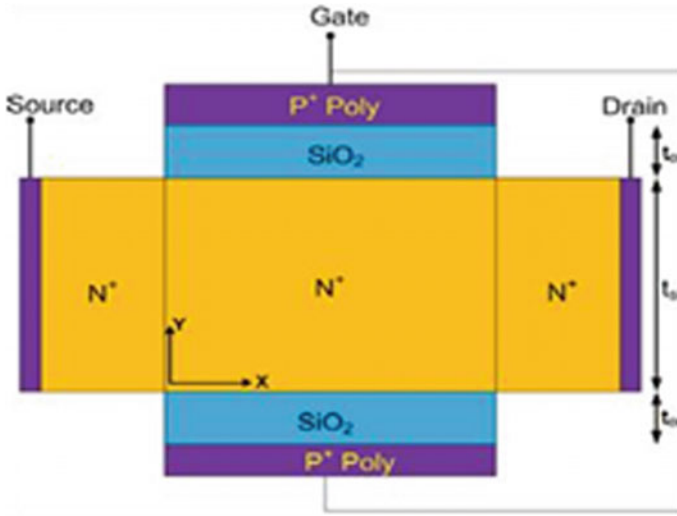
**Fig. 2** The 3D view of SG MOSFET

40 nm respectively. SG MOSFET has advantages like improve transconductance, enhanced current and good control of short channel effects due to shortest channel length [10]. SG MOSFET required smaller voltage means that gate electrodes have higher command over the conducting Channel [11]. As channel width is reduced the threshold voltage reduces due to which the improved current possibility achieved within the channel. SG transistor is solely the cylindrical part of the transistor [12]. The cylindrical Structure specially reduces corner effects and short channel effects [13] (Table 1).

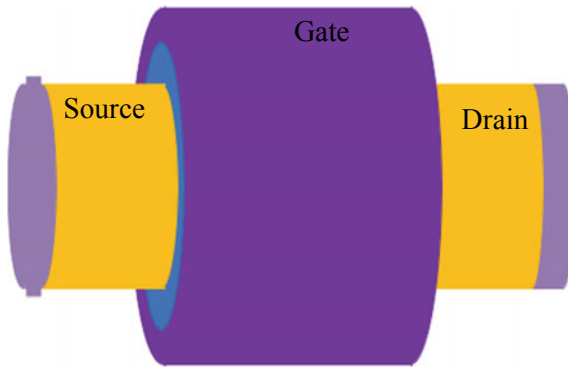
The interpretive 2D and 3D view of JLSG MOSFET shown in Figs. 3 and 4 respectively. The thickness of gate oxide, gate length and channel radius are 2 nm, 5 nm, 20 nm respectively. This is a symmetrical structure around the axis which helps channel passing through with origin. Consequently the symmetrical structure, electric field and electrostatic potential is not varying in the angular direction. JLSG MOSFET

**Table 1** Device parameters of SG MOSFET

Symbol	Quantity	Value
$t_{ox}$	Gate oxide thickness	3 nm
$N_D$	Doping concentration	$5 \times 10^{19} \text{ Cm}^3$
$R$	Radius of channel	15 nm
$L$	Channel length	40 nm
$W$	Gate work function	5.2 eV



**Fig. 3** The 2D view of JLSG MOSFET



**Fig. 4** The 3D view of JLSG MOSFET

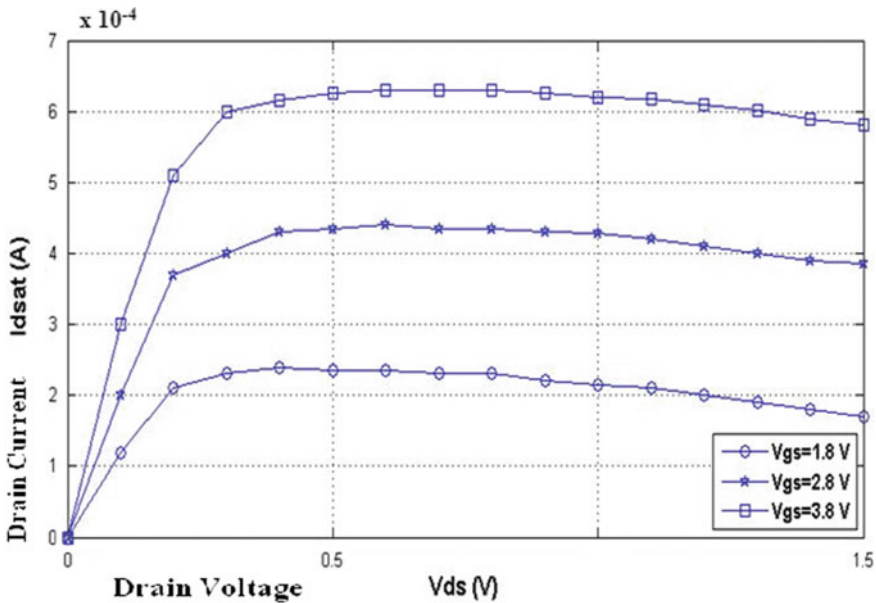
has excellent short channel immunity so more variation in electrical parameters occurs [14]. The drain current found higher in JLSG MOSFET cause of less carrier mobility degradation. In JLSG transistor has more doping concentration around the channel, source and drain areas which considerably improves fabrication processes and exhibits improve transistor electrical behavior throughout small subthreshold swing and very large ON state current [14, 15] (Table 2).

**Table 2** Device parameters of JLSG MOSFET

Symbol	Quantity	Value
$t_{ox}$	Gate oxide thickness	2 nm
$N_D$	Doping concentration	$1 \times 10^{19} \text{ Cm}^{-3}$
R	Radius of channel	5 nm
$L$	Channel length	20 nm
$W$	Gate work function	5.1 eV

### 3 Result and Discussions

In Fig. 5, shows Drain Current of junction based Surrounding Gate MOSFET, Fig. 6, Drain current of Junctionless Surrounding Gate MOSFET and in Fig. 7, Comparison of Drain current for Surrounding Gate MOSFET junction based and Junctionless Surrounding Gate MOSFET with different parameters. This model provides constant carrier mobility. We can observe that short channel immunity and reliability of long channel is main regard for highly scaled devices. JLSG MOSFET device will provide better results which indicate favorable results for future VLSI devices.



**Fig. 5** Surrounding gate MOSFET drain current ( $I_D$ ) versus drain to source voltage ( $V_{DS}$ )

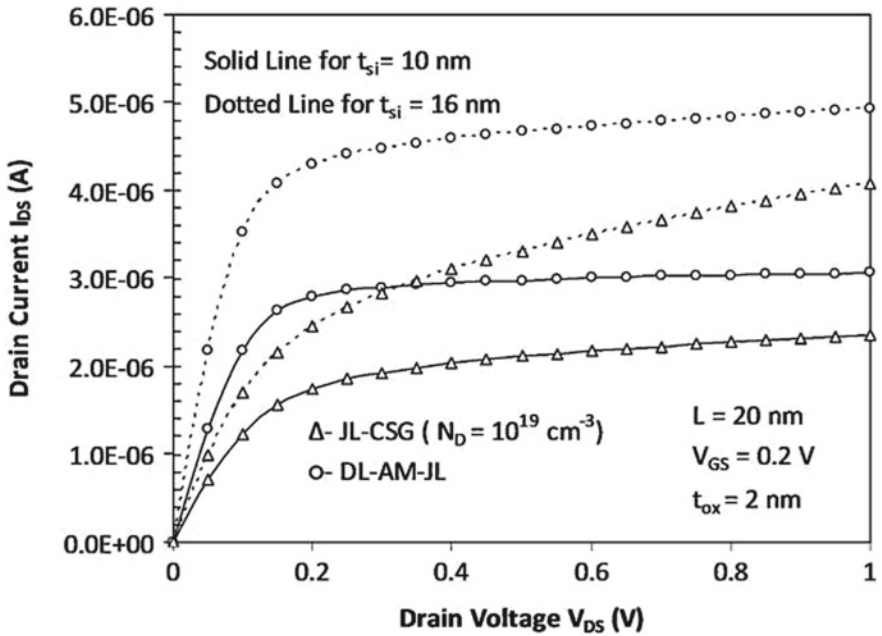


Fig. 6 Junctionless surrounding gate MOSFET drain current ( $I_D$ ) versus drain to source voltage ( $V_{DS}$ )

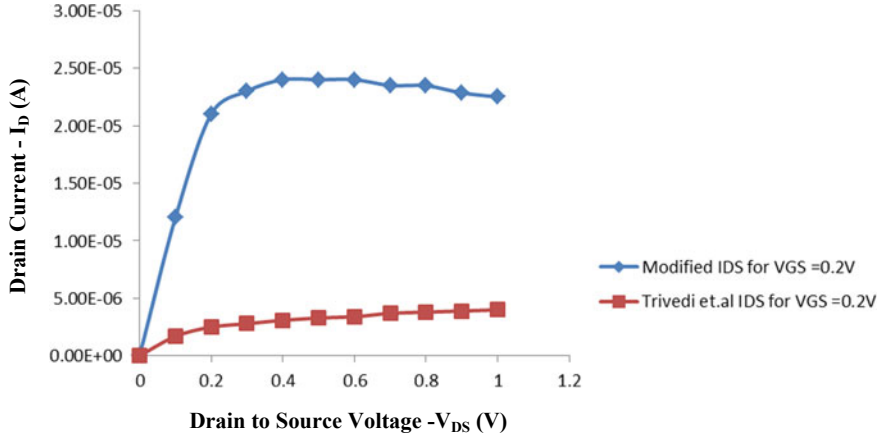


Fig. 7 Junctionless surrounding gate MOSFET and surrounding gate MOSFET, drain current ( $I_D$ ) versus drain to source voltage ( $V_{DS}$ ) when  $V_{GS} = 0.2$  V



## References

1. Srivastava MA, Kumar A, Rai S (2017) Analytical model and performance investigation of electric potential for junctionless cylindrical surround gate (JLCSG) MOSFET. In: 4th international conference signal processing and integrated networks 2017, pp 256–261. IEEE
2. Pratap Y, Halder S, Gupta R, Gupta M (2014) Performance evolution and reliability issues of junctionless CSG MOSFET for RFIC Design. *IEEE Trans Device Mater Reliab* 14(1):418–425
3. Trivedi N, Kumar M, Halder S, Deswal S, Gupta M, Gupta R (2017) Charge plasma technique based dopingless accumulation mode junctionless cylindrical surrounding gate MOSFET: analog performance improvement. *Appl Phys A Mater Sci Process Cross Mark* 123:564. Springer
4. Duarte J, Choi S, Moon D, Choi Y (2011) Sensitivity of threshold voltage to nanowire width variation in junctionless transistor. *IEEE Electron Device Lett* 32(2):125–127
5. Ansari L, Feldman B, Fags G, Colinge J, Career J (2010) Simulation of junctionless si nanowire transistors with 3 nm gate length. *Appl Phys Lett* 97(6): 062105-1–062105-3
6. Gnudi A, Reggiani S, Gnani E, Baccarani G (2012) Analysis of threshold voltage variability due to random dopant fluctuations in junctionless FETs. *IEEE Electron Device Lett* 33(3):3336–3338
7. Auth C, Plummer J (1997) Scaling theory for cylindrical, fully depleted, surrounding gate MOSFET's. *IEEE Electro Device Lett.* 18(2):74–76
8. Sallese J, Chevillon N, Lallement C, Iniguez B, Pregaldiny F (2011) Charge based modeling of junctionless double gate field effect transistors. *IEEE Electron Devices* 58(8):2628–2637
9. Gnani A, Gnudi A, Reggiani S, Baccarani G (2011) Theory of junctionless nanowire FET. *IEEE Trans Electron Device* 58(9):2903–2910
10. Agarwal A, Sharma R, Mani P (2018) Analytical modeling electrical conduction of nano scaled surrounding gate MOSFET. In: International conference on micro-electronics and telecommunication engineering (ICMETE) 2018, pp 331–334. IEEE, Ghaziabad
11. Agarwal A, Sharma R, Mani P (2018) A study of characterization of surrounding gate MOSFET. *VSRD Int J Techn NON Tech Res* 9(1):67–70
12. Christopher P, James P (1998) A simple model for threshold voltage of surrounding gate MOSFET's. *IEEE Trans Electron Devices* 45(11):2381–2383
13. Srivastava V, Yadav K, Singh G (2013) Explicit model of cylindrical surrounding gate MOSFET. *WSEAS Trans Circ Syst* 12(3):81–90
14. Colinge JP, Lee CW, Afzalian A (2010) Nanowire transistors without junctions. *Nat Nanotechnol* 5(3):225–229
15. Guangxi H, Xiang P, Ding Z, Liu R, Wang L, Tang T (2014) Analytical models for electric potential, threshold voltage and subthreshold swing of junctionless surrounding gate transistors. *IEEE Trans Electron Devices* 61(3):688–694

# Improving the Performance of Video Content Genuineness Using Convolution Neural Network



Bharat Gupta, Vasvi Bajaj, Rajat Bhusan Panda and Lalit Garg

**Abstract** Video searching in search engines uses metadata information to find the relevant videos according to the search queries. Metadata information mainly comprises the title and description of the video. The major drawback of this approach is that it overlooks whether or not the content of the video is genuine or not. Since the metadata information is provided by the uploader, the person may provide false information about it. Therefore, there is a need of improving the results of video searched. The proposed work classifies the video in different categories and then compares the tag provided to each video with the tags that were extracted from the metadata of the video. The other factor like views count, likes and dislikes, comments is also considered for the ranking of the video searched. It improves the genuineness of the content of the video searched.

**Keywords** Convolution neural network · Deep neural network · Image classification · Video classification

## 1 Introduction

Human brain is a unique entity. It registers and processes information within split seconds. However, these remarkable skills of a human brain are difficult to reciprocate with machines. Machines are more complex than anyone could imagine. Several

---

B. Gupta (✉) · V. Bajaj · R. B. Panda  
Department of Computer Science and Engineering, Jaypee Institute of Information Technology,  
Noida, India  
e-mail: [bharat.gupta@jiit.ac.in](mailto:bharat.gupta@jiit.ac.in)

V. Bajaj  
e-mail: [vasvibajaj@gmail.com](mailto:vasvibajaj@gmail.com)

R. B. Panda  
e-mail: [rajatpandajit8@gmail.com](mailto:rajatpandajit8@gmail.com)

L. Garg  
Faculty of Information & Communication Technology, University of Malta, Msida, Malta  
e-mail: [lalitgarg@um.edu.mt](mailto:lalitgarg@um.edu.mt)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_61](https://doi.org/10.1007/978-981-15-2329-8_61)

researches have been carried out to reach that level of accuracy of machines that is there with the human brain. There is a lot of video data available on Internet. Pictures and recordings have turned out to be inescapable on the web, advancing the advancement of calculations that can examine their semantic substance for various applications, including pursuit and synopsis.

Videos are a grid medium to understand a scenario or gain knowledge, but they too have drawbacks. Like the video suggestion given by the search, result may not be the exact one which is desired, and it may contain some unwanted things which may cause misunderstanding of the given topic or cause a distraction from the searched on. Other problem faced, due to the high volume of the video data, is the presence of unwanted clickbait videos. Video searching on popular search engines utilizes the meta-information to find the relevant videos according to the search queries. The major drawback of the approach is that it overlooks whether or not the content of the video is genuine or not. Since the meta-information used to generate a search result is provided by the uploader, the uploader could easily provide false meta-information to it that leads to the creation of click baits.

According to the Oxford dictionary “clickbait is the content on the Internet with the main purpose of attracting attention and encouraging visitors to click on a link to a particular website”. Huge amount of clickbait videos present on the Internet which attracts user by giving flashy names and thumbnails but have a video with either no relevance or just a blank screen in the video. Since YouTube videos are commercialized, i.e. people get paid, for the number of clicks their videos receive, a lot of people have started uploading these Clickbait’s to lure in clicks of the user. Using these, people earn money, and the users lose their precious time and are left with a bad experience.

Convolution neural networks (CNNs) [1] are a successful model for understanding picture content, giving condition of-the-workmanship [2, 3] results on picture acknowledgement, division, discovery and retrieval. The key empowering factors behind these outcomes were procedures that could bolster the learning procedure by scaling up the systems to a huge number of parameters and gigantic marked informational collections. A video is actually only a heap of pictures.

Since a video contains large amount of data in them, it becomes a tedious task to obtain the right video result whose contents match what we searched for. So, the proposed work makes videos searchable through the contents and the metadata associated with it and gives it appropriate ranking in the searched result list.

## 2 Background and Related Work

An examination accentuates [4], the age of portrayals of normal language for human activity and its association with video streams. Ordinary strategies for picture preparing were utilized to extricate abnormal state video highlights. These highlights are changed over into common language portrayals utilizing setting free sentence structure. Base up methodology has been utilized to depict video substance in normal

language. The formalism of the characteristic language permits familiar, rich portrayals to be produced, taking into account nitty-gritty and refined articulations. This strategy could be utilized for gathering identification, expansion of social models, progressively complex human and other item communications.

As per the existing literature [5], apparently, there seems a rise in the accuracy for CNN-based architecture for the classification of images. Since a video is stack of images, a CNN-based classifier can be used to predict what are the contents inside the video.

The initial CNN-based network, i.e. AlexNet gives an error of only 15.3% in the ImageNet Large-Scale Visual Recognition Challenge in 2012. It was considered to be the state of the art technology of that time.

The graphical processing unit (GPU) [6] helps a major time in reducing the time to train the model. GPUs are less flexible, but GPUs are designed to calculate the same instructions in parallel. Deep neural networks (DNNs) are structured in a very uniform way so that thousands of identical artificial neurons perform the same calculation at each layer of the network. GPUs have additional advantages over CPUs, including more computing units and a higher bandwidth for memory retrieval. Furthermore, in applications requiring more computations, GPU can be exploited to further speed up calculations [7].

The CNN-based model that showed a high level of accuracy is the VGG-based model. In the ImageNet Large-Scale Visual Recognition Challenge in 2014, two models by VGG were released, the 16-layer-based model VGG16 having an error of 7.8% and the 19-layer-based model VGG19 having an error of only 7.3% [1]. These VGG-based models took an input image of shape  $224 \times 224$  and extract features from it. The AlexNet has a filter size of  $7 \times 7$ , while VGG networks use a filter size of  $3 \times 3$ .

Clickbait is a marketing tool [8] used by many social media publishers that entices and misleads users to click on a certain link through the use of eye-catching teaser content, exaggerated descriptions, omission of essential information or even outright deception—whether or not users are actually interested in the content topic. This usually serves the purpose of maximizing the revenue generated through display advertisement on the content's page.

Publishers on social media platforms are increasingly using clickbait to spark the natural curiosity of their users and to get clicks on their content. Each click earns revenue from the advertisement display. Clickbait is big issue faced by the YouTube search engine [8, 9]. The proposed CNN-based video classifier can be used to classify the videos and provide improved results for the searched term. The VGG16 model is used to make a video classifier.

### 3 Problem Statement

Video searching on popular search engines utilizes the meta-information to find the relevant videos according to the search queries. Meta-information mainly comprises

the title and description of the video. The major drawback of this is that it overlooks whether or not the content of the video is genuine or not. Since the meta-information used to generate a search result is provided by the uploader, the person can easily provide false meta-information to it.

The proposed work classifies the video in five different sports-related categories and provides tags to this video. The classification of videos can then further help us to check whether the content of the video is what we actually desired or not.

## 4 Experimental Design

### 4.1 Training Data Set Selection

The SVW [10] sports video dataset is selected for training the model for classification. SVW consists of 4200 videos captured by users of the Coach's Eye smartphone app, a leading sports training app developed by TechSmith Corporation, solely with smartphones. SVW comprises 30 sports categories and 44 different actions. Due to the constraints on the hardware to train the model, five categories of sports were taken, namely Pole vault, Golf, Football, Basketball and Archery. The videos of SVW are broken down into frames per second and then it is used as a training data.

The number of videos in each category is

- Archery: Consists of 127 videos with an average duration of 11.98 s.
- Basketball: Consists of 111 videos with an average duration of 11.18 s.
- Football: Consists of 137 videos with an average duration of 12.01 s.
- Golf: Consists of 118 videos with an average duration of 11.55 s.
- Pole vault: Consists of 126 videos with an average duration of 9.75 s.

### 4.2 Proposed Model

The Visual Geometry Group (VGG) network architecture was introduced by Simonyan and Zisserman [1] related to very deep convolutional networks for large-scale image recognition (Fig. 1).

This network is characterized by its simplicity, using only  $3 \times 3$  convolutional layers stacked on top of each other in increasing depth. The convolution layer is the core building block of a CNN. The layer's parameter consists of a set of learnable filter, which has a small receptive field, but extends through the full depth of the input volume. Max pooling is used to reduce volume size (Fig. 2).

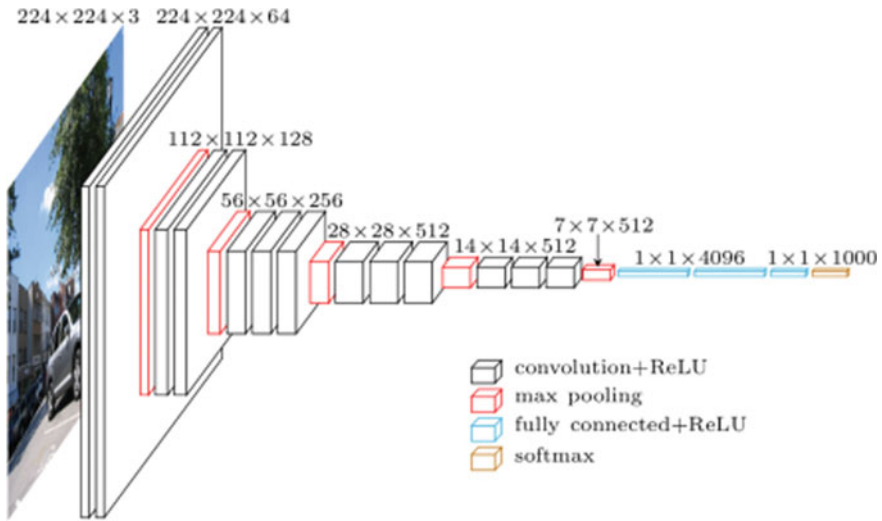


Fig. 1 Design for VGG16 ImageNet classifier [1]

The feature extraction is been done by a model similar to that of VGG16. Total number of parameters learned from the step of feature extractions is 138,357,544. The feature extracted from the model is then passed to a five-layer fully connected network having an output layer of six nodes with a “softmax” classifier (Fig. 3).

The architecture has an input shape of image of size  $224 \times 224$ . A shape of the features extracted from the VGG16-like layers is (1L, 7L, 7L, 512L). The extracted features from the convolution layer are 13,968,502,069,114. These features are then passed through the fully connected neural network. Trainable parameters present in the models are 26,380,934. For the training of the model, the dataset consists of 1962 images and 841 images for the validation set.

### 4.3 Approach

A video is really just a stack of images. So, the approach used to classify the video is mainly based on the concept that videos are a stack of images. If we know how to classify an image, we can classify the images present in the stack.

The number of frames of a video are predicted into different classes instead of determining a single class for a video. After getting the frames classified to their respective classes, a summarization of the result is done to get a final result.

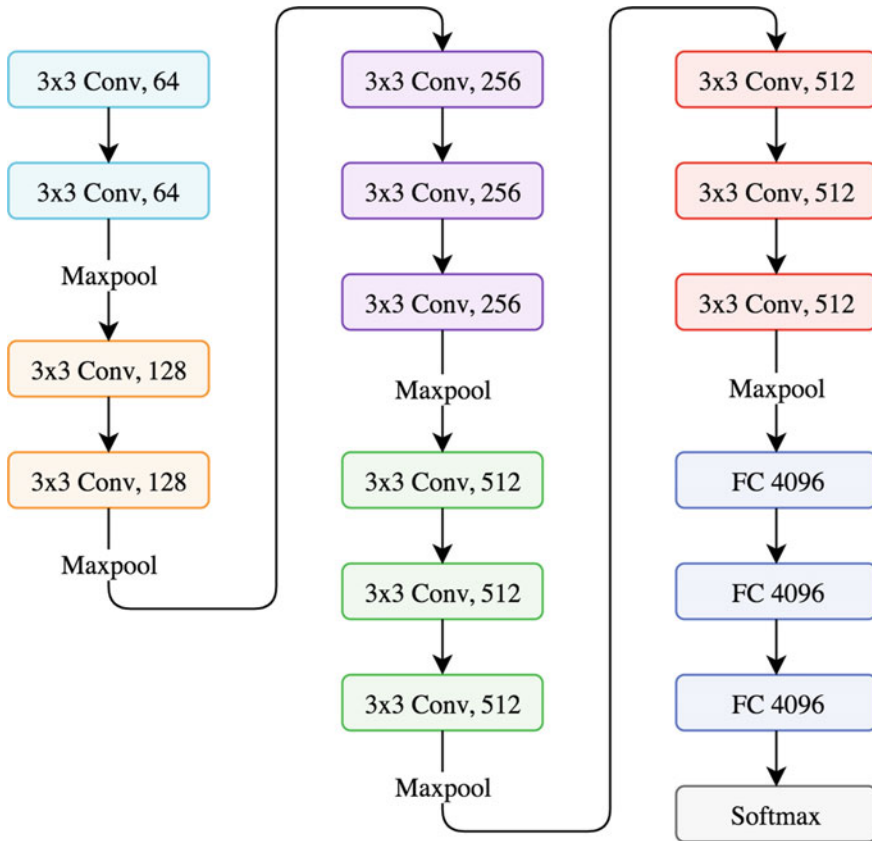


Fig. 2 Description of different layers of modified VGG16 model

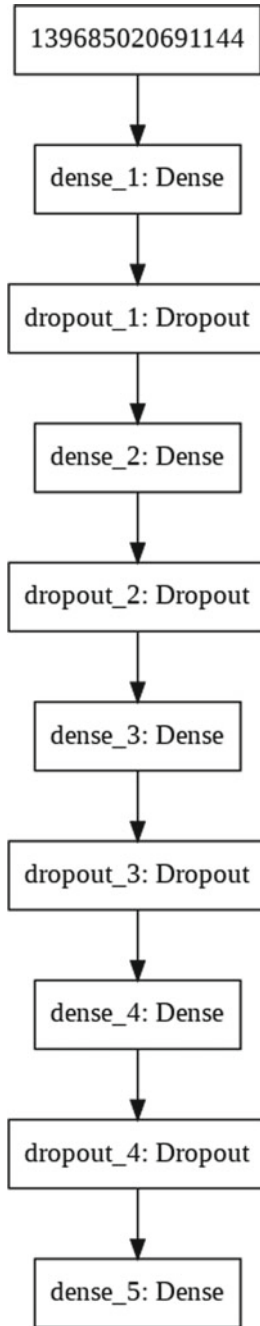
## 5 Experimental Results

This proposed work performed in Python programming language using Google Colaboratory cloud platform [11]. Colaboratory is a free Jupyter notebook environment that requires no set-up and runs entirely in the cloud. The model was tested against two types of data: the testing data available to us and the videos which were searched on YouTube.

### 5.1 Accuracy for Different Training Sizes

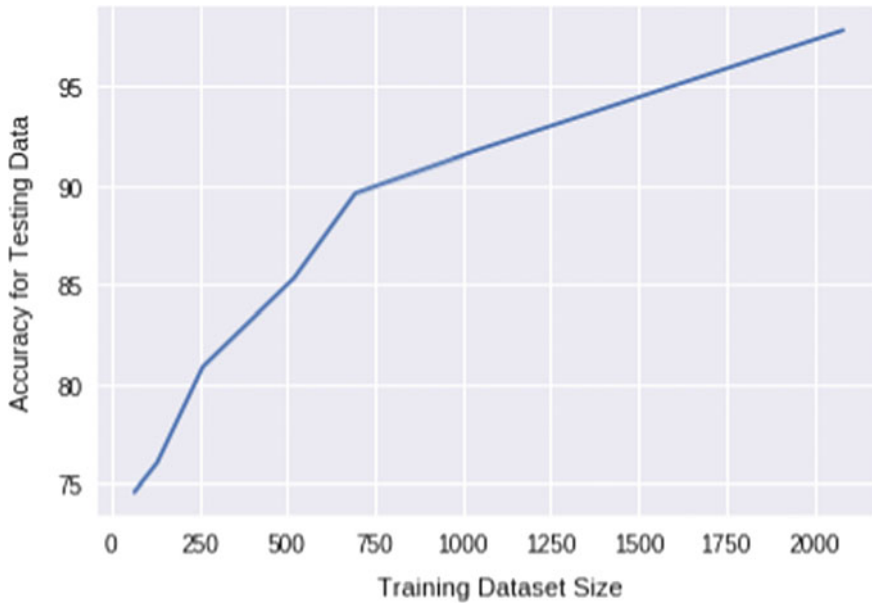
The accuracy is shown for varying training sizes as follows (Fig. 4):

The graph shows that there is a significance decrease in accuracy as the size of training set starts decreasing. For the training of the model, 2083 images were taken



**Fig. 3** Fully connected layer for classification of the image based on the features extracted





**Fig. 4** Graph for size of training dataset (number of videos) versus the accuracy (%) for a given dataset size

with a validation set of size 893 images. The testing accuracy for 1570 images passing in the model is 97.13%.

## 5.2 Testing on Real-Time Video Data

For the real-time testing of the model, sports-related videos were searched on the YouTube and were then used to predict which sport they belonged to. In total, 17 different sports videos were taken. Presently, due to hardware restriction, the maximum length of the video that could be taken is 10 min (Fig. 5).

The above graph shows a relation between the duration of the live YouTube testing data with the accuracy of the prediction. If the video is predicted for a specific duration, the prediction would not be accurate. As we can see from the graph, there is no clear relationship between the length of the video and accuracy of the prediction.

The prediction for real-time YouTube videos tested against the video prediction model trained is shown in Fig. 6. The testing was done for the 17 different sports video available on YouTube.

The graph shows that rather than taking a fixed number of frames if the whole video is taken into consideration for prediction, the video would be classified. The graph also shows the correct prediction for real-time YouTube videos.

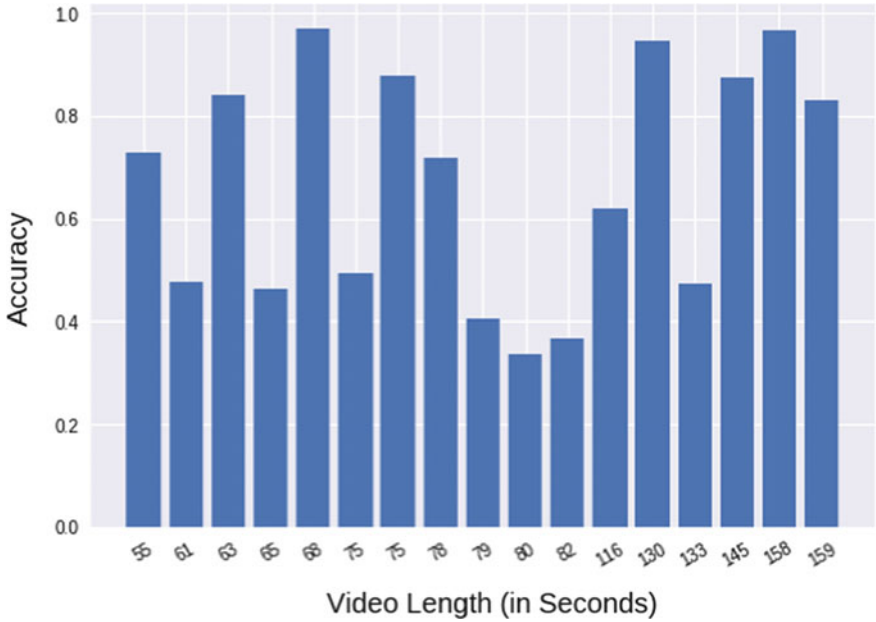


Fig. 5 The comparison graph of accuracy and length of the video

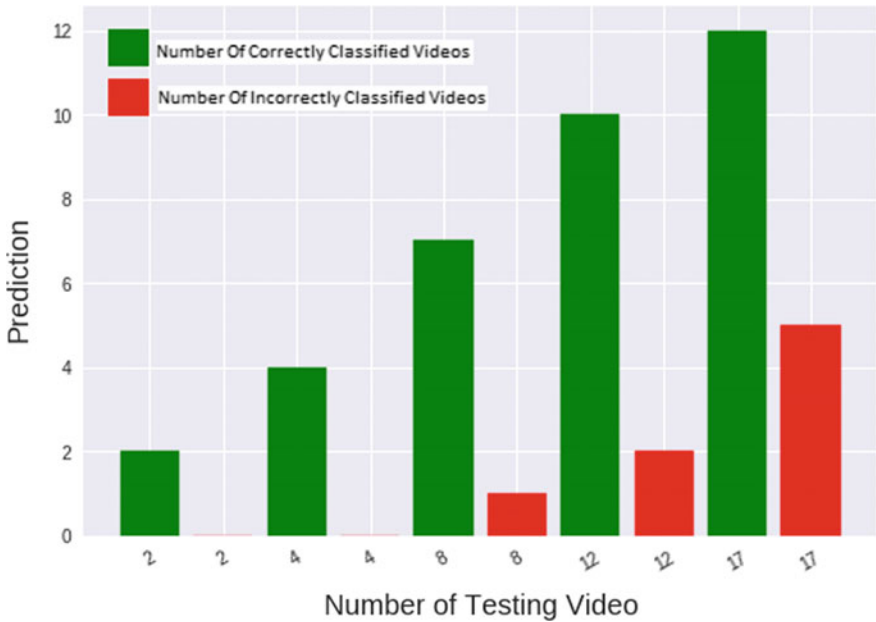


Fig. 6 The graph for number of videos tested vs correctly predicted (green-coloured bar shows correctly predicted, whereas red-coloured bar shows incorrectly predicted)

## 6 Conclusion

The paper has presented convolution neural network based on VGG16 to classify the video in sports domain. The proposed model has improved accuracy of 82.3%. Static image classification approach has been adopted instead of action classification by identifying the objects present in the video. The results provide better accuracy since various parameters for checking genuineness have been used: Frames division and metadata, comments, title, description, likes and dislikes. Presently, the classifier designed is designed and implemented to categorize the videos into five classes. This method of classification could be used to check the genuineness of the video content with respect to the searched query.

## References

1. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations. CA, USA
2. Yao K, Zweig G (2015) Sequence-to-sequence neural net models for grapheme-to-phoneme conversion. In: Interspeech 2015. Dresden, Germany
3. Pinto H, Almeida MJ, Gonçalves M (2013) Using early view patterns to predict the popularity of youtube videos. In: Proceedings of the 6th ACM international conference on web search and data mining
4. Khan MUG, Gotoh Y (2012) Describing video contents in natural language. In: Proceedings of the workshop on innovative hybrid approaches to the processing of textual data, pp 27–35
5. Covington P, Adams J, Sargin E (2016) Deep neural networks for youtube recommendations. In: Proceedings of the 10th ACM conference on recommender systems
6. Coates A, Huval B, Wang T, Wu DJ, Ng AY, Catanzaro B (2013) Deep learning with COTS HPC systems. In: 30th international conference on machine learning, pp 2374–2382
7. Kaiser L, Sutskever I (2016) Neural GPUs learn algorithms. In: CoRR abs/1511.08228
8. Qu J, Hißbach AM, Gollub T, Potthast M (2018) Towards crowdsourcing clickbait labels for youtube videos. In: HCOMP
9. Zannettou S, Chatzis S, Papadamou K, Sirivianos M (2018) The Good, the bad and the bait: detecting and characterizing clickbait on youtube. In: IEEE security and privacy workshops (SPW). San Francisco, CA
10. Sports Videos in the Wild (SVW) (2009) A video dataset for sports analysis. <http://cvlab.cse.msu.edu/project-svw.html>, last accessed 2019/05/09
11. Google Colaboratory Homepage. <https://colab.research.google.com/>, last accessed 2019/05/09

# A Novel IN-Gram Technique for Improving the Hate Speech Detection for Larger Datasets



Bharat Gupta, Nikita Goel, Dhruv Jain and Namita Gupta

**Abstract** Hate speech is a type of written or a spoken statement that is used to demean or humiliate a person or a community. In this era of new age socialism, this type of speech is prevalent on social media platforms, where certain groups of people display offensive behaviour towards some people that may be distributed over gender, religion, nationality, etc. These kinds of activities must be avoided or suspended on social media platforms. Therefore, it is necessary to automate the detection of hateful content that gets circulated on the social media. The research work provides an enhanced technique as compared to the existing techniques with improved performance. The proposed model of IN-Gram compares the performance of detection of hateful content on social media with the traditional TF-IDF, N-Gram and PMI techniques. The proposed approach improves the hate speech detection rate by 10–12% for larger datasets as compared to existing approaches.

**Keywords** Hate speech detection · Machine learning · NLP · PMI · N-Gram · TF-IDF

---

B. Gupta (✉) · N. Goel · D. Jain  
Department of Computer Science and Engineering, Jaypee Institute of Information Technology,  
Noida, India  
e-mail: [bharat.gupta@jiit.ac.in](mailto:bharat.gupta@jiit.ac.in)

N. Goel  
e-mail: [nikitagoel11@gmail.com](mailto:nikitagoel11@gmail.com)

D. Jain  
e-mail: [dhruvjain7777@gmail.com](mailto:dhruvjain7777@gmail.com)

N. Gupta  
School of Business Management, Sharda University, Greater Noida, India  
e-mail: [Go2namita@gmail.com](mailto:Go2namita@gmail.com)

## 1 Introduction

Hate speech detection involves tremendous amount of space and computation power. In order to enhance the performance of detection, the algorithms should be robust and reliable. In the proposed approach, a systematic approach is proposed by comparing the solutions obtained after implementation of improved N-Gram technique with the existing TF-IDF, N-Gram and PMI techniques. The proposed model trains the tweets extracted from Twitter using Twitter APIs. The data is cleaned to remove unnecessary tokens. The PMI technique focuses on the detecting the hateful tweet by using the frequency distribution of tokens used in positive and negative context. The N-Gram technique focuses on forming bigrams and fits the model in order to classify bigrams having negative tokens. The TF-IDF technique finds the term frequency-inverse document frequency of every word in the tweet dataset. The proposed technique improves the detection of hate speech using existing N-Gram as the base technique. The length of the dataset is varied, and performance in terms of accuracy of each algorithm is measured.

The proposed model not only detects the hateful content but also flags a red signal to the users entering the hateful content. It also helps users to identify the hateful content in their written text.

The IN-Gram technique helps in reducing the computation and increasing the efficiency of computation. The model aims at increasing the performance of detection of hateful content encountered by the system. This work aims in expanding the scope of usability for the benefit of the society.

## 2 Related Work

Hate speech detection has been widely studied and applied all over the world to several areas like reviews of people on a movie, product, service, etc. Feature extraction and classification techniques have been used for the detection of hateful content on the social media.

Fortuna et al. [1] have done a comparative analysis of the existing feature extraction and classification techniques.

Djuric et al. [2] detect the hate speech in the text through the bag-of-words approach by generating all the tokens and applying the k-nearest neighbour classifier to generate the result with the help of word cloud library. It identifies the nearest neighbour for a particular token from the corpus and groups them into offensive or non-offensive.

Watanabe et al. [3] classify the tweets into clean, offensive and hateful content by making unigrams and patterns derived from the training set as dictionaries and applying syntactical analysis to detect hate speech.

Schmidt et al. [4] have done a survey on different techniques used to detect hate speech content on social media such as using metadata, bag-of-words approach, sentiment analysis, lexicon-based approach and knowledge-based features. The classification models used for hate speech detection are support vector machines (SVM) and recurrent neural network (RNN) language models.

Bindal et al. [5] proposed point-wise mutual information (PMI) approach as a technique to detect hateful content in a textual speech. PMI score is calculated for each and every tweet for both hate, non-hate tweets, and then, a final score for every tweet is generated.

Gamback et al. [6] predict the hate speech in a multilinguistic data of English, German and Portuguese languages through N-Grams as the feature extraction technique. Characters as well as word N-Grams are used to represent as features to predict the hate speech in the context.

Bernard et al. [7] use binary and multilabel classification to predict hateful and non-hateful content. TF-IDF, Word2Vec are used as feature extraction techniques for syntactical analysis, while N-Grams is used as feature extraction technique for semantic analysis.

Davidson et al. [8] make use of crowd-sourcing as a method to generate labels for a sample of the tweets: hate speech, offensive language and other. The model involves training a multi-class classifier to differentiate between these categories. The model first used logistic regression with L1 regularization to reduce the dimensions of the data. Test was performed among a variety of models that have been used in prior work: logistic regression (LR), Naive Bayes, decision trees (DT), random forest and linear SVMs.

Akram et al. [9] have done comparative accuracy analysis of different classifiers like support vector machines, Naïve Bayes and K-nearest neighbour. The paper discusses data pre-processing steps, tokenizing the words and assigning TF-IDF scores to the words and then classifies it in categories with the help of the classifiers.

Gautam et al. [10] discuss a methodology of classifying text using SVM classifier. The text classification performed using natural language processing (NLP) involves tokenization and stemming, feature selection, training, testing, probabilistic model by learning. The paper suggests an algorithm for unstructured data which is at first pre-processed, and then, TF-IDF is used to assign weights to each individual tweet. After pre-processing, features are extracted from the text, and the text is henceforth converted into a trainable classifier model using SVM classifier.

### 3 Model Assumptions

Python is used as the programming language for the implementation of the model. The tweets are categorized as sexist, racist or none. The dataset contains a number of tweets extracted from Twitter database including the details regarding the tweets like usernames, locations, date of posting, etc., of users.

### 3.1 Model Setup

The experiment dataset consists of 16,353 tweets which are extracted from Twitter. These are used as a database [11] that consists of sexist, racist, or none tweets, usernames, authors, retweets with other features.

### 3.2 Model Libraries

Some of the libraries used in the proposed model to predict hate speech in a given context are as follows:

- NLTK: Natural language toolkit (NLTK) possesses many features like word tokenization, removal of stop words, different stemmers such as Porter Stemmer and Snowball Stemmer. It also helps in validating the results obtained from the training and test sets through many performance measures.
- SkLearn: SkLearn helps in splitting up of the database into training and test sets and converts a collection of text documents into a matrix of tokens.
- Pandas: Pandas library is used in making data frames out of the dataset stored in a .csv file. It makes utilization and processing of tweets easier for further calculations.
- Gensim: Gensim library is used to load bag-of-words and creates Word2Vec dictionaries.

## 4 Proposed Methodology

This section discusses the different existing and the proposed improved approach used for hate speech detection. The proposed model improves the existing approaches used in N-Gram, PMI and TF-IDF techniques.

### 4.1 PMI Technique

This algorithm focuses on categorizing the tokens into binary sentiments. The list storing the PMI scores is called the sentiment lexicon. The list identifies the sentiment as negative and positive by observing the sign of the score assigned. The existing model [5] focuses on using the score calculation by using the following formula:

$$\text{Score}(t) = \log \left( \frac{\text{frequency}(t, \text{hate}) * \text{frequency}(\text{non-hate})}{\text{frequency}(t, \text{non-hate}) * \text{frequency}(\text{hate})} \right)$$

where  $\text{Score}(t)$  is the lexicon sentiment score of the term  $t$ ,  $\text{frequency}(t, \text{hate})$  represents frequency or number of occurrences of term  $t$  occurring in hateful context in the dataset of tweets,  $\text{frequency}(t, \text{non-hate})$  refers to frequency of term  $t$  occurring in non-hateful context,  $\text{frequency}(\text{hate})$  represents the total number of hateful tweets in datasets, and  $\text{frequency}(\text{non-hate})$  represents the total number of non-hate tweets.

## 4.2 TF-IDF Technique

TF-IDF scores are used to calculate and match the words within the training set and then it predicts the result with better accuracy and precision. The traditional technique uses TF-IDF calculation and bag-of-words (BOW) approach.

The TF-IDF calculation is done as follows:

$$\text{TF}(i, j) = \frac{\text{Term } i \text{ frequency in the } j\text{th document}}{\text{Total words in } j\text{th document}}$$

$$\text{IDF}(i) = \log_2 \frac{\text{Total documents}}{\text{Documents with term } i}$$

TF-IDF score for term  $I$  in the document  $j$  can be given as

$$\text{TF-IDF}(i) = \text{TF}(i, j) \times \text{IDF}(i)$$

The TF-IDF technique cross validates the results obtained by estimating the accuracy of prediction with the validation set. TF-IDF score [2] is assigned to each of the word in the bag-of-word corpus of training set. The same method is applied in the testing phase on validation set, and TF-IDF scores are assigned to each of the word in the bag-of-words corpus of the validation set.

The last step involves obtaining the training and test sets containing TF-IDF scores for each of their words, and classification is performed by using SVM classifier by passing both the training and test sets as parameters into the model.

## 4.3 N-GRAM Technique

The existing N-Gram technique involves forming multi-grams and comparing these bigrams on the basis of their occurrences. It involves a high computation in comparing bigrams, trigrams and poly-grams. There is no mechanism of filtering the occurrences of most occurring words and least significant words. The existing models of N-Gram make bigrams, trigrams and poly-grams out of consecutive words. This technique involves high computation power and complexity to execute.



The bigrams for smaller datasets are less in number but as the size of dataset increases, the number of bigrams increases substantially. It increases the latency of the model, and the model results in degraded performance.

For example, considering a sentence, “I do not like fish” will form bigrams: “I do + not”, “do + not like”, “not + like fish”. This model first removes stop words from the context like (“a”, “an”, “the”, etc.), and then, the model considers proper nouns and uncommon words as the bigrams. Hence, for large number of tweets, the model would generate large number of bigrams thereby increasing time and space complexity and decreasing the performance of the model.

#### ***4.4 Improved N-Gram Technique (IN-Gram)***

N-Gram technique is used to find the occurrence of words in the form of bigrams, trigrams, etc., in the document. In the improved approach, each of the classified labels is separately categorized as hateful and neutral labels. The bigrams are selected in such a manner that frequent occurring words, i.e. like, as, of, for, etc., and the least significant words, e.g. Allen, George, Caroline, etc., in the document are removed. After cleaning each tweet, the bigrams are formed by passing the mapped datasets into Count Vectorizer where the vectorized bigrams are fitted, transformed and densed in order to form a matrix which counts occurrences of each bigram. The processed bigrams are further assigned score on the basis of their occurrences in the dataset. After training these bigrams, these are validated by the tweets of the validation set and new tweets also given scores in the testing phase.

The flowchart shown in Fig. 1 represents an architecture designed to develop the improved N-Gram model. For both training and validation sets, a dictionary of tweets is maintained which stores the occurrences of their respective bigrams. The training is performed using support vector machine (SVM), and performance is predicted by calculating the accuracy of the enhanced model.

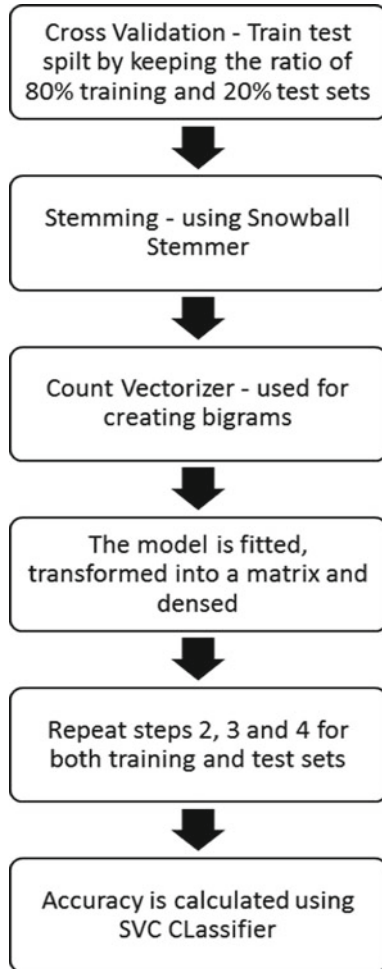
## **5 Experimental Results**

The accuracy metrics is used as performance measure to detect hate speech content in the performed experiments after applying different feature extraction and classification techniques.

Table 1 shows the permissible range of accuracy, which is used presently for the hate speech detection when data size is varied.

The graph in Fig. 2 provides a comparative analysis between improved (IN-GRAM) and existing traditional techniques (TF-IDF, PMI, N-GRAMS) when dataset size is varied and small.

For the tweets ranging from 2000 to 4000, the permissible range allows the accuracy to be greater than 45%, while the model accuracy obtained by applying the



**Fig. 1** Flowchart of the improved N-Gram approach

**Table 1** Permissible range according to varying size of dataset

Size of dataset	Permissible range	Model accuracy
2000	>45.5	65.01
4000	>55	56.09
6000	>70	66.94
8000	>72	61.92
10,000	>74.5	77.41
12,000	>80	91.41
14,000	>77.5	93.41
16,000	>86.5	95.45

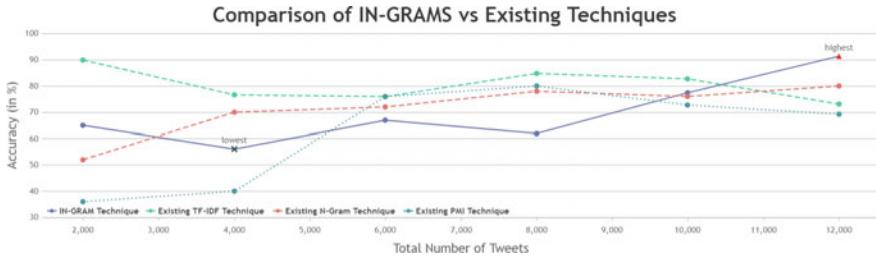


Fig. 2 Accuracy prediction for existing and improved models of IN-Gram for smaller datasets

improved IN-GRAM technique as the feature extraction technique accounts to nearly 65%.

It can also be inferred from Fig. 2 that when the size of the dataset varies in the range of 2000–6000 tweets, the TF-IDF technique has good accuracy as compared to others. When the size of the dataset is in the range of 6000–10,000, PMI technique has good accuracy.

The sizes of datasets are increasing significantly in social networking sites. The proposed technique has better accuracy for larger datasets as per Fig. 2. The experimental results of comparison of IN-Gram and other existing techniques for 12,000–16,000 tweets are further explored in Fig. 3.

The graph in Fig. 3 shows that the proposed technique has good accuracy when the size of the dataset is above 12,000 tweets. The accuracy has improved by 10–12% than the existing approaches. The proposed IN-GRAM has low accuracy when the size of the dataset is small.

For 16,000 tweets, i.e. for larger datasets, permissible range allows the accuracy to be greater than 86.5%, while the model accuracy obtained by applying proposed IN-GRAM technique is 95%. The existing N-Gram model has an accuracy of 87% for this dataset.

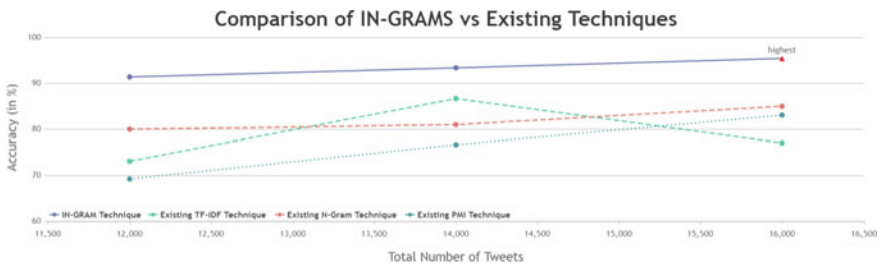


Fig. 3 Accuracy prediction for existing and improved models of IN-Gram for larger datasets

As the size of the dataset increases above 12,000 tweets, IN-GRAM technique gives us the good accuracy, i.e. 10–12% improvements as compared to others techniques. So, it can be inferred that the proposed improved model IN-GRAM performs well for larger datasets, while the existing techniques have good accuracy for relatively smaller datasets.

## 6 Conclusion

Detection of hate speech manually is a time-consuming process. It requires an automated approach for detecting hate speech by using different techniques. The proposed research uses accuracy as a performance measure for the hate speech detection. The traditional TF-IDF technique has good accuracy for 2000–6000 tweets dataset. The existing PMI technique gives high accuracy for the range of 6000–10,000 tweets dataset. The IN-GRAM technique has improved the performance of accuracy by 10–12% for larger datasets. The size of dataset in the social networking sites is increasing drastically. The proposed IN-GRAM approach will be more beneficial in detecting hate speech content on social media sites.

## References

1. Fortuna P, Nunes S (2018) A survey on automatic detection of hate speech in text. *ACM Comput Surv CSUR J* 51(4):85
2. Djuric N, Zhou J, Morris R, Grbovic M, Radosavljevic V, Bhamidipati N (2015) Hate speech detection with comment embeddings. In: *Proceedings of the 24th international conference on world wide web*. NY, USA, pp 29–30
3. Watanabe H, Bouazizi M, Ohtsuki T (2018) Hate speech on twitter: a pragmatic approach to collect hateful and offensive expressions and perform hate speech detection. *IEEE Access* 6:13825–13835
4. Schmidt A, Wiegand M (2017) A survey on hate speech detection using natural language processing. In: *Proceedings of the fifth international workshop on natural language processing for social media, association for computational linguistics*, pp 1–10
5. Bindal N, Chatterjee N (2016) A two-step method for sentiment analysis of tweets. In: *International conference on information technology (ICIT)*. Bhubaneswar, India, pp 218–224
6. Unsvag EF, Gamback B (2018) The effects of user features on twitter hate speech detection. In: *Proceedings of the second workshop on abusive language online (ALW2)*. Brussels, Belgium, pp 75–85
7. Jansen BJ, Salminen J, Almerexhi H, Milenkovic M, Jung S, An J, Kwak H (2018) Anatomy of online hate: developing a taxonomy and machine learning models for identifying and classifying hate in online news media. In: *International conference on web and social media*
8. Davidson T, Warmsley D, Macy M, Weber I (2017) Automated hate speech detection and the platform of offensive language. In: *Proceedings of the eleventh international AAAI conference on web and social media*
9. Zia T, Akram MS, Nawaz MS, Shahzad B, Abdulllatif M, Mustafa RU, Lali MI (2017) Identification of hatred speeches on twitter. *Int J Adv Electron Comput Sci* 4(1)
10. Gautam N, Bhardwaj A (2016) Proposed approach for optimize the text classification using SVM-RBF Kernel. *Int J Comput Sci Inf Technol Secur IJCSITS* 6

11. Hatespeech. <https://github.com/ZeeraK/hatespeech>, last accessed 2019/03/21
12. Pak A, Paroubek P (2010) Twitter as a corpus for Sentiment Analysis and opinion mining. In: Proceedings of the seventh conference on international language resources and evaluation, European languages resources association

# Design of Configurable Analog Block-Based Oscillator and Possible Applications



Kushaagra Maheshwari, Sudhanshu Maheshwari and Piyush Yadav

**Abstract** This article introduces a configurable analog block with wide functionality which is then used for designing a new sinusoidal signal generator with three outputs. The circuit uses three current feedback operational amplifiers and provides control over the frequency of oscillation independent of condition of oscillation. The new circuit is experimentally tested using AD844 ICs and also simulated using capture CIS tool. Both experimental and simulation results have good agreement. Possible applications are suggested, and one such application is demonstrated in generation of square and triangular waveforms.

**Keywords** Analog circuits · Sinusoidal oscillators · CFOAs · Waveform generation

## 1 Introduction

The design of analog circuits is a challenging step in the study of electronic and communication systems. Analog design can be better understood if modular design approach is used. This approach enables the design of larger systems using smaller and identical sub-systems. This approach has also been referred to as the one employed for field-programmable analog arrays (FPAA). A configurable analog

---

K. Maheshwari (✉) · P. Yadav

Department of Electronics and Communication Engineering, G.L. Bajaj Institute of Technology and Management, Greater Noida, India  
e-mail: [kushaagramaheshwari02@gmail.com](mailto:kushaagramaheshwari02@gmail.com)

P. Yadav

e-mail: [piyushyadav1985@gmail.com](mailto:piyushyadav1985@gmail.com)

S. Maheshwari

Department of Electronics Engineering, AMU, Aligarh, India  
e-mail: [maheshwarispm@gmail.com](mailto:maheshwarispm@gmail.com)

block (CAB) is used at the heart of FPAA. Recent designs allow the benefits of current mode building blocks to be exploited for the purpose [1]. One such current mode building block is current feedback opamp (CFOA), which is also commercially available. This becomes a motivating factor to introduce a simple generalized analog block using single CFOA. The functionality of analog block is discussed. These blocks are used for designing new quadrature oscillator circuit. The proposed circuit uses three such blocks and provides three outputs. Therefore, the new proposed quadrature oscillator employs three CFOAs and is superior to voltage opamp-based circuit which uses four opamps [2]. The new circuit is advantageous due to the higher performance of CFOAs over voltage operational amplifiers (VOAs) used in previous work [2]. On the other hand, when compared to the large number of exemplary circuits, the new circuit is a feasible addition to the existing ones [3–12]. Whereas the CFOA applications were reported long back [13], the active-RC oscillators have also been studied decades back [14] and continue to appear in very recent literature [15]. Next, the proposed oscillator is further used to implement a system-level application as function generation. The ideas proposed are experimentally verified and also supported by simulation studies.

Section 2 introduces the CAB, its functionalities, and the design of quadrature oscillator circuit. Section 3 is on verification results; some possible applications are given in Sect. 4. The concluding comments are given in Sect. 5.

## 2 Proposed Idea

### 2.1 CFOA-Based CAB

A current feedback operational amplifier (CFOA) has the following terminal characteristics and can be implemented using commercial chips by analog devices, namely AD844.

$$I_y = 0; V_x = V_y; I_z = I_x; V_w = V_z; \quad (1)$$

The simple configurable analog block using one CFOA is shown in Fig. 1. Various functionalities are listed below:

- (A) For input at  $V_1 = V_{in}$ ;  $V_o = -Z_1/Z_2$ .
- (B) For input at  $V_1 = V_2 = V_{in}$ ;  $V_o = 1 - Z_1/Z_2$ .

From the above, various simple functions can be obtained like amplifier, integrator, differentiator from the first-mentioned option (A), with appropriate choice of  $Z_1$  and  $Z_2$ . From the second-listed option (B), several functions can be realized with appropriate selection of  $Z_1$  and  $Z_2$ . Therefore, Fig. 1 is called a configurable analog block, capable of realizing many simple analog functions under different conditions.

### 2.2 Quadrature Oscillator Using CAB

The use of Fig. 1 is now shown for suggesting a new quadrature oscillator. The proposed circuit is shown in Fig. 2. It can be visualized as a cascade three simple circuits obtained from the CAB of Fig. 1. As per Fig. 2, above-mentioned choices (listed below) and an appropriate selection of  $Z_1$  and  $Z_2$  in Fig. 1 result in the new oscillator circuit.

- (i) Choice (B):  $Z_1 = \left(\frac{R_1}{1+sR_1C_1}\right)$  and  $Z_2 = R_2$ .
- (ii) Choice (A):  $Z_1 = \left(\frac{1}{sC_2}\right)$  and  $Z_2 = R_3$
- (iii) Choice (B):  $Z_1 = 2R'$  and  $Z_2 = R'$

The characteristic equation of proposed oscillator circuit is given below

$$s^2 + s\left(\frac{1}{R_1C_1}\right) - s\left(\frac{1}{R_3C_2}\right) + \frac{1}{R_3C_1C_2}\left(\frac{1}{R_2} - \frac{1}{R_1}\right) = 0 \tag{2}$$

The frequency of oscillation (FO) and condition of oscillation (CO) are found as below.

$$f_o = \frac{1}{2\pi} \sqrt{\frac{R_1 - R_2}{R_1R_2R_3C_1C_2}} \tag{3}$$

$$R_3C_2 \geq R_1C_1 \tag{4}$$

Equations (3, 4) show that the FO can be varied by resistor ( $R_2$ ), without disturbing the CO. This is a very important characteristic of an oscillator. The CO can be maintained by  $R_3$  although it may also be maintained by  $R_1$ ; the former is a better design option.

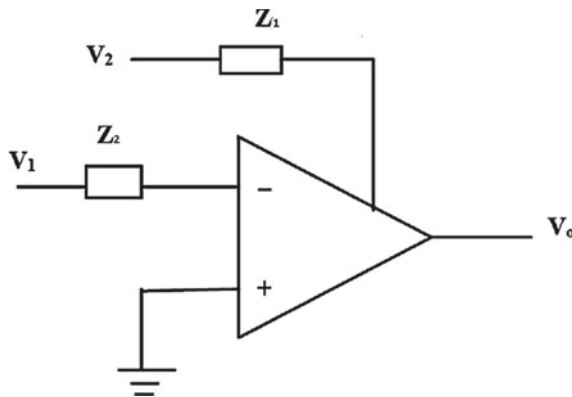


Fig. 1 Generalized configurable analog block



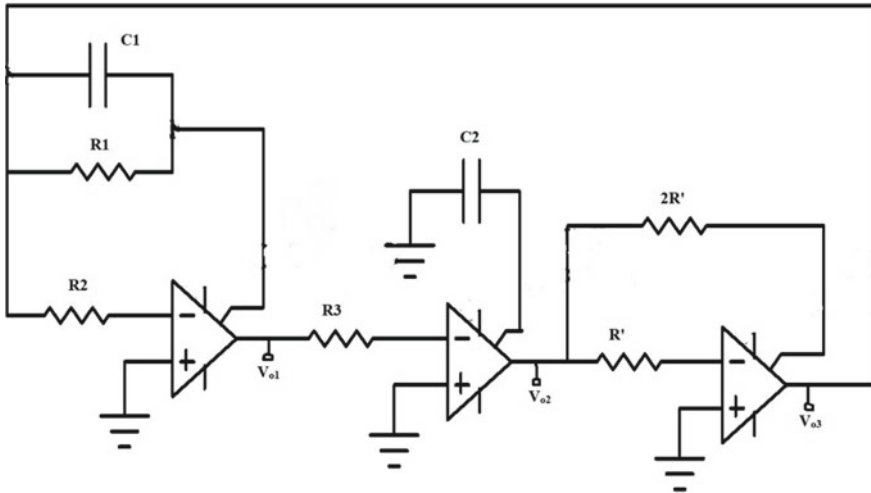


Fig. 2 New quadrature oscillator using generalized CABs

### 2.3 Oscillator Design

A simple design could be used with  $C_1 = C_2$ , which results in the following design equations.

$$f_o = \frac{1}{2\pi C} \sqrt{\frac{R_1 - R_2}{R_1 R_2 R_3}} \tag{5}$$

$$R_3 \geq R_1 \tag{6}$$

Now for condition of oscillation,  $R_3$  may be selected the same as  $R_1$  by making  $R_3$  variable and adjusting it close to  $R_1$ . Let these two resistors be ‘ $R$ ’. As another design step, if  $R_2 = 0.5R_1$  then the frequency of oscillation becomes:

$$f_o = \frac{1}{2\pi C R} \tag{7}$$

Equation (7) gives the simplified design for the oscillator circuit.

## 2.4 Oscillator Outputs

The circuit provides three outputs as shown in Fig. 2. In terms of circuit elements, the three outputs are related as

$$V_{02} = -\left(\frac{1}{sR_3C_2}\right)V_{01}; \quad V_{03} = -V_{02} \quad (8)$$

From Eq. (8), it can be seen that  $V_{02}$  leads  $V_{01}$  by  $90^\circ$  and  $V_{03}$  is in-phase opposition to  $V_{02}$ . Therefore, it seems that the three outputs of the proposed oscillator are in quadrature relationship with progressive  $90^\circ$  shift. Now for the frequency of oscillation as obtained for the simple design as mentioned in the preceding (Eq. 7), the three outputs are also equal in their magnitudes.

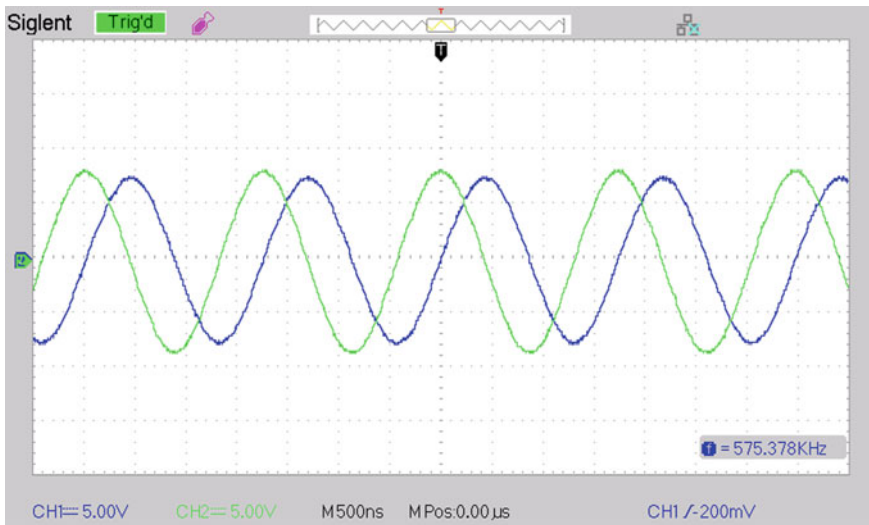
## 2.5 State of the Art Compared

The new quadrature oscillator is compared with very recent circuits published in the literature. The new circuit requires only three ICs when compared to the circuit of Ref. [3], which uses a very large number of ICs for realizing quadrature oscillator. Similarly, the new circuit uses fewer ICs when compared another good work [4], where the number of ICs used are four, as compared to three in the proposed work. The circuits of Ref. [5] present single-phase or three-phase oscillators, where two or six chips are employed, respectively. So the three-phase circuit in this paper uses fewer ICs than the recently appeared work [5]. Another oscillator required six ICs of two types for generating quadrature sinusoidal waveforms [8]. One very rich work using current conveyors was built using eight AD-844 chips, which is also quite large number, as compared to present work [11]. This study on very recent literature is presented in light of voluminous literature on the subject, which has been very elegantly compiled in the recent monograph by Senani et al. [6] and as review by Abuelma'atti [9]. Coming to the passive components, it seems that five resistors are used along with two capacitors, but this needs more elaboration. Since the circuit is realized using generalized analog block, the same demands use of an inverting stage formed by third CFOA and two resistors with 1:2 matching. Therefore, these resistors do not appear in oscillator FO and CO equations. Effectively, three resistors decide the FO and CO, and the benefit of independent tuning of one another still prevails. Hence, it can be concluded that the oscillator parameters actually depend on three resistors and two capacitors, which compare well with most of the circuits found in

the open literature, which use four or five resistors. A very recent study shows the use of four resistors and two capacitors, built using three CFOAs [15]. The purpose of this study does not target oscillator bibliography, rather a contribution to the new circuit design, which is not found in the literature.

### 3 Experimental and Simulation Results

The designed quadrature oscillator of Fig. 2 which has been build using the CAB of Fig. 1 is experimentally tested using three AD844 chips. The ICs are biased at  $\pm 10$  V power supply. The simple design presented in Sect. 2.3 is used with capacitor values as 50 pF.  $R_1$  and  $R_2$  are selected as 4.4 k $\Omega$  and 2.2 k $\Omega$ , respectively. The condition of oscillation is set by using variable  $R_3$  which is adjusted close to  $R_1$  as per Eq. (6). Using these selected values, the frequency of oscillation is found close to 575 kHz, as shown in the results of Fig. 3. The two quadrature outputs are seen in Fig. 3 followed by their XY plot in Fig. 4, which is a circle, thus suggesting quadrature relationship. The proposed quadrature oscillator is next simulated using ‘Capture CIS Cadence’



**Fig. 3** Experimentally obtained quadrature waveforms of new circuit

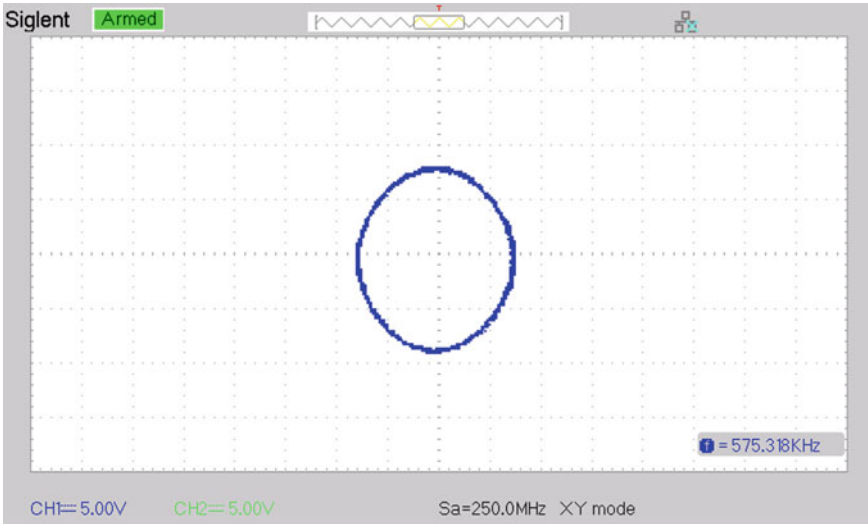


Fig. 4 XY plot showing quadrature relationship

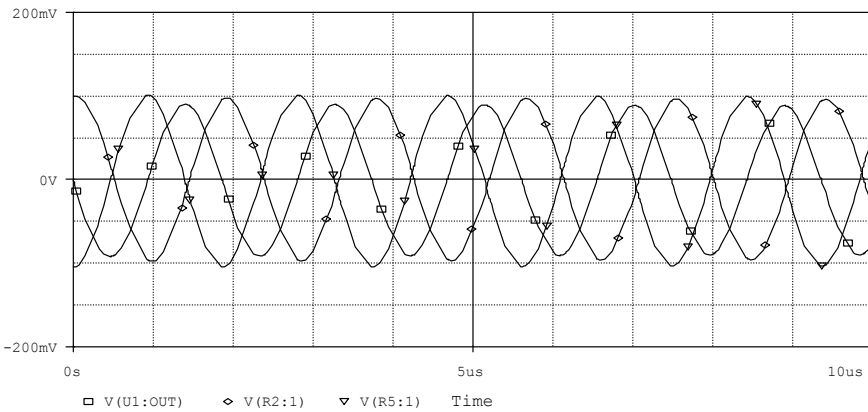


Fig. 5 Simulated waveforms of new circuit

tool for the same designed values as used in experimentations above. The simulation results are shown in Fig. 5, where the three outputs at  $f_{O} = 552 \text{ kHz}$  are obtained. The spectrum is seen in Fig. 6, with very good performance, and THD was less than 0.5%. Therefore, it is clear that the experimental and simulated  $f_{O}$  is in close agreement.

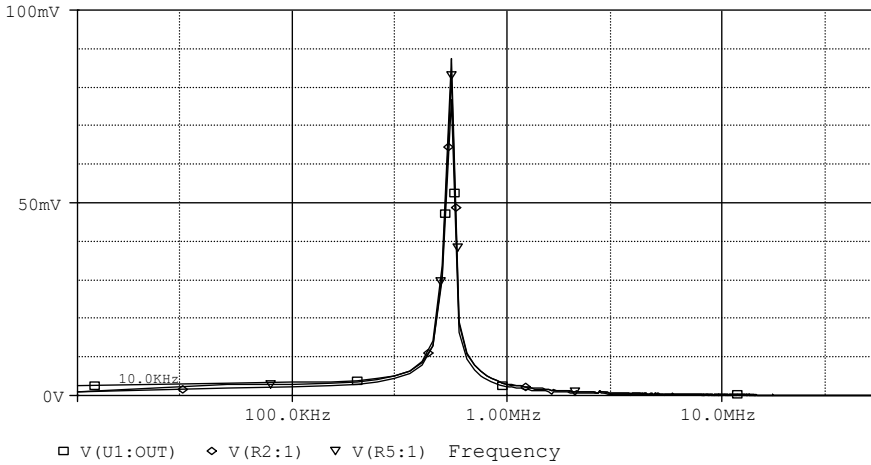
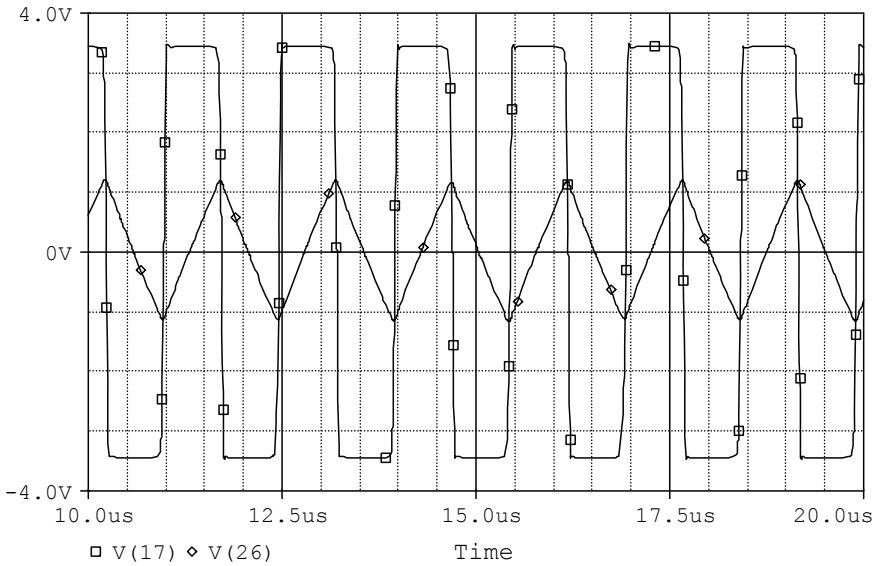


Fig. 6 Spectrum of new oscillator

### 4 Possible Applications

The new CFOA-based circuit with three outputs is useful for realizing many applications in electronics and communication engineering [16]. It is well known that communication systems need carrier signal generator with appropriate quadrature property. In digital communication, signals of sinusoidal nature carry the digital message using ASK, BPSK, QPSK, etc. [12]. Another very interesting application of sinusoidal signals is in the generation of other waveforms. For example, when passed through a comparator a sinusoidal signal gets converted into square wave signal. Next, with square wave as an input signal and the processing analog block as integrator, a triangular signal can be generated. Thus, the oscillator circuit can be easily extended to the generation of square and triangular wave signals. This particular application is further demonstrated in this paper by using CFOA-based comparator and integrator circuit. The comparator and integrator can be realized using generalized CAB as shown in Fig. 1. To demonstrate this, one of the output signals of the oscillator is passed through a comparator, whose output is then integrated. The resultant waveforms in simulation setup are shown in Fig. 7.



**Fig. 7** Square and triangular waveforms obtained from the proposed oscillator

## 5 Conclusion

This paper deals with the design of a new sinusoidal oscillator with quadrature outputs, which is realized using a generalized configurable analog block. The new circuit benefits from the use of identical blocks enabling design ease. The circuit uses three CFOAs and enjoys control of the frequency of oscillation without affecting the condition of oscillation. Both experimental and simulation results are given with good agreement to one another. As one application, the generation of square and triangular waveforms is also demonstrated. Further applications may be explored as future work.

## References

1. Premont C, Grisel R, Abouchi N, Chante JP (2009) A current conveyor based FPAA. *Analog Integr Circ Sig Process* 17:105–124
2. Maheshwari K (2019) A new sinusoidal quadrature oscillator for electronics engineering. *Int J Electron Inf Eng* 10(1):45–50
3. Chaturvedi B, Kumar A, Mohan J (2019) Low voltage operated current-mode first-order universal filter and sinusoidal oscillator suitable for signal processing applications. *Int J Electron Commun (AEU)* 99:110–118
4. Arora TS, Gupta S (2018) A new voltage mode quadrature oscillator using grounded capacitors: an application of CDBA. *Eng Sci Technol Int J* 21:43–49

5. Yuce E, Verma R, Pandey N, Minaei S (2019) New CFOA based first order all pass filters and their applications. *Int J Electron Commun (AEU)*. <https://doi.org/10.1016/j.aeue.2019.02.017>
6. Senani R, Bhaskar DR, Singh VK, Sharma RK (2016) *Sinusoidal oscillators and waveform generators using modern electronic circuit building blocks*. Springer International Publishing, Switzerland
7. Maheshwari S, Ansari MS (2012) Catalog of realizations for DXCCII using commercially available ICs and applications. *Radioengineering* 21(1):281–289
8. Kumar A, Chaturvedi B (2018) Novel CMOS dual-X current conveyor transconductance amplifier realization with current-mode multifunction filter and quadrature oscillator. *Circuits Syst Signal Process* 37:2250–2277
9. Abuelma'atti MT (2017) Recent developments in current-mode sinusoidal oscillators: circuits and active elements. *Arab J Sci Eng* 42:2583–2614
10. Biolkova V, Bajer J, Biolek D (2011) Four phase oscillators employing two active elements. *Radioengineering* 20(1):334–339
11. Sharma RK, Arora TS, Senani R (2017) On the realisation of canonic single resistance-controlled oscillators using third generation current conveyors. *IET Circuits Devices Syst* 11(1):10–20
12. Ali I, Maheshwari S (2017) Multiple PSK modulator design using current mode building. In: *IMPACT-2017*, pp 262–266. <https://doi.org/10.1109/mspct.2013.6782128>
13. Soliman AM (1996) Applications of the current feedback operational amplifiers. *Analog Integr Circ Sig Process* 11(3):265–302
14. Bhattacharyya BB, Sundaramurthy M, Swamy MNS (1980) Realization of tunable RC-active oscillators using grounded capacitors and voltage amplifiers. *Int J Circuit Theory Appl* 8(4):355–371
15. Maheshwari S (2019) Realization approach for sinusoidal signal generation and circuit with easy control. *J Circuits Syst Comput*. <https://doi.org/10.1142/S0218126620500310>
16. Kashish K, Priya M, Yadav P (2016) Design of low power pulse oximeter for early detection of hypoxemia. In: *IEEE international conference on micro-electronics and telecommunication engineering (ICMETE)*, Technically co-sponsored by IEEE Electron Devices Society. <https://doi.org/10.1109/icmete.2016.29>, pp 600–605

# Enhanced Reliability of Polarity Controllable–Ferroelectric–FETs under the Impact of Fixed Trap Charges



Priyanka Pandey and Harsupreet Kaur

**Abstract** In the present study, a detailed analysis has been done to investigate the device performance of Polarity Controllable–Ferroelectric–Field Effect Transistors (PC–FE–FET) under the influence of fixed trap charges. It has been observed that by virtue of ferroelectric layer, the proposed device shows improved device characteristics over the conventional device for both n- and p-modes of operation even in the presence of traps. Here, NTC (PTC) present in PC–FE–FET device operating in n- (p-) mode exhibits superior subthreshold characteristics over the conventional device. Moreover, due to higher on-state current obtained for PC–FE–FET, devices with PTC (NTC) in n- (p-) mode demonstrate increased values of transconductance, cut-off frequency and  $I_{\text{on}}/I_{\text{off}}$  ratios.

**Keywords** Ferroelectric · Negative capacitance · Polarity Controllable FET · Transconductance

## 1 Introduction

The continuous demand for ultra-low power CMOS devices/circuits has led to exhaustive miniaturization in device dimensions which has also resulted in detrimental effects such as short-channel effects (SCE), hot-carrier effect (HCE), low current drivability, increased parasitic resistance, increased power dissipation, etc., that degrade device performance in sub-nanometer regime. In order to combat these issues, various innovative device designs have been introduced and one of them is negative capacitance (NC)-based devices [1, 2] that employ ferroelectric (FE) layer as gate insulator. These devices are capable of providing steep-switching characteristics along with high current drivability [1, 2].

---

P. Pandey · H. Kaur (✉)

Department of Electronic Science, University of Delhi South Campus, Benito Juarez Road, New Delhi 110021, India

e-mail: [harsupreetkaur@gmail.com](mailto:harsupreetkaur@gmail.com)

P. Pandey

e-mail: [priyanka.pandey004@gmail.com](mailto:priyanka.pandey004@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication*

*Engineering*, Lecture Notes in Networks and Systems 106,

[https://doi.org/10.1007/978-981-15-2329-8\\_64](https://doi.org/10.1007/978-981-15-2329-8_64)



Further, in order to overcome the critical issue of parasitic resistance, Schottky-barrier field-effect transistors (SBFETs) have also gained worldwide interest [3, 4]. In these devices, the conventional doped source/drain (S/D) regions are replaced by metal silicides such as NiSi, PtSi, etc., due to which overall parasitic resistance reduces [4]. Furthermore, many research groups have also demonstrated that with proper tuning of Schottky junctions present at metal S/D-silicon interface along with application of appropriate external bias signal, Polarity Controllable-Field Effect Transistors (PC-FET) can be designed which hold the potential to be operated both as n-FET or p-FET [5, 6].

Moreover, as dimensions reach nanometer regime, the effect of fixed trap charges (FTC) at oxide-silicon interface also plays a dominant role and thus cannot be neglected. FTC is basically either positive trap charges (PTC) or negative trap charges (NTC) that result either during the fabrication procedure or due to HCEs or may also result when device is subjected to positive or negative stress bias [7–9], and result in shift in the threshold voltage thereby affecting device behavior.

In recent years, various studies have been reported on PC-FET [1, 2, 10, 11]. However, till now, to the best of our knowledge, no work has been attempted to study the effect of FTC on the device performance of PC-FETs. Hence, in view of this, in the present work, a comprehensive study has been carried out to extensively investigate the impact of FTC on the device characteristics of Polarity Controllable-Ferroelectric-Field Effect transistors (PC-FE-FET). The role of FE in improving immunity against the FTC has also been thoroughly studied to assess the reliability of the proposed device against HCEs. A comparative analysis with conventional PC-FETs has also been done and various device characteristics such as transfer characteristics, drain characteristics, transconductance, transconductance efficiency,  $I_{on}/I_{off}$  ratio and cut-off frequency have been compared for both devices with and without the presence of FTC.

## 2 Device Structure and Simulation Scheme

Figure 1 shows the schematic diagram of Dual-Gate Polarity Controllable-Ferroelectric-Field Effect Transistor (PC-FE-FET) with channel length  $L$  as 40 nm and silicon thickness  $t_{Si}$  as 12 nm. Nickel Silicide with work function 4.45 eV is taken as source and drain regions.  $L_1$ , and  $L_2$  represent the lengths of control gate (CG) and polarity gate (PG),  $L_3$  denotes the length of ungated region present between CG and PG and  $L_4$  represents the distance present between CG or PG edges and Schottky junctions. Here,  $L_1$ ,  $L_2$ , and  $L_3$  are 10 nm each and  $L_4$  is taken as 5 nm, respectively. The thickness  $t_{IFL}$  of interfacial layer (IFL) present between silicon channel and FE layer is assumed as 2 nm. In the present case, Yttrium-doped HFO<sub>2</sub> is chosen as FE insulator with thickness  $t_{FE}$  4 nm [12], and the concentration of NTC is taken as  $-1 \times 10^{12} \text{ cm}^{-2}$  and that of PTC is taken as  $+1 \times 10^{12} \text{ cm}^{-2}$  according to reported range. The device characteristics of PC-FE-FET have been investigated extensively using

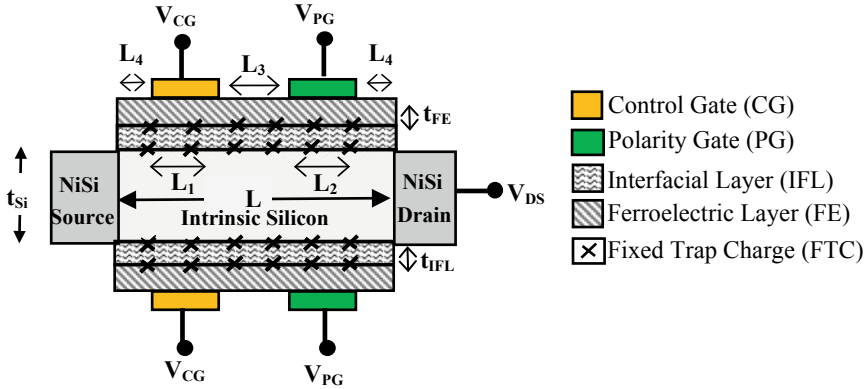


Fig. 1 Schematic of Dual-Gate PC-FE-FET

TCAD Atlas 2D simulation models [13] along with 1-D Landau–Khalatnikov equation [1]. The simulation model involves Shockley-Read-Hall recombination model, auger recombination model, barrier lowering model, non-local band-to-band tunneling model and universal Schottky tunneling model. The effect of FTC has been accounted for in simulation by using interface statement.

Further, in the presence of FTC, the total charge in the channel can be calculated [14] as

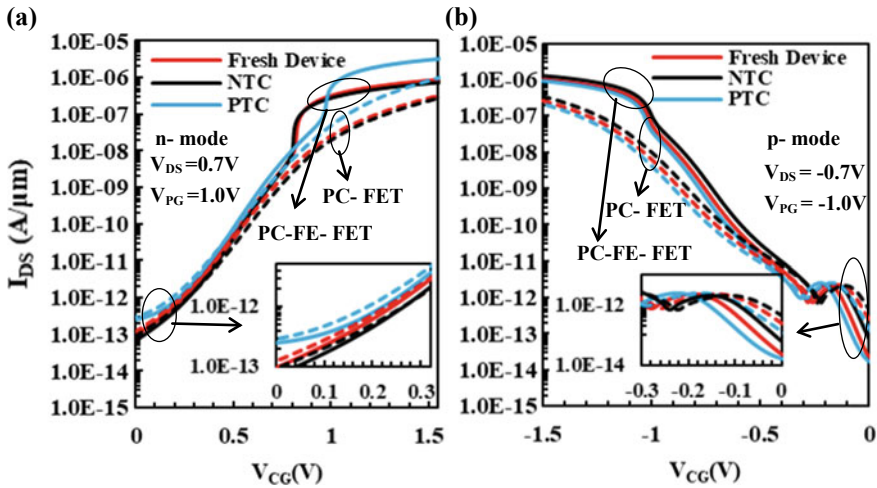
$$Q_t = Q_b + C_{pd}(V_{CG} - V_{FE} - V_{DS}) + C_{ps}(V_{CG} - V_{FE}) + qN_{FTC} \quad (1)$$

where  $Q_b$  is the bulk charge present at internal metal gate electrode,  $V_{CG}$ ,  $V_{DS}$  are the voltages applied at control gate and drain terminals, respectively,  $C_{ps}$ ,  $C_{pd}$  are drain and source capacitances,  $q$  is electronic charge,  $N_{FTC}$  is the concentration of FTC and  $V_{FE}$  is the drop across FE layer [1].

### 3 Results and Discussions

Figure 2 shows the transfer characteristics for both modes of proposed and conventional devices i.e., PC-FE-FET and PC-FET, respectively. It can be seen that for the proposed device, NTC in n-mode and PTC in p-mode leads to minimum subthreshold current in comparison with the other two. Further, the proposed device, PC-FE-FET is offering higher on-state current and lower subthreshold current in comparison with the conventional PC-FET which is due to the presence of FE layer that enhances the device reliability toward FTC and holds potential in digital applications.

Figure 3 shows the output drain characteristics for both the devices for different overdrive voltages  $|V_{ov}| = 0.5 \text{ V}$ ,  $1.0 \text{ V}$  and  $1.5 \text{ V}$ . It can be seen that the proposed device is showing higher drain current values in comparison with conventional device



**Fig. 2**  $I_{DS}-V_{CG}$  for PC-FE-FET and PC-FET devices with **a** n-FET, **b** p-FET modes of operation

for all values of  $|V_{ov}|$ . It is due to the fact that FE layer present in PC-FE-FET enhances the overall gate capacitance which leads to higher values of drain current. Further, for low value of  $|V_{ov}|$  i.e., 0.5 V, it is also noticed that for PC-FE-FET, drain current drops down which implies negative differential output resistance [15].

Figure 4 shows the variation in transconductance ( $g_m$ ) with control gate voltage for both modes of proposed and conventional devices. As earlier seen in Fig. 2, PC FeFET delivers higher on-state current, as a result of which higher  $g_m$  values are obtained for proposed device in comparison with conventional PC-FET. Further, it can also be seen that both PC-FE-FET and PC-FET, in the presence of PTC (NTC) show higher peak  $g_m$  values when operated in n- (p-) mode. It is due to the fact that PTC in n-mode and NTC in p-mode exhibits much higher on-state current than the other two.

Figure 5 shows the variation in transconductance efficiency ( $g_m/I_{DS}$ ) with control gate voltage for both the devices. Higher  $g_m$  values exhibited by PC-FE-FET also translate into higher  $g_m/I_{DS}$  values in comparison with PC-FET. Further, due to the presence of FE layer, the increase in  $g_m/I_{DS}$  in PC-FE-FET is much more prominent as compared to conventional device where the change in peak values is marginal. Moreover, it has also been observed for both the devices that NTC in n-mode and PTC in p-mode are delivering higher  $g_m/I_{DS}$  values over the fresh device and thus enhances immunity toward HCEs.

Figure 6a shows  $I_{on}/I_{off}$  ratio for different fixed trap charge densities for both devices.  $I_{on}$  is calculated at  $V_{CG} = 11.11$  V,  $V_{PG} = 11.01$  V and  $V_{DS} = 10.71$  V whereas  $I_{off}$  is calculated at  $V_{CG} = 0$  V,  $V_{PG} = 11.01$  V and  $V_{DS} = 10.71$  V. It can be observed that for both PC-FE-FET and PC-FET, device with NTC in n-mode operation and PTC in p-mode operation offers higher values of  $I_{on}/I_{off}$  along with least value of off-state current ( $I_{off}$ ) which is quite evident from Fig. 6b. Further, it can also be noticed that

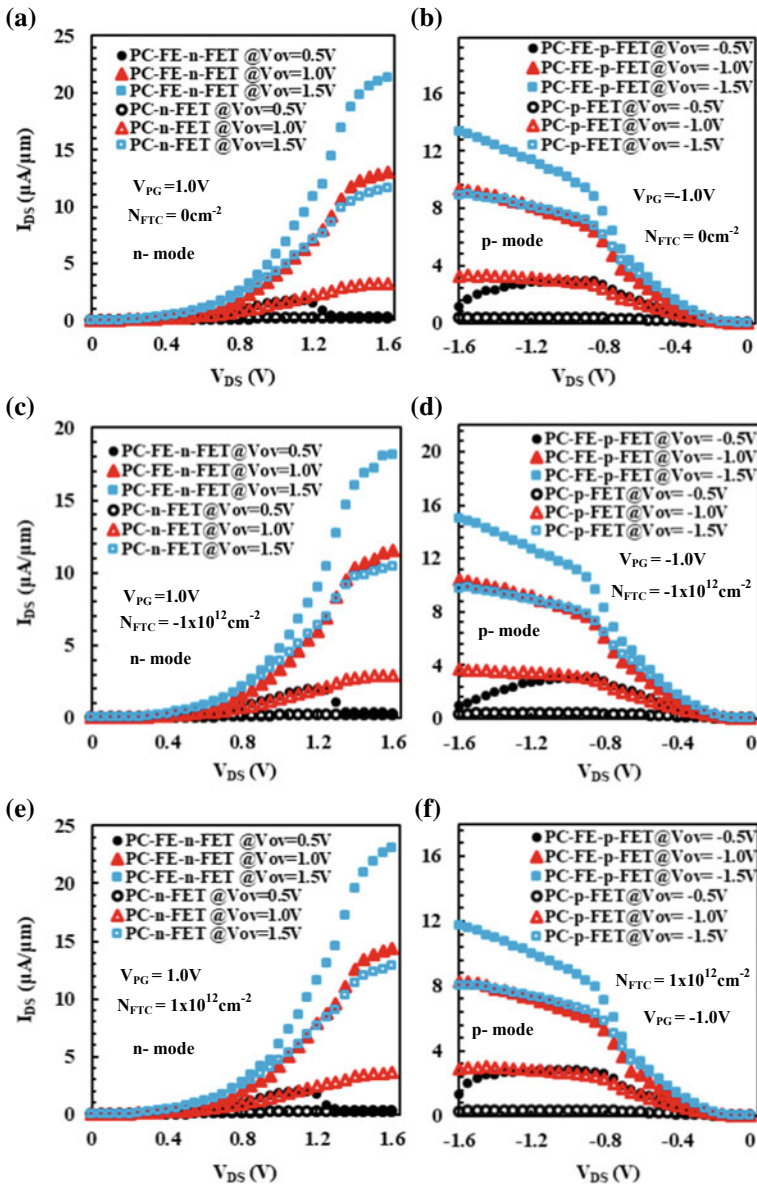
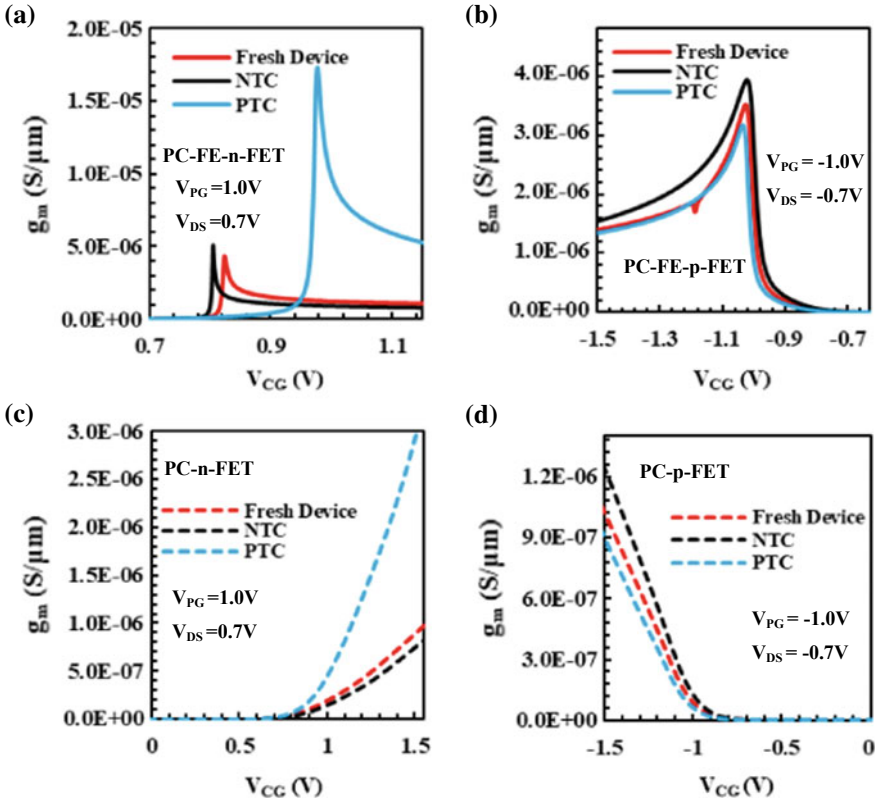


Fig. 3 Output drain characteristics for both PC-FE-FET and PC-FET devices for a, b fresh device, c, d NTC, e, f PTC respectively

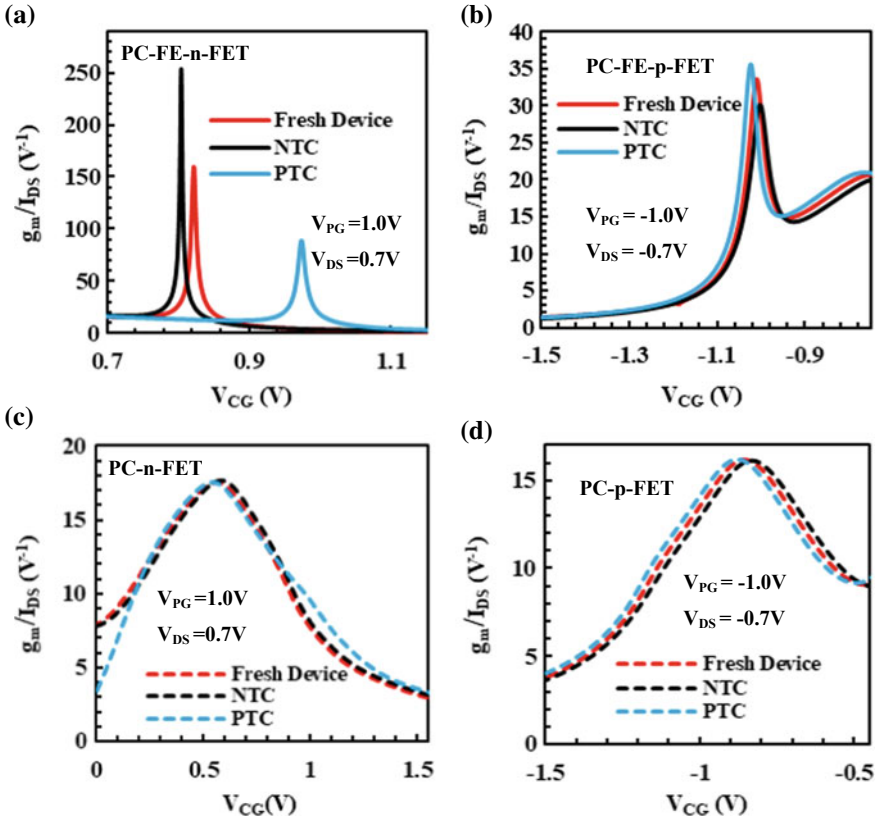


**Fig. 4** Transconductance with control gate voltage in **a** PC–FE–n–FET, **b** PC–FE–p–FET, **c** PC–n–FET and **d** PC–p–FET for fresh and damaged devices

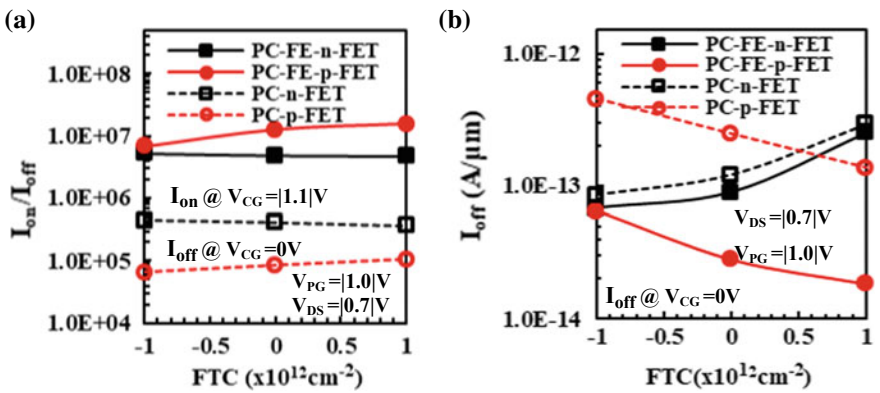
due to the presence of FE layer, proposed device exhibits higher current drivability in comparison with the conventional device and thus, PC–FE–FET in both modes of operation offer higher values of  $I_{on}/I_{off}$  and lower values of  $I_{off}$  in comparison with PC–FET.

Figure 7 shows the variation in cut-off frequency ( $f_T$ ) with drain current for both the devices. It is quite clear that due to the FE layer, higher  $g_m$  values are obtained as can be seen from Fig. 4 and this translates to higher  $f_T$  values which are obtained for PC–FE–FET in comparison with PC–FET. Further, it can also be noted for both PC–FE–FET and PC–FET that devices with PTC (NTC) in n- (p-) mode exhibits higher  $g_m$  values and thus demonstrates increased  $f_T$  values in comparison with the other two and thus can be extended to high frequency-switching applications.

Thus, it can be summarized that even in the presence of traps, the polarity controllable device with FE layer, i.e., PC–FE–FET exhibits superior subthreshold characteristics along with output drain characteristics and shows potential for low power digital/analog applications.



**Fig. 5** Transconductance efficiency with control gate voltage in **a** PC–FE–n–FET, **b** PC–FE–p–FE, **c** PC–n–FET and **d** PC–p–FET for both fresh and damaged devices



**Fig. 6** Variation of **a**  $I_{on}/I_{off}$ , **b**  $I_{off}$  with fixed trap charge density for both PC–FE–FET and PC–FET



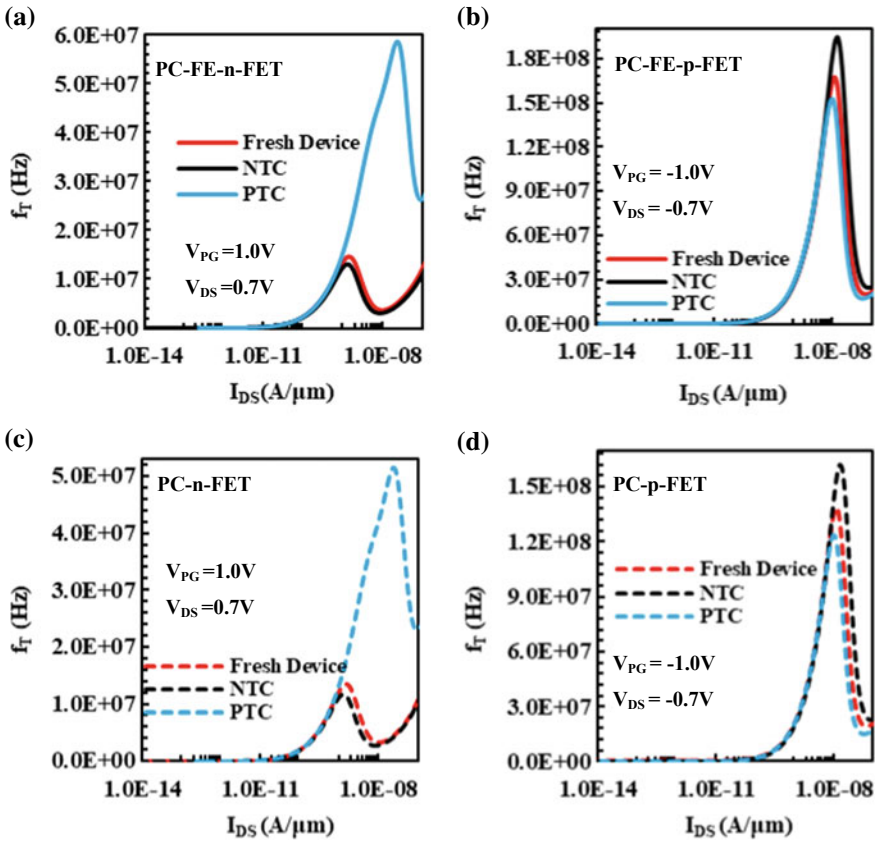


Fig. 7 Cut-off frequency with drain current for a PC-FE-n-FET, b PC-FE-p-FET, c PC-n-FET and d PC-p-FET for both fresh and damaged devices

## References

1. Salahuddin S, Dutta S (2008) Use of negative capacitance to provide voltage amplification for low power nanoscale devices. *Nano Lett* 8(2):405–410
2. Rusu A, Salvatore GA, Jimenez D, Ionescu AM (2010) Metal-ferro-metal-oxide-semiconductor field effect transistor with sub-60 mV/decade subthreshold swing and internal voltage amplification. In: 2010 international electron devices meeting, pp 16.3.1–16.3.4, IEEE, San Francisco, CA, USA
3. Zhu G, Zhou X, Ang LK, Lee TS, See GH, Lin S, Chin YK, Pey KL (2009) A compact model for undoped silicon-nanowire MOSFETs with Schottky barrier source/drain. *IEEE Trans Electron Devices* 56(5):1100–1109
4. Larson JM, Snyder JP (2006) Overview and status of metal S/D Schottky barrier MOSFET technology. *IEEE Trans Electron Devices* 53(5):1048–1058
5. Heinzig A, Slesazek S, Kreupl F, Mikolajick T, Weber WM (2011) Reconfigurable silicon nanowire transistors. *Nano Lett* 12(1):119–124
6. De Marchi M, Sacchetto D, Frache S, Zhang J, Gaillardon P-E, Leblebici Y, De Micheli G (2012) Polarity control in double-gate, gate-all-around vertically stacked silicon nanowire

- FETs. In: 2012 international electron devices meeting, IEEE, San Francisco, CA, USA
7. Kang H, Han J-W, Choi Y-K (2008) Analytical threshold voltage model for double-gate MOSFETs with localized charges. *IEEE Electron Device Lett* 29(8):927–930
  8. Sato M, Kamiyama S, Matsuki T, Ishikawa D, Ono T, Morooka T, Yugami J, Ikeda K, Ohji Y (2010) Study of a negative threshold voltage shift in positive bias temperature instability and a positive threshold voltage shift the negative bias temperature instability of yttrium-doped hfo2 gate dielectrics. *Jpn J Appl Phys* 49(4S):04DC24-1–04DC24-5
  9. Bracchitta JA, Honan TL, Anderson RL (1985) Hot-electron induced degradation in MOSFET's at 77 K. *IEEE Trans Electron Devices* ED-32(9):1850–1857
  10. Yoon JS, Jeong J, Lee S, Baek R-H (2019) Metal source-/drain-induced performance boosting of sub-7-nm node nanosheet FETs. *IEEE Trans Electron Devices* 66(4):1868–1873
  11. Li Y, Zhang J, Yang J, Yuan Y, Hu Z, Lin Z, Song A, Xin Q (2019) Complementary integrated circuits based on n-type and p-type oxide semiconductors for applications beyond flat-panel displays. *IEEE Trans Electron Devices* 66(2):950–956
  12. Schroeder U, Yurchuk E, Muller J, Martin D, Schenk T, Polakowski P, Adelman C, Popovici MI, Kalinin SV, Mikolajick T (2014) Impact of different dopants on the switching properties of ferroelectric hafnium oxide. *Jpn J Appl Phys* 53(8S1):08LE02-5–08LE02-5
  13. ATLAS: 2-D device simulator (2014), Version 5.19.20.R, Silvaco, Santa Clara, CA, USA
  14. Dong Z, Guo J (2017) A simple model of negative capacitance FET with electrostatic short channel effects. *IEEE Trans Electron Devices* 64(7):2927–2934
  15. Zhou J, Han G, Li J, Liu Y, Peng Y, Zhang J, Sun Q-Q, Zhang DW, Hao Y (2018) Negative differential resistance in negative capacitance FETs. *IEEE Electron Device Lett* 39(4):622–625



# Particle Swarm Optimization for Training Artificial Neural Network-Based Rainfall–Runoff Model, Case Study: Jardine River Basin



Vikas Kumar Vidyarthi and Shikha Chourasiya

**Abstract** The use of artificial neural network (ANN) in estimating runoff of a river is popular among hydrologists and scientist from a long time. The classical gradient descent algorithm (GD) is the most commonly used algorithm for training the ANN runoff models so far. The performance of GD algorithm, however, is affected by chances to get stuck at the local minimum. In this paper, one of the popular evolutionary optimization algorithms, known as particle swarm optimization (PSO), has been explored to train the ANN rainfall–runoff model. The superiority of the PSO over the GD method in training ANN rainfall–runoff model is illustrated using data from a real catchment. On the basis of various error statistics, it has been observed that particle swarm optimization can be very effective optimizer in developing ANN-based models for water resources applications, especially in modeling rainfall–runoff process.

**Keywords** Water resources · Rainfall–runoff model · Artificial neural network

## 1 Introduction

Most of the water resources problems are highly non-linear and dynamic in nature. One of them is estimation of runoff in a river. The physics-based techniques, which are used for modeling these processes, provide better comprehensibility, but are poor in accuracy. The data-driven techniques, on the other hand, are highly accurate for modeling such processes [9]. The artificial neural network (one of the data-driven techniques) is very popular among hydrologists and scientists for estimation and forecasting different water resources variables [2, 6] from last couple of decades.

---

V. K. Vidyarthi (✉)  
SRM IST NCR Campus, Ghaziabad, India  
e-mail: [vikas.civil@gmail.com](mailto:vikas.civil@gmail.com)

S. Chourasiya  
Department of Civil Engineering, IIT Kanpur, Kanpur, India

Feed-forward and backpropagation (FFBP) network is one of the most popular networks in the development of ANN models so far. For optimal set of ANN parameters, the classical gradient descent optimization technique is used in the standard FFBP network [8]. Maier et al. [6] in review article revealed that more than 85% of the previously published works on ANN models in hydrology use the popular gradient descent (GD) algorithm for network training. The gradient-based optimization techniques have always a chance to get stuck in the local minima [1] resulting in poor performance of ANN models. There are several optimization techniques available, which are based on evolutionary principle and reaches to the region of global optimum [1]. There is a strong need to explore these evolutionary optimization techniques in training the ANN to enhance the performance of runoff estimation [6].

Particle swarm optimization (PSO) is one of the evolutionary optimization algorithms used to optimize non-linear system with high accuracy. PSO is invariably used in many fields of science and technology [7].

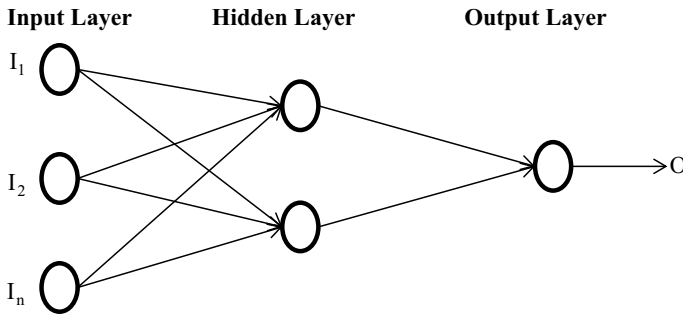
The objectives of the present work are to (a) develop ANN model for rainfall–runoff process and (b) compare the strength of PSO as training algorithm with classical GD algorithm in ANN rainfall–runoff model. The daily rainfall and runoff data derived from Jardine River basin, Australia, have been employed to develop all the models in this study. The paper starts with the brief introduction of ANN and PSO techniques followed by the details of the model development. The paper ends with the concluding remarks.

## 2 Modeling Techniques

In this study, the ANN is used for modeling the rainfall–runoff process. Two algorithms, namely gradient descent and particle swarm optimization, were used for determining the optimized set of ANN parameters. A brief discussion of these techniques is provided in this section.

A simple feed-forward ANN as shown in Fig. 1 consists of three layers: input, hidden, and output layers were used in this study. The neurons in the adjacent layers are connected with ‘weight’ which represents the strength of the connection.

There are two basic steps in ANN: (1) feed-forward step in which the inputs are fed in the forward direction through non-linear activation function at hidden and/or output layers and (2) backpropagation in which the errors at the output layer are propagated back in reverse direction to obtain the optimal set of parameters that produces outputs equal to or closer to the targets. To obtain the optimal set of parameters, the GD algorithm (used in backpropagation algorithm in ANN) has been extensively used so far, but any other suitable optimization techniques can be used.



**Fig. 1** ANN architecture

## 2.1 Gradient-Descent Algorithm

The gradient descent, used for training of the ANN, is a first-order iterative optimization algorithm for finding the minimum of cost/objective function. In this method, as the local minimum is searched in the direction negative of the gradient of the function, the objective function needs to be continuous and differentiable. The GD-based ANN is also known as the standard backpropagation algorithm and the detailed description can be found in [8].

## 2.2 Particle Swarm Optimization (PSO)

The PSO was first proposed by Kennedy and Eberhart [4] meant for simulating social behavior. It is a meta-heuristic evolutionary search optimization algorithm. It can search very large spaces of candidate solutions without making any assumptions about the problem being optimized. As PSO is population-based algorithm, it does not use gradient for optimum search, henceforth, the condition of continuity and differentiability for cost/objective function is not required as required in the gradient-based classical optimization methods. Basically, the PSO algorithm works by having a population, called as swarm, of candidate solutions, called as particles. At every iteration, each particle moves in the direction of its own best position, as well as in the direction of the global best position discovered by any of the particles in the swarm, meaning thereby, all the particles will move closer to the particle, which discovers a new solution better than the previous solutions, resulting in movement toward the global best. The detailed description of PSO algorithm and its application can be found in Kennedy and Eberhart [5], Van den Bergh and Engelbrecht [10], and Poli [7].

### 2.3 Model Evaluation Statistics

The performances of the models developed in this study are evaluated using standard error statistics. The various error statistics used in this study are average absolute relative error (AARE), correlation coefficient ( $R$ ), normalized root mean square error (NRMSE), and threshold statistics (TS10, TS25, TS50, TS75, and TS100) also used extensively by many researchers in the past [3, 9].

## 3 Model Development

Two ANN models have been developed in this study: First uses the GD algorithm for its training referred as ANN-GD, and another uses PSO technique referred as ANN-PSO. The explanation of these algorithms is provided in the previous section. The explanation of the study area and the data used in this study are as follows.

#### *Study Area and Data*

The rainfall and runoff data derived from Jardine River basin, Queensland, Australia, were used to illustrate the proposed methodology in this study. The area of the Jardine River basin is 2500 km<sup>2</sup>.

The daily rainfall and runoff data of 16 years from January 1974 to December 1989 were used for development of all the models in this study. The division of the data was obtained by hit and trial basis, and it was found that the division of the entire data: 60% for training, 10% for validation, and rest 30% for testing was found to be suitable for development of all the models in this study. The basic statistics of the data set are presented in Table 1. The data are normalized between 0.1 and 0.9.

In the first step of ANN model development, the significant input variables were selected on the basis of autocorrelation function (ACF) and cross-correlation function (CCF) values between the runoff at the outlet of Jardine River basin and the lagged rainfall and runoff data.

**Table 1** Basic statistics of the data set

	Flow (mm/d)		Rainfall (mm)	
	Train	Test	Train	Test
Minimum	0.37	0.18	0.00	0.00
Maximum	21.41	12.09	162.60	135.80
Average	2.53	2.21	4.59	4.42
Std. dev.	2.60	2.14	12.13	12.17

The lag up to 10 days was considered. The inputs having  $ACF > 0.9$  and  $CCF > 0.39$  were included in the input variable set. Thus,  $P(t - 1)$ ,  $P(t - 2)$ ,  $P(t - 3)$ ,  $P(t - 4)$ ,  $Q(t - 1)$ , and  $Q(t - 2)$  were considered as input variables for ANN model development in this study.

*Model Training*

Three-layered ANN consists of one input, one hidden, and one output layers was used for developing all the models in this study. The sigmoid function is used as activation function. In model training, the connection weights of the ANN are iteratively adjusted such that the error at the output layer is minimized resulting in the best agreement between the predicted and observed data using any optimization algorithm. In this paper, PSO (an evolutionary optimizer) is used for training the ANN-PSO model, and the strength of PSO is compared with that of GD (a classical gradient-based technique) algorithms in ANN-GD model. The details of PSO and GD algorithms are already discussed previously.

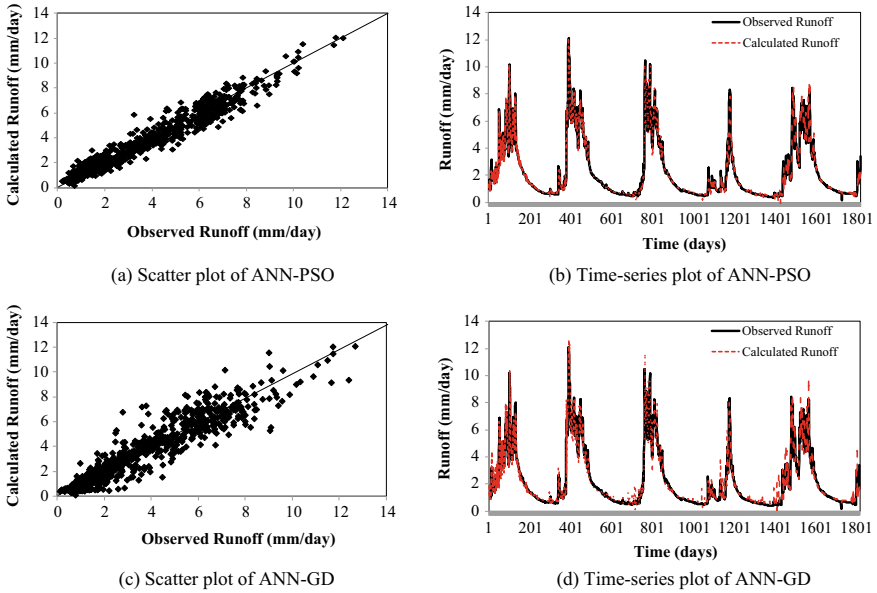
Three parameters of PSO: inertia weight ( $w$ ), social parameter ( $P_1$ ), and cognitive parameter ( $P_2$ ) were determined by hit and trial method. The GD algorithm parameters—learning rate ( $\eta$ ) and momentum correction factor ( $\alpha$ )—were also determined by the hit and trial method. The population or swarm is kept as 400. The values of  $w = 0.7298$ , and  $P_1 = P_2 = 1.49618$ , thus, obtained for PSO parameters, were found to be the best for training ANN-PSO model. Likewise, the values of  $\eta = 0.9$  and  $\alpha = 0.01$  of GD algorithm were found to be the best for training ANN-GD model in this study. The maximum iteration was fixed at 50,000. The architectures for both ANN-GD and ANN-PSO were fixed by trial and error procedure. The number of neurons in the hidden layer was varied from 1 to 20, and the architecture with the minimum mean square error (MSE) so obtained during training was considered as the best architecture and further considered for testing. The best architectures for ANN-GD and ANN-PSO so obtained were 6-4-1 and 6-2-1, respectively.

*Results and Discussion*

The strength of the two training algorithms used in developing ANN-GD and ANN-PSO was compared with each other. The model performance in terms of various error statistics during training and testing is presented in Table 2. Analyzing the

**Table 2** Performance statistics of various models

Models	AARE	R	NRMSE	TS25	TS50	TS75	TS100
<i>Training performance</i>							
ANN-PSO	14.0	0.98	0.120	30.0	51.1	81.5	98.0
ANN-GD	17.5	0.97	0.135	23.1	46.2	78.9	95.7
<i>Testing performance</i>							
ANN-PSO	15.0	0.98	0.110	23.0	48.1	81.3	96.0
ANN-GD	18.5	0.96	0.124	20.9	43.2	77.9	94.2



**Fig. 2** Scatter and time-series plots during testing period

results from Table 2 during training, it can be observed that the ANN-PSO model provides the best AARE (14.0),  $R$  (0.98), NRMSE (0.120), and various TS values. During testing also, the ANN-PSO performed better as compared to ANN-GD model in terms of all the error statistics. Overall, it can be statistically observed that the ANN model which uses PSO as training algorithm performed better than the ANN model which uses GD algorithm for training. The scatter and time-series plots during testing period are shown in Fig. 2, which shows that the runoff obtained from ANN-PSO model provides better agreement with the observed runoff which justifies the performance statistics obtained.

## 4 Summary and Conclusion

This paper investigated the strength of one of the evolutionary optimization algorithms, known as particle swarm optimization (PSO) for training artificial neural network in modeling rainfall–runoff process. The performance of PSO in training ANN was also compared with the results obtained from the ANN trained with classical gradient descent (GD) technique. The daily rainfall and runoff data derived from Jardine River basin, Australia, were employed to develop all the models in this study. The performances of the models were evaluated on the basis of various statistical measures.

The findings of this paper suggest that the runoff predicted by ANN model trained with PSO was significantly better than the ANN rainfall–runoff model trained using GD algorithm. This paper illustrated the application of PSO for training ANN model in solving one of the complex problems in water resources, but it can be further utilized in any other field to solve complex processes/systems. The effectiveness of PSO and other similar evolutionary algorithms for training the ANN models need to be explored further in order to exploit their advantages over the classical techniques in ANN training. It is hoped that the future research efforts will focus in improving the accuracy of the models using the evolutionary algorithms.

## References

1. Deb K (2001) Multi-objective optimization using evolutionary algorithms. Wiley, Chichester, UK
2. Govindaraju RS, Rao AR (2000) Artificial neural networks in hydrology. Springer, Netherlands
3. Jain A, Srinivasulu S (2006) Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques. *J Hydrol* 317(3–4):291–306
4. Kennedy J, Eberhart RC (1995) Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks, NJ, vol 4, pp 1942–1948
5. Kennedy J, Eberhart RC (2001) Swarm intelligence. Morgan Kaufmann, Burlington. ISBN 978-1-55860-595-4
6. Maier HR, Jain A, Dandy GC, Sudheer KP (2010) Methods used for the development of neural networks for the prediction of water resources variables in river systems: current status and future directions. *Environ Model Softw* 25:891–909
7. Poli R (2008) Analysis of the publications on the applications of particle swarm optimisation. *J Artif Evol Appl* 1–10
8. Rumelhart DE, Hintont GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323(6088):533–536
9. Srinivasulu S, Jain A (2009) River flow prediction using an integrated approach. *J Hydrol Eng ASCE* 14(1):75–83
10. Van den Bergh F, Engelbrecht AP (2004) A cooperative approach to particle swarm optimization. *IEEE Trans Evol Comput* 8(3):225–239

# Text Generation Using Long Short-Term Memory Networks



Ishika Dhall, Shubham Vashisth and Shipra Saraswat

**Abstract** The domain of natural language processing has lately achieved exceptional breakthroughs especially after the origination of the deep neural networks. This has enabled the machine learning engineers to develop such deep models that are capable of performing high-level automation, empowering computer systems to interact with the humans in a competent manner. With the usage of special types of deep neural networks known as recurrent neural networks, it is possible to accomplish various applications in the domain of natural language processing including sentiment analysis, part-of-speech tagging, machine translation, and even text generation. This paper presents a deep, stacked long short-term memory network, an advanced form of recurrent neural network model which can generate text from a random input seed. This paper discusses the shortcomings of a conventional recurrent neural network hence bringing forward the concept of long short-term memory networks along with its architecture and methodologies being adopted.

**Keywords** Long short-term memory networks · Recurrent neural networks · Natural language processing · Text generation · Machine learning

## 1 Introduction

Text generation is one of the most significant applications that the domain natural language processing and machine learning is aiming to crack. Text generation or natural language generation is a process of generating a meaningful text deliberately to achieve some specific communication goals. Text generation techniques are used to perform automatic letter writing, automatic report generation, automatic documentation systems, etc.

The goal of text generation is to empower computer machines to recognize data patterns of text vocabulary which further produces understandable human language. Information (nonlinguistic) is fed to system as input and the expected output can

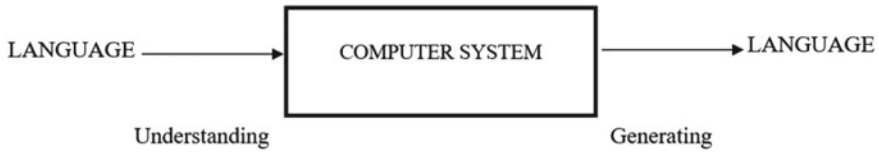
---

I. Dhall (✉) · S. Vashisth · S. Saraswat  
Department of CSE, Amity University, Sector-125, Noida, Uttar Pradesh, India

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_66](https://doi.org/10.1007/978-981-15-2329-8_66)

649





**Fig. 1** Basic model of text generating systems

simply be the text, tables, graphics, plain ASCII or formatted LaTeX, RTF and HTML.

On a higher abstraction level, it is easy to visualize the working of a text generating system (see Fig. 1). However, the traditional methods like Pollen Forecast as presented in [1] for Scotland system and others were not able to show many promising results as they worked on simpler information and were not robust to complex data. The recent development in the field of machine learning and deep learning has given rise to various advance forms of deep neural network like for instance various forms ANNs which are being used to crack problems of almost every domain from classifying ECG signals [2] to problems that come under the domain of computer vision using CNNs and NLP using RNNs.

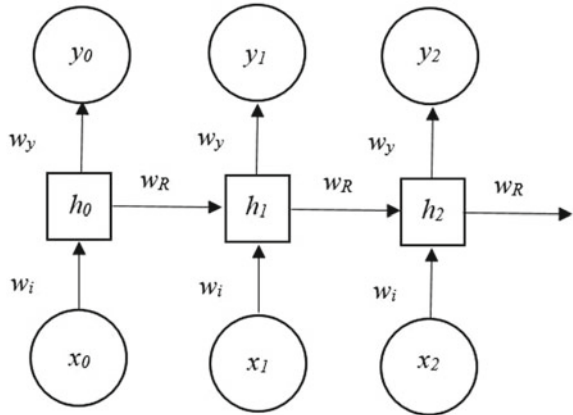
This paper contributes by presenting an approach to text generation for a random seed by constructing a deep and stacked long short-term memory network model. The trained model flexibly predicts the text as per the number of characters that are provided to it as input. The paper also includes the methodology and architecture of the LSTM model. Another contribution of the paper includes the complications with the conventional RNN model which resulted in the LSTM networks coming into picture.

The rest of the paper is organized as mentioned: Sect. 2 presents the preliminaries; Sect. 3 presents the related works; Sect. 4 is proposing the methodologies and Sect. 5 discusses the result. Finally, conclusion is discussed in Sect. 6.

## 2 Preliminaries

### 2.1 Recurrent Neural Networks

An recurrent neural network (RNN) is special type of an artificial neural network (ANN). It was premeditated to accomplish pattern recognition on sequential form data; an example of such data can be text or numerical time-series or genomes and handwriting, etc. RNN can work in such situations where feed-forward neural network fails, in case of a feed-forward neural network, all set input ( $X$ ) and output ( $Y$ ) are independent concerning each other. Therefore, in scenarios where the need is to predict the following word in sentence, a traditional feed-forward neural network

**Fig. 2** Working of an RNN

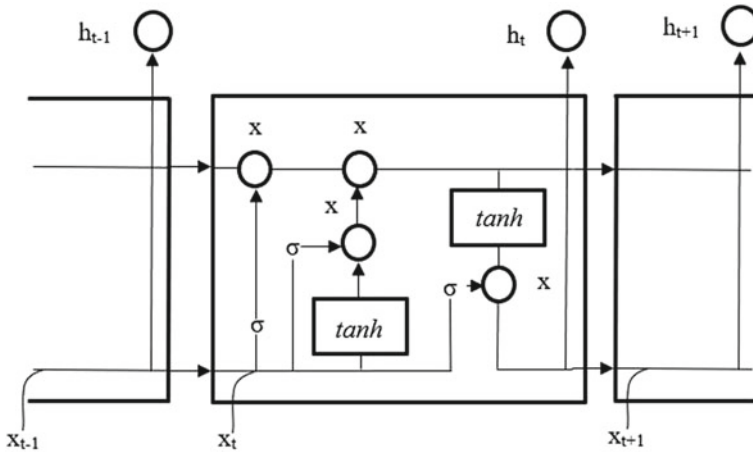
will not consider the previous set of output to forecast the very subsequent word in a sentence.

As we know, a deep neural network consists of many hidden layers and each hidden layer comprises of its own set of weights and biases, e.g.,  $(w_0, b_0)$ ,  $(w_1, b_1)$ , etc. On the contrary (see Fig. 2), RNN can change the independent activations to dependent activations by providing all the layers the same set of weights and biases. This allows the joining of layers together with the same set of weights and biases into a single recurrent layer.

In order to train the RNN as presented in [3], input is provided to the RNN in the form of a solitary time step then the current state is calculated using set of previous states and the current input. For the next time stamp ( $h_t \rightarrow h_{t-1}$ ) depending upon the problem for which the network is being designed, one can take multiple steps and join the information from all the previous states. After the completion of the time steps, the output is evaluated using the final current state. The error is evaluated by comparison with the output to the target output. The use of back-propagation algorithm is considered to update the weights of the RNN for each and every iteration.

## 2.2 Long Short-Term Memory Networks (LSTMs)

LSTMs or long short-term memory network [4] is a unique and superior category in recurrent neural networks (RNNs). It is proficient to acquire long-term dependencies, i.e., in the context of text generation, while predicting a word in the sentence which uses information that is not dependent on the very previous word rather uses some information in the text that is way too far from the current word position, such dependency is known as long-term dependency. LSTM solves the issue of “vanishing gradient” that is faced in the case of standard recurrent neural network (RNN).



**Fig. 3** Working of a long short-term memory network

An RNN follows a simple chain structure which consists of a simple repeating module consisting of a squashing activation function like  $\tanh$ . On the other hand, LSTM although also follows a chain alike construction but the repeating modules consist of a dissimilar structure pattern (see Fig. 3).

### 3 Related Work

Recurrent neural networks (RNNs) are one of the most prevailing models in the domain of deep learning which do not exhibit a generic nature. The conventional architecture of RNNs is not suited for tasks of character-level modeling. Therefore, to overcome the problems [5] proposed a modified variant of RNN which resolved their training problems by application of Hessian-free optimization technique and introducing gated or multiplicative connections. A long short-term memory network (LSTM) is a special form of recurrent neural network that is designed for the task of compound sequence generation. These networks follow a long-range of construction which works by prediction of one data-point at a time. The model demonstrated in [6] can perform efficient synthesis of cursive handwriting in an extensive diversity of styles. Another approach to text generation is by using neural checklist models [7]. This model generates output by vigorously adjusting the interpolation into a model. RNN is difficult to train and it is dubious to show the full potential of a given RNN model and in order to address these problems, [8] used the long short-term memory neural network architecture to examine it on a task of modeling an English and a French language. It showed an improvement of around 8% over other existing RNNs. [9] proposed two innovative approaches to text generation which encodes the contexts into a continuous semantic illustration and then decodes the semantic

illustration into text categorizations with RNNs. Since RNNs have its own hitches like vanishing gradients, the following research will depict how using LSTMs can overcome its shortcomings.

## 4 Methodology

The initial phase in an LSTM is to evaluate that form of information which is not needed and will be neglected from the state of the cell. The second phase is to identify whether the recent information will be maintained to the cell state which is carried out by following some steps. Further, a tanh layer creates new candidate values in a form of that could be updated to the state. Hence, using this mathematical model, LSTMs can perform text generation precisely and efficiently.

### 4.1 Data Collection and Preprocessing

To practically implement the same, data acts the initial step. This data must be in the format of ASCII text and will act as the fuel for our long short-term memory network. The type of text data that is used to train the data will be the governing factor while performing text generation. For this model “Alice in Wonderland.txt” file is provided as input to the network which comprises of 1,63,780 characters.

Data preprocessing is required in order to have filtered and clean data that can directly be fed to our neural network. For example, the text data is converted to lower case and mapping of unique characters to integers is performed before designing a long short-term memory network.

### 4.2 Constructing the LSTM Model

The goal is to build a deep network by stacking LSTM layers in order to enable the model to learn complex and long sentences in an efficient manner. This deep LSTM model can be built by using various libraries and its functions like Keras, TensorFlow, Theano, Cntk, etc. The LSTM as discussed in [10] will consist of many different layers along with activation function and a type of learning optimizer which is adaptive momentum estimation (Adam) in our case. Figure 4 describes the model summary of the LSTM model that is used to perform text generation.

```
Model: "sequential_1"
Layer (type)                Output Shape                Param #
=====
lstm_1 (LSTM)                (None, 100, 300)          362400
batch_normalization_1 (Batch Normalization) (None, 100, 300)          1200
lstm_2 (LSTM)                (None, 100, 300)          721200
batch_normalization_2 (Batch Normalization) (None, 100, 300)          1200
lstm_3 (LSTM)                (None, 100, 300)          721200
batch_normalization_3 (Batch Normalization) (None, 100, 300)          1200
lstm_4 (LSTM)                (None, 100, 300)          721200
batch_normalization_4 (Batch Normalization) (None, 100, 300)          1200
lstm_5 (LSTM)                (None, 100, 300)          721200
batch_normalization_5 (Batch Normalization) (None, 100, 300)          1200
lstm_6 (LSTM)                (None, 100, 300)          721200
batch_normalization_6 (Batch Normalization) (None, 100, 300)          1200
lstm_7 (LSTM)                (None, 100, 300)          721200
batch_normalization_7 (Batch Normalization) (None, 100, 300)          1200
lstm_8 (LSTM)                (None, 100, 300)          721200
batch_normalization_8 (Batch Normalization) (None, 100, 300)          1200
lstm_9 (LSTM)                (None, 100, 300)          721200
batch_normalization_9 (Batch Normalization) (None, 100, 300)          1200
lstm_10 (LSTM)               (None, 300)                721200
batch_normalization_10 (Batch Normalization) (None, 300)                1200
dense_1 (Dense)              (None, 58)                  17458
=====
Total params: 6,882,658
Trainable params: 6,876,658
Non-trainable params: 6,000
```

Fig. 4 Model summary

### 4.3 Training and Hyper-Parameter Tuning

To train the long short-term memory network, the preprocessed data is fed to our designed LSTM model. The LSTM fits the data and accordingly updates its weights and biases. Training phase is the crux of Deep learning task as this is where learning happens. Therefore, this phase is computationally expensive and thus requires a lot of resources and computational power. In order to avoid overfitting batch normalization is performed after every LSTM layer. The model was trained on Google Colab platform using the GPU support; it took the model approximately 22 h to complete its training. While training a LSTM, we need to provide the optimal values of number of epochs and batch size.

Hyper-parameter tuning needs to be done in order to evaluate the optimal values for various parameters such as learning rate, no. of epochs, batch size, quantity of nodes, and quantity of layers.

## 5 Results

The trained stacked long short-term memory model was successfully able to generate text for a randomly generated seed with a testing accuracy of 71.22% as shown in Fig. 5.

The data generated by the LSTM model is quite realistic in nature and interpretable in nature. The final long short-term memory network consisted of the following parameters:

Using the hyper-parameter mentioned in Table 1, long short-term memory can predict next word in a sentence maintaining its memory and history. The achieved testing accuracy of 71.22% of the constructed model may be increased by increasing the no. of epochs and adding a greater quantity of layers/nodes to the current network.

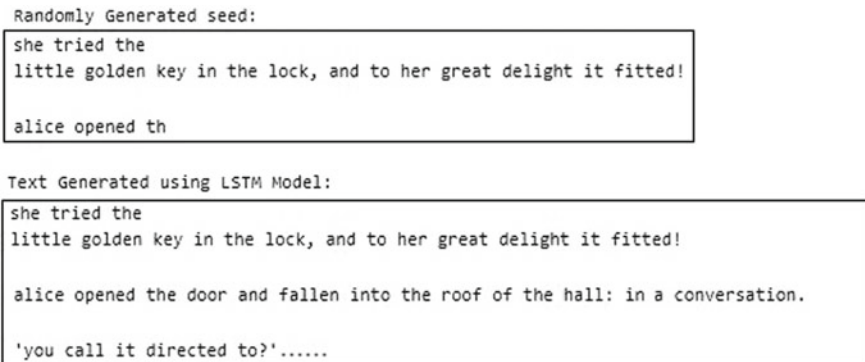


Fig. 5 Sample text generated using LSTM

**Table 1** Hyper-parameters for LSTM model

Parameters (deep LSTM model)	Configuration
Total number LSTM layers	10
Number of nodes (LSTM)	300
Kernel initializer	he_uniform
Optimizer	Adam
Learning rate	0.001
Batch size	1000
Number of epochs	100

## 6 Conclusion and Future Work

LSTM networks have proved to be the best type of model existing till date to perform prediction and classification over text-based data. LSTM is successfully able to resolve the problem faced by the standard recurrent neural networks, i.e., the problem of vanishing gradient. LSTM is an efficient model but it is overall computationally expensive and requires high processing power, i.e., use of GPUs to fit and train the model. They are currently used in various applications like voice assistants, smart virtual keyboards and automated chatbots, sentiment analysis, etc. As future research, the model accuracy of current LSTM models can be surpassed by appending more layers and nodes to the network and applying the notion of transfer learning on the same problem domain.

## References

1. Reiter E (2007) An architecture for data-to-text systems. In: Proceedings of the eleventh European workshop on natural language generation. Association for Computational Linguistics
2. Saraswat S, Srivastava G, Shukla S (2018) Classification of ECG signals using cross-recurrence quantification analysis and probabilistic neural network classifier for ventricular tachycardia patients. *Int J Biomed Eng Technol* 26(2):141–156
3. Mikolov T et al (2010) Recurrent neural network based language model. In: Eleventh annual conference of the international speech communication association
4. Sak H, Senior A, Beaufays F (2014) Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: Fifteenth annual conference of the international speech communication association
5. Sutskever I, Martens J, Hinton GE (2011) Generating text with recurrent neural networks. In: Proceedings of the 28th international conference on machine learning (ICML-11)
6. Graves A (2013) Generating sequences with recurrent neural networks. arXiv preprint [arXiv:1308.08502013](https://arxiv.org/abs/1308.0850)
7. Kiddon C, Zettlemoyer L, Choi Y (2016) Globally coherent text generation with neural checklist models. In: Proceedings of the 2016 conference on empirical methods in natural language processing
8. Sundermeyer M et al (2012) LSTM neural networks for language modeling. In: INTER-SPEECH

9. Tang J et al (2016) Context-aware natural language generation with recurrent neural networks. arXiv preprint [arXiv:1611.09900](https://arxiv.org/abs/1611.09900)2016
10. Li X, Wu X (2015) Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition. In: 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE



# Socio-medic Drone with Integrated Defibrillator



Shivam Pandey, Rahul Kumar Barik, Aritra Karan, P. Phani Kumar, D. Haripriya and N. Kapileswar

**Abstract** The purpose of this paper is to develop a socio-medic drone which can act as a life savior for accident victims in remote areas such as highways or maybe for a person who is dealing with heart issues. During a fatal accident or a cardiac arrest, the first few minutes after the accident or cardiac arrest are the most important which decides between life and death of the individual. This socio-medic drone will be equipped with a first aid kit and also an integrated defibrillator which can act as a great life support to the individual till the ambulance actually arrives through the busy traffic to the ground zero. Real-time parameters can also be checked by the drone such as temperature, heart rate and heartbeat. The values of these parameters can be transmitted to the monitoring hospital database which will initiate the recovery process accordingly.

**Keywords** Socio-medic drone · First aid kit · Integrated defibrillator · Sensors

## 1 Introduction

The increasing population has led to many negative attributes in the normal day-to-day life of a human being. This includes traffic congestion issues, serious hike in health issues, increase in stress levels, etc. This project specially caters to boosting the medical facilities in a more efficient and proficient way with the help of technological advancements in the field of aerial unmanned vehicles.

During a fatal accident, the victim in most of the cases gets unconscious and due to the shock from the accident, the victim may lose his life if the basic measures and first aid are not carried out during the first few minutes after the mishap takes place. The same case appears in terms of cardiac arrests, where the victim should be given the proper facility during the first few minutes otherwise the condition may get worse. The drone with the following facilities can reach the accident zone very quickly and carry out the process till the ambulance actually arrives through the traffic.

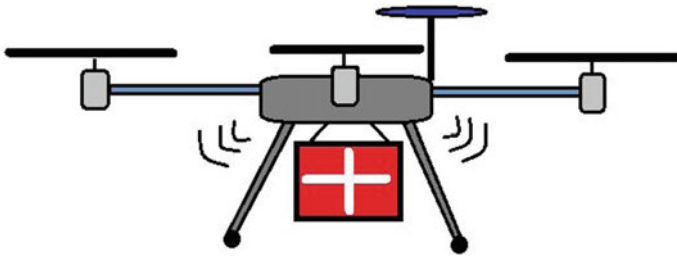
---

S. Pandey · R. K. Barik · A. Karan · P. Phani Kumar (✉) · D. Haripriya · N. Kapileswar  
SRM Institute of Science and Technology, Ramapuram, Chennai, India  
e-mail: [phanikup@srmist.edu.in](mailto:phanikup@srmist.edu.in)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_67](https://doi.org/10.1007/978-981-15-2329-8_67)

659



**Fig. 1** Arrangement of the first aid kit in the socio-medic drone

The speed of the drone is attained by the use of four high-speed brushless motors (900 kV) attached with efficient propellers that produce the required thrust. The drone is powered with the help of one 2200 mAh 3-cell lithium polymer battery with a maximum voltage of 11.1 V. Four ESC's, 30 Amps each are connected to the four rotors for the controlling purpose of its movement. The full system is GPS enabled and also telemetry is used for better understanding of the exact location of the drone. This whole system is designed to be controlled from a control room of the monitoring hospital. This becomes easy due to the installation of the FPV camera in the drone which makes live streaming possible. Figure 1 is a schematic of the socio-medic drone with the arrangements of the first aid kit and integrated defibrillator attached to it.

The paper contains the introduction to the socio-medic drone where we get to know about the actual application of the technology. Section 2 gives a list of all the related works that are being used in this paper to make it more informative. Section 3 contains the methodology and the usage of components that have been used in the following system. Section 4 contains results and discussions where an overall end result has been discussed on the possibilities and limitation of the drone. Section 5 contains the conclusion of the paper.

## 2 Related Works

R. Pahonie proposed a paper with biomechanics of flexible wing drones used for emergency medical transport operations where the usage of CFD simulations to underline the aerodynamic characteristics of a custom-designed small aircraft wing [1]. G. C. Bravo proposed a paper with first aid drone for outdoor sports activities [2], where the drone was to be used for first aid facilities happening during any outdoor activities.

Judy E. Scott proposed the current status of innovative drone delivery with a particular emphasis on healthcare [3]. Joseph Christopher Fancher prepared a paper in the field of current drone technologies where it is reviewed, optimized and used

to demonstrate the feasibility of medical supply delivery to remote areas of China via unmanned aerial vehicles (UAV) [4].

A. Claesson put forward a project on unmanned aerial vehicles (drones) in out-of-hospital-cardiac-arrest where the use of automated external defibrillator (AED) prior to Emergency arrival of ambulance is done through drones which can increase 30-day survival possibilities [5, 6]. Kelly Daly presented, 'How Drones are modernizing the Healthcare Industry'. In this paper, it is clearly defined as to how supplies of blood and medical supplies to organ transplants, drones are helping to save lives [7].

James C. Rosser proposed an article providing the comprehensive review of current and future drone applications in medicine and also how drones can be used for surveillance of disaster sites and areas with biological hazards [8]. Vangara Vamsi Krishna presented a paper on building of a prototype of drone ambulance for assisting the ambulance in saving people's lives. They show how a drone can tackle traffic congestion delays of the ambulance, hence reaching to the accident area and then report to the ambulance by assisting it with necessary parameters such as patient's temperature, state of consciousness, and heart rate. [9].

### 3 Methodology

The world of automation brings up an open mind toward developments in the area of aerial unmanned vehicles such as drones. Now, UAV combined with health care can do wonders in terms of speed, efficiency and new implications. The socio-medic drone is equipped with an integrated defibrillator besides having a first aid kit which works as a life support system.

The main function of an integrated defibrillator is to revert the adverse condition when an un-rhythmic pattern of the heart is observed. This device has the capability to put the heart rhythm to the correct pace and correct pattern. Due to the shock, the heart starts behaving anomalously which to some extent can be cured by the process of defibrillation. Regarding the medics, the quadcopter is attached with a MED-box, temperature or sensor, and basic other sensors that measure the parameters of the patient along with an integrated defibrillator. The drone is also equipped with latest technologies such as Wi-Fi systems and Bluetooth module which keeps it on track in terms of connectivity and file transmitting, receiving and sharing.

The drone can be used as a guide for the ambulance reaching the accident zone at remote locations; an emergency air ambulance can be used to transport blood. A system was also introduced where an air ambulance was used to deliver oxygen cylinders in emergency situations. Figure 2 is a schematic representation of the block diagram of the whole connectivity system wherein all the sensors are connected to the microprocessor.

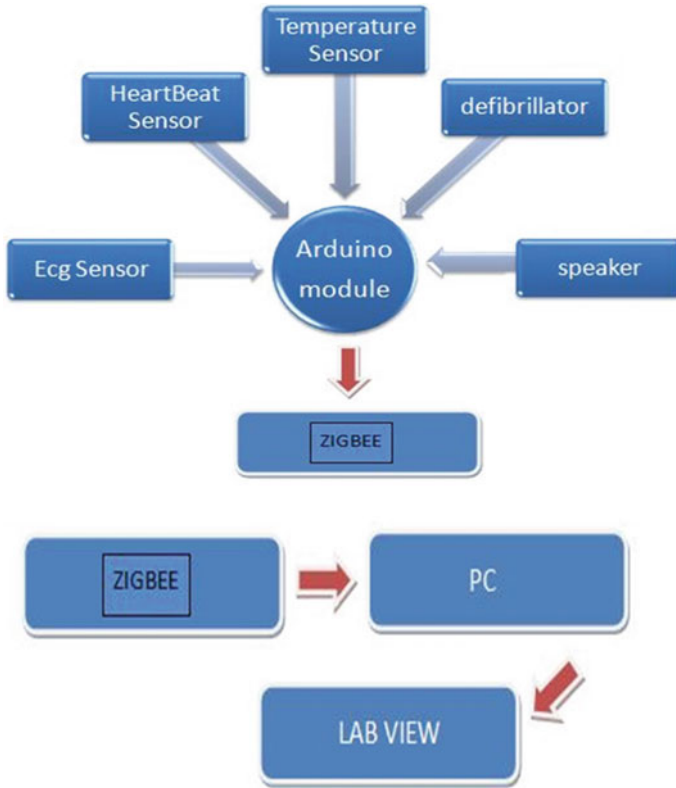


Fig. 2 Block diagrams of the system

### 3.1 Hardware's Required

Figure 3a, b represent a drone frame and the transmitter set respectively to display that is used for controlling the drone. The following are the important elements of the application and their descriptions.

**Drone Frame:** A drone frame is a special frame embossed on a plastic foundation.

**Brushless Motors:** Also known as an EC motor, they produce AC electric current to drive each phase of the motor.

**ESC:** Electronic speed controller is an electronic circuit that controls and regulates the speed of the electric motor.

**Flight Controller:** This is the heart of the flight control system and controls most of the electrical components with the assistance on an Arduino microprocessor.

Figure 4a, b represent a GPS module and the interfacing process simultaneously.

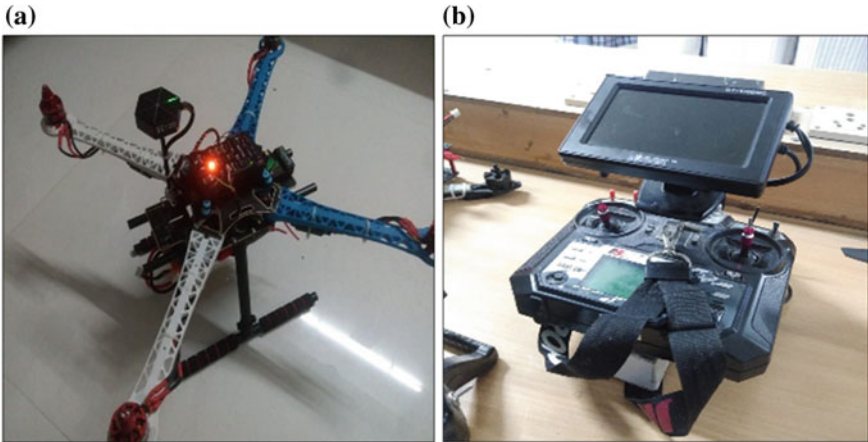


Fig. 3 a Drone frame. b Transmitter set

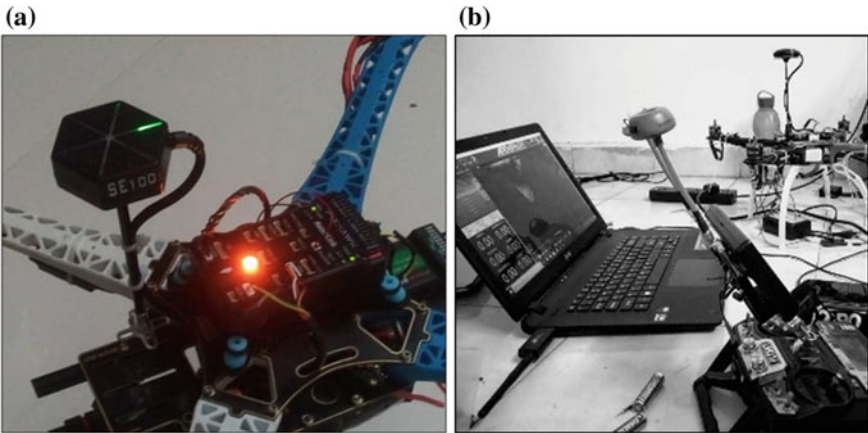


Fig. 4 a GPS module. b Interfacing process

The GPS is used to track the exact position of the drone. The interfacing process is done so as to get all the data from the corresponding sensors without any hindrance to the system.

**GPS for Drone:** It uses the radio receiver to collect signals from the orbiting satellites to determine the speed and position.

**Telemetry Rx Tx:** Used to transfer one or two digital inputs.

**ZigBee Rx Tx:** It is a wireless Tx and Rx module used to trans receive signals.

**Li-on Battery:** This has 3C continuous discharge current suitable for long-time flying.

**Drone Propeller:** Ensures that the drone advances in high power density.

**Rx Tx:** These are the receivers and transmitters.

**Defibrillator:** The life-saving component used in this drone can help people revive during a fatal shock or a cardiac arrest.

**First Aid Kit:** In case of any emergency, these kits will be carried weighing 2 kg approximately.

**Ultrasonic Sensors:** Emits the sound waves which travel through the air and reflects by the objects when it hits.

**Power Distribution for Drones:** Also known as panel board used for electricity supply.

**Camera Assembly:** Used for monitoring the places and upcoming obstacles.

### List of key sensors and devices used for healthcare:

**Integrated Defibrillator** Device used in emergency condition when the victim is facing cardiac arrest, this device has two terminal plates which are known as touch-pads which are stuck on the chest of the victim in a diagonal manner to the heart. It passes an electric pulse through the heart and thus clears the blockage for blood flow, thus the heart starts working again. This small defibrillator is made of electrode pads and is powered by the drone battery through modulated circuit. This defibrillator is integrated within the drone body. It can easily and safely be used by the third party.

The use of the installed defibrillator is monitored by the medical experts and the control the current output from the base station. This communication takes place using ZigBee module used in the communication sector for trans-receiving signals and then the signal is fed into programming board (Arduino) and the signal is modulated or demodulated.

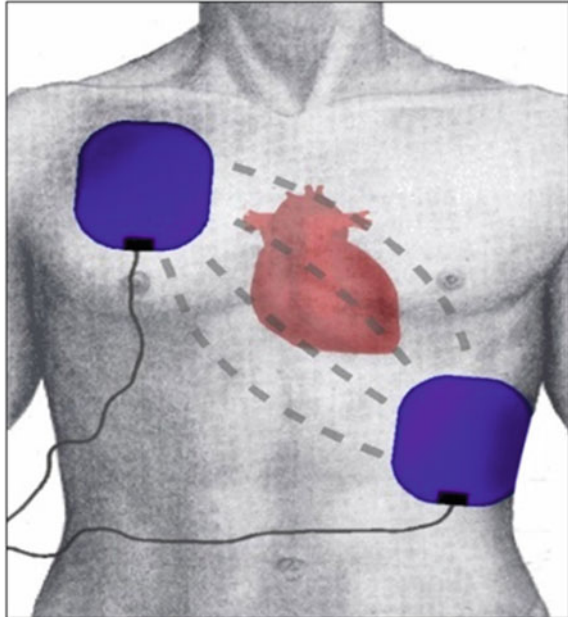
Figure 5 represents a schematic representation of the working of a defibrillator showing the placement of the electrode pad and its placement in the cardiac region.

**Heartbeat Sensor** The condition of the patient can be monitored by analyzing his heartbeat. Thus, we are adding heartbeat sensor with our medical drone. The heartbeat can be felt from the radial nerve in the wrist. Instructions shall be given by the professionals, where to attach the sensor and thus by using the principle of photoplethysmography we can calculate the pulse rate of the person. The sensor works using an optical power variation as light is scattered or absorbed during its path through the blood as the heartbeat changes.

The heartbeat sensor consists of light-emitting diode and a detecting photodiode. The signal sends to the Arduino for analyzing and filtering the signal; the signal is then modulated and transmitted using a ZigBee module to the base station where a medical expert is guiding through other end. The heartbeat sensor module is very easy to place on the human body for monitoring the heart rate it can be either placed on fingers or wrist surface since it has dedicated power supply there are very less chances of external noise and disturbing frequency.

**Temperature Sensor** The body temp of the patient must be monitored for knowing his condition. Thus, we are using infrared temperature sensors for this purpose.

**Fig. 5** Defibrillator electrode pad placement. <https://en.wikipedia.org> [10]



The sensor provides an accurate non-contact measurement. The sensor is operated from the station to know the condition of the temp of the patient and the values are taken and analyzed and appropriate suggestion is given the experts on the other hand monitoring suggest medic to give from the smart first aid box attach with the drone.

The sensor is made of multiple thermocouples on a silicon chip to measure an objects infrared energy. The GTPCO-003 thermopile temperature sensor is used for sensing the temperature. The sensor produces accurate value and the signal is delivered to the modulating unit (Arduino) placed in the drone the signal is then given to the ZigBee for transmitting the information to the base station.

The ZigBee receiving module then receives the data and demodulates and sends the signal to a computer for further process and to display the output in an understandable interface. This module as the other module has a dedicated power supply with a capacitive relay so that the module does not effected by any other power flow or external frequency. This module is designed to a miniature size so that it can be installed very easily and had a minimum weight so that it does not affect the payload of the drone.

**ECG Sensor** (Electrocardiogram) sensor records the pathway of electrical impulses through the heart muscle and can be recorded on resting and ambulatory subjects, or during exercise to provide information on the heart's response to physical exertion. AD8232 is a mini-sized module to monitor the heart condition when the victim is in tough condition in response to a physical move. Usually the device has three inputs, the inputs are connected to the electrode pads have a defined position to place and

obtain an accurate reading the input pads are placed in the configuration of (RA, LA, RL). Which is right arms, left arms, right leg, this module monitors the heart condition by sending signals through electrodes and receiving it through other electrode. This device is very small in size that it can be placed in drone and can be operated by any person without any experience. This device inputs are connected with Arduino to modulate the signal without any noise and then the signal is transmitted to the base station using ZigBee the output is then monitored using ECG viewer scale this will give a brief information of the patient to the medical expert. This module has dedicated power supply so that it does not mix with any other signal to and result in noise.

**Speaker** An audio system is placed in the drone through which the base team can communicate with the person present in the site; the speaker is installed with LM1875 which has a crystal clear audio output and decent amount of amplified audio. The audio system is enabled with microphone module used to create a two-way communication system. The audio signal is connected to ZigBee for transmitting the signal and receiving the audio signal as the other system this system retains power from the drone power distribution system.

### *3.2 Software Required*

**Arduino IDE:** This software is used for interfacing all the various components with the microcontroller, i.e., the Arduino Uno board. This enables an easy link between the data collected and the data received. For various functions to be performed, different commands are given to the respective functions to be carried out.

**Google Earth:** As we have already introduced telemetry function, Google earth is needed by the drone for navigation purpose, so that it can reach the assigned destination within minutes.

**Lab VIEW:** This software is basically used for the components such as the ECG sensor, where after getting the parameter reading we get a waveform representation of it.

## **4 Results and Discussion**

The yellow block beneath the drone carries all the components, including various sensors and the defibrillator summing up the total weight. The drone was able to carry all the weight of the external sensor and the first aid kit box. The recorded flight time of 20 min in a decent altitude which could be extended by supplying external battery was achieved. All the sensor had a good response time and had a good reading output. Control of the base station was very easy and effective with





**Fig. 6** Socio-medic drone lifting off for emergency

**Table 1** Tabulation of the flight time and distance with respect to weight

S. No.	Weight (g)	Flight time (min)	Distance (km)
1.	500	31	26
2.	750	20	17
3.	1000	15	13
4.	1500	12	10
5.	1750	10	8

no signal loss. The camera had a good response and perfect picture quality video. ZigBee module was perfectly installed in the drone (Fig. 6).

Motors installed in the drone are 940 kV with a battery (power supply) of 8000 mAh. Table 1 represents the trials done with the drone weight and noted flight time with maximum distance covered with the drone, the average speed of the drone is fixed to 50 km/h.

## 5 Conclusion

The technological advancements in this world have provided us with many positive aspects, but it is also the base for few of the negative sides happening in the current time. Traffic congestion being one of them, which also leads to the delay of ambulances during emergency situations. This paper deals with the solution for such an issue and gives a clear vision of how this can be taken care by the usage of a medic drone which will be fast as well as efficient. It also gives an insight into the idea of using a defibrillator that can be a life-saving component for the victim of an accident or for a person who suffers from a cardiac arrest.

## References

1. Pahonie R, Mihai R, Barbu C (2015) Biomechanics of flexible wing drones usable for emergency medical transport operations. In: International conference on e-health and bioengineering conference (EHB), Iasi, pp 1–4. <https://doi.org/10.1109/ehb.2015.7391538>
2. Bravo GC, Parra DM, Mendes L, de Jesus Pereira AM (2016) First aid drone for outdoor sports activities. In: 2016 1st international conference on technology and innovation in sports, health and wellbeing (TISHW), Vila Real, pp 1–5. <https://doi.org/10.1109/tishw.2016.7847781>
3. Scott CH (2017) Drone delivery models for healthcare. In: Proceedings of the 50th Hawaii international conference on system sciences, University of Colorado Denver, USA
4. Fancher JC, Huang ZM, Kui CX, Zhi LY (2017) Unmanned drones for medical supply delivery in china major qualifying project, WPI. A major qualifying project submitted to the Faculty of Worcester Polytechnic Institute in partial fulfillment of the requirements for the degree of Bachelor of Science, 20 Apr 2017
5. Claesson A et al (2016) Unmanned aerial vehicles (drones) in out-of-hospital-cardiac-arrest. *Scand J Trauma Resuscitation Emerg Med*
6. Sanfridsson J et al (2019) Drone delivery of an automated external defibrillator—a mixed method simulation study of bystander experience. *Scand J Trauma Resuscitation Emerg Med* 27:40. <https://doi.org/10.1186/s13049-019-0622-6>
7. Daly K (2010) Article 08 legal issues affecting supply chain. Article 08—the future is here. How drones are modernizing the healthcare industry
8. Rosser JC Jr, Vignesh V, Terwilliger BA, Parker BC (2018) Surgical and medical applications of drones. *JSLs* 22(3):e2018.00018. [www.SLS.org](http://www.SLS.org)
9. Krishna VV (2018) Design of drone ambulance. *Int J Pure Appl Math* 119(15):1813–1818. ISSN: 1314-3395 (online version), special issue
10. <https://en.wikipedia.org/wiki/Defibrillation>

# Big Data Processing Based on Machine Learning for Multi-user Environments



Kamel H. Rahouma and Farag M. Afify

**Abstract** Many sources of data yield non-structured data like the Internet of things (IoT), geospatial data, E-commerce, social media, and scientific research that is not appropriate in to traditional, structured warehouses. Nowadays, sophisticated analytical techniques allow companies to obtain perspicacity from data with earlier unachievable levels of accuracy and speed. Real-time analytics for big data is the capability to achieve the most suitable decisions and get significant actions at the best time. First, we present a survey of processing the big data (BD) in real time (RT) and focus on its challenges. Then, we propose an algorithm to handle BD by integration with machine learning operations in multi-user environment optimization operations, reduce maintenance costs and better speed of fault detector and provide common operations necessary to process unstructured information. There are important conditions that have been taken into a concern to guarantee the quality of services (QoS) and transmission velocity and ensure the system's physical time synchronization and the correctness of the data processing.

**Keywords** Big data · Multi-user environment · Real-time processing · Machine learning

## 1 Introduction

The developments in the application of scientific knowledge are for practical purposes, especially in an industry in the last few years driven to the accumulation of huge quantities of information. BD brings recognition from business, governments' operations, research, etc., because BD tools let users treat big amounts of information to obtain the best judgment through the analytics of information. Machine learning (ML) is a vital element in the analytic of information as it prepares the machines to obtain knowledge by using trained information, then it makes efficient decisions by using experiments that decreases the time of data processing.

---

K. H. Rahouma (✉) · F. M. Afify  
Electrical Engineering Department, Faculty of Engineering, Minia University, Minia, Egypt  
e-mail: [kamel\\_rahouma@yahoo.com](mailto:kamel_rahouma@yahoo.com)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_68](https://doi.org/10.1007/978-981-15-2329-8_68)

To obtain certain decisions, ML tools were employed, and it can increase the performance of other components of knowledge innovation.

BD [1] has grown within the presence for a few years. BD is related to traditional database systems, though the distinction is that it exceeds in volume and processing.

BD is very huge, transfers so quick including unorganized and organized information, while traditional data contain the only structured information [2].

The digitization of all operations now produces different kinds of huge and RT data across a wide range of processes. Several applications produce BD, for example, social media, scientific experiments, cloud applications, E-government services, monitoring programs, and information warehouses. Data expand quickly since operations generate continuously rising amounts of unorganized and structured information [3]. The impact on information storage, processing, and transfer is necessary to re-evaluate the methods and clarifications to great result the user wants. In that context, processing types and algorithms have a great role. A great quality of answers for special applications and programs exist, so an accurate and systematic analysis of present solutions for processing standards, algorithms, and techniques applied in BD storage and processing environments has great importance. Nowadays, sophisticated analytical techniques allow companies to obtain perspicacity from data with earlier unachievable levels of accuracy and speed. Real-time analytics for BD is the capability to achieve the most beneficial judgments and get significant actions at the best time [4].

In this paper, we present an overview of processing the BD in RT and focus on its challenges. We suggest an algorithm to handle BD in a multi-user framework, and we provide common operations necessary to process unstructured data. The paper also provides an algorithm to schedule multiple functions of users simultaneously in a distributed environment. The rest of the paper introduces a background of the big data in Sect. 2. In Sect. 3, we give a literature review, and in Sect. 4 we explain the methodology of this paper. In Sect. 5, we depict the outcome of our work and compare our results and the previous work results. In Sect. 6, we introduce some of the conclusions, and a list of the used references is given at the end of the paper.

## 2 A Background About BD

### 2.1 What Is BD?

Among all the descriptions given for “BD,” commonly it means information that is very huge, very fast, or very difficult for existing tools to process. “Very huge” shows that companies frequently and necessarily agree to deal with petabyte-scale quantities of information that become of sensors, transaction histories, clicks streams, and outside. “Very fast” shows that not only is information huge, but also it is to be processed immediately. For example, it is to make cheating detection at a location of sale or to determine which ad to display to a person on a webpage. “Very difficult”

is a collection of information that does not match cleverly into a current processing tool or that needs some kind of analysis that current tools cannot offer [5].

## 2.2 Characteristics of BD

There are three Vs which may shortly describe the characteristics of the BD. These are the variety, velocity, and volume. The volume means the number of datasets and storage. The variety means the information types. The velocity means the rapidity of incoming information. Growth of research and discussions concern with BD, and the three Vs were expanded to five Vs as shown in Fig. 1 These are the veracity, variety, volume, velocity, and value [6].

Today’s BD plays an important function for a lot of fields like a business, online purchasing, banking, astronomy, health care, and finance. BD gives great advantages to business activities. BD is produced by a huge quantity of data. These data extend to rise every day because information arrives permanently of various operations.

BD has a huge quantity of unwanted data in both unorganized and organized structures. In unorganized, the information is stored in undefined and unsystematic ways, while organized information is saved into well-defined structures [7].

Facebook, Wikipedia, and Google produced unorganized information, while E-commerce operations produced organized information. Since the appearance of organized and unorganized information, some difficulties appear in BD such as data collecting, data transfer, sharing, storage, privacy, analysis, search, handling of information, fault tolerance, and visualization. It is impossible to handle these difficulties in regular ways. Regular information administration mechanisms were unable to process, analyze, and schedule jobs in BD. Therefore, BD uses different tools to handle these challenges. This paper aims to study these challenges [8].

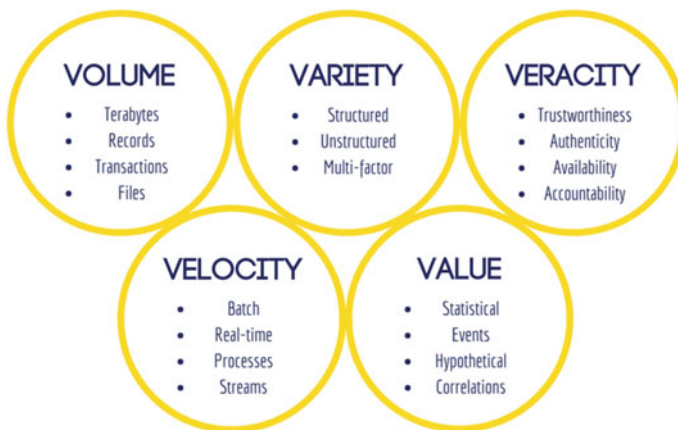


Fig. 1 Five Vs of big data

## 2.3 *BD Software*

### 2.3.1 **Hadoop (HD)**

The appropriate software for managing BD difficulties is HD. HD is software that allows distributed handling of huge information across clusters of machines applying simple principles of programming [9]. It contains two parts. The first part is used for storage, and it is named Hadoop Distributed File System (HDFS). The second part is employed for processing, and it is named MapReduce (MR). The HDFS has a master structure and slave structure. The master is a process named (NameNode) which controls the operations on files and manages the global namespace. The slave is a process named (DataNode), and it saves the data in the structure of data blocks and operations as instructed by the NameNode. The NameNode handles the information replication and arrangement for reliability, fault tolerance, and performance. The NameNode divides and saves files into 64 MB data blocks over the DataNodes. Typically, there are three duplicates of all data blocks which are saved in the HDFS. The failure detection is implemented in the framework of regular duplicates of DataNodes to NameNode. If there is no heartbeat from a DataNode for a long time, it is marked as lost not employed for new processes and if required, additional duplicates of its data are implemented [10].

### 2.3.2 **MapReduce (MR)**

MR is an information handling standard for dealing with various challenges of computing. The MR concept is stimulated through the map, and it decreases functions, which are usually used in functional languages. The MR allows administrators to simply display their process like a map and decrease operations [11].

## 2.4 *BD Processing Framework Challenges*

Increasing the number of researches in the BD field does not express that we understand all BD processes, so we do not have a common definition of BD and we have a lot of differences about tools and applications [12]. Moreover, there is a big problem that researchers are currently interested, this problem is process the information in RT and the process by which the data we need to make a quick decision is processed. like monitoring and protecting systems, Electrical grid and Smart Cities [13–15]:

1. Manage energy of all sectors.
2. Solve any faults problems.
3. Save energy as much as possible.
4. Distribute power reasonably.

To execute these Jobs, fault analysis and error status must perform during a low period; oppositely that make more failures in the electric network. RTBD has important requirements in analysis, acquisition, security, data management, and benchmarking. These challenges are briefly described in the following [16–18]:

1. The RT processing speed.
2. The RT systems stability.
3. The large-scale applications.
4. Data collection.
5. Data analytics.
6. Data security.
7. Usability issue of data management.
8. Test benchmark of performance.

## 2.5 Technical Challenges

Added to the last challenges, there are some technical challenges which may be faced by BD systems. From these challenges [19, 20]:

1. Fault tolerance

New incoming technologies like cloud computing and BD it is regularly that whenever the failure happens the damage is done should be within acceptable threshold rather than beginning the whole job from scratch. Fault-tolerant computing is greatly complicated, containing complex algorithms. Hence, the main responsibility is to reduce the possibility of error to the minimal level.

2. Quality of data

Storage and collection of a huge amount of information are more costly. More information is applied for predictive analysis and decision making in business will give to best results. BD focuses on the quality of information storage rather than having very great irrelevant information so that excellent result and conclusion can be formed.

3. Heterogeneous data

In BD, unstructured information describes almost every sort of data being generated by fax transfer, social media interaction, and handles various kinds of document and more. Transforming unorganized information into organized information one is also not possible. Structured information is organized but unstructured information is raw and unorganized. In addition, two important challenges are visualization and hidden BD.

### 3 A Literature Review

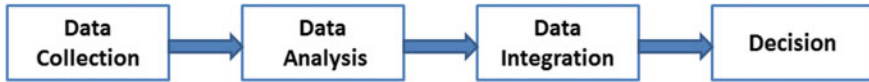
Data is one of the most important things in all parts of our lives, and every day large quantities are produced. Many international reports indicate that data quantities are multiplying daily and very quickly. Despite the huge flood of data, they are only used to a limited extent where they are managed, processed, and analyzed for their use [21]. These huge quantities need to be dealt quickly, and some of them require real-time processing, for example, detect fraud and data for security institutions in general, and there are companies benefited from these operations, especially fraud operations such as PayPal. As a result of the great need and difficulties to deal with these huge amounts of data appeared, many tools to deal with HD and MR; the major companies are working to take advantage of these quantities like Google and Yahoo. In the past few years, the research has focused on big data operations, and a large number of tools have been proposed, for example, Storm [22] and Spark [23]; these tools are able to face many operations challenges dealt with on BD. These tools help companies in data processing and obtaining the desired results, and it has been used in many large companies such as Netflix which give them the best recommendations to deal with customers and address errors.

Feldman et al suggested suitable methods for handling big data and provided a suitable solution for the large storage areas that are used daily to store this data. Data compression has become an important element that has proven to be effective when used by Feldman and all data still without a loss. Faced with the difficulties of handling BD was behind the suggestion that was introduced by Jiang et al. A tool capable of handling and retrieving big data, increasing processing speed, analyzing, and assisting in decision making was introduced. Cuzzocrea et al proposed a tool capable of handling graphs and data for BigWeb Data. This tool has proven to be effective and has provided a solution to most of the problems facing companies collecting data from the web [24–26].

### 4 The Methodology

Our methodology depends on using ML; it is a combination of a set of sciences which able us to get benefit on existing data and learn and act on previous experiences. At the moment, ML is used in all aspects of life such as health, industry, and trade. ML technologies enable the handling of layers, devices, and data that were impossible to access, make decisions, and avoid previous errors. Data processing is usually required before it can be used to remove unnecessary data and incomplete data through which systems can learn to avoid past errors, increase productivity, and develop products [27, 28]. ML tools have been able to gain experience by learning and benefit from the experiences built by the systems and the way we can take advantage of the data. If we say that the big data helps us to store large amounts of data, through ML and





**Fig. 2** Real-time data processing stages

integration with BD as shown in Fig. 4, we can get useful information that is easy to use in all areas [29].

#### ***4.1 The Proposed Algorithm***

We proposed an algorithm with a high-level programming interface which achieves integration between BD and ML and provides common operations necessary to process unstructured data. It also supports an algorithm to schedule multiple functions of users simultaneously in a distributed environment. It has also been taken into account robust computing capability; a large RT information processing necessitates a powerful timing, and this means the system must respond to any request in the shortest possible time. So in the beginning, the RT data processing system must have powerful data computing capability. The traditional way of processing large data is depending on the robust computing abilities of a cloud to obtain the desired goal, while at the RT, it necessitates depending on the strength to rapidly exchange information between devices.

The RT data processing framework is split into four stages: collection, analysis, integration of data, and resolution as shown in Fig. 2.

Data collection is responsible for information collecting plus storage and includes information cleaning and information preparation for analysis.

Data analysis is the core of the large real-time data processing system and the critical stage to determine system performance. This phase is mainly responsible for modeling data structures, clearing data, processing, and making information ready for integration layer.

Data integration: This phase plays a related role in these processes. At this stage, the combination is between more information processing algorithms and providing support for other layers.

Decision making: In this layer, decisions are made based on the data coming from other layers, from which the final objective of the information analysis process is produced.

This algorithm will help to find the similarities between the documents to make informed decisions about clustering. Many distance metrics can be used to find similarities between documents.

The framework described generically in nature works with any set of documents that make up large data. The frame takes large data as inputs and produces clusters. The processes involved in the framework include keyword selection, creation of feature space, calculation of similarity, and aggregation.

The algorithm is responsible for terminating specific tasks to improve the productivity of large data processing. The distributed environment takes several useful functions and arranges them correctly in a way that is handled optimally. The concept of waiting time is used to ensure that jobs are given their role, and large data processing is performed efficiently.

#### 4.1.1 Algorithm

```

Initialize files = Get list of files in storage(Partition)
Initialize filesData = Get last running data result(DB)
Initialize newFileData<File, Hashcode>.
foreach file in files
    newHashCode = generateFileHashCode(file)
    oldHashCode = getSavedHashCode(file, filesData)
    AddRecode(newFileData, file, newHashCode)
    if oldHashCode = nothing
        Report("New Added File");
    elseif oldHashCode = newHashCode
        Report("no Change on File")
    Else
        Report ("File Updated")
    EndIf
next file
forEach file in filesData
    If files not contains file
        Report("File Deleted")
    End if
Next file
newFileData Overwrite fileData
    Save(fileData)

```

#### 4.1.2 The Scheduling

We suggest Responsive Job Scheduler that depends on the position of reference. The suggested tool intended for a group contains machines or nodes (N) from 1 to  $n$ , and these machines contain what it needs to work efficiently. The group contains a multi slave node (SN) and one master node (MN). MN contains operations list and handling these operations according to SN specification.

The workflow of operations list is as follows:

1. The new job comes to operations list (OL).
2. Task is split into two tasks (Map and Reduce) and map split to local non-local job.
3. When a new slave task is added, then it checks for free slot count.
4. If any job arrives at SN, it sends a signal to MN which manages operations list for SN.
5. MN inspects the arriving job to determine the appropriate time for implementation and waiting for finishing tasks or arriving jobs.

$$\text{Data locality} = \text{no. of local map/task} / \text{Total map task}$$

The fairness of a job

$$\text{fairshare} = \text{job weight} / \sum \text{job weight} * \text{capacity of taktracker}$$

Average acknowledgment time

$$T_{\text{avg}} = Rl * T_{\text{avg}}^1 + (1 - Rl) T_{\text{avg}}^{\text{nth}}$$

Performance =

Map task = no. of pending map tasks / no. of currently running map task

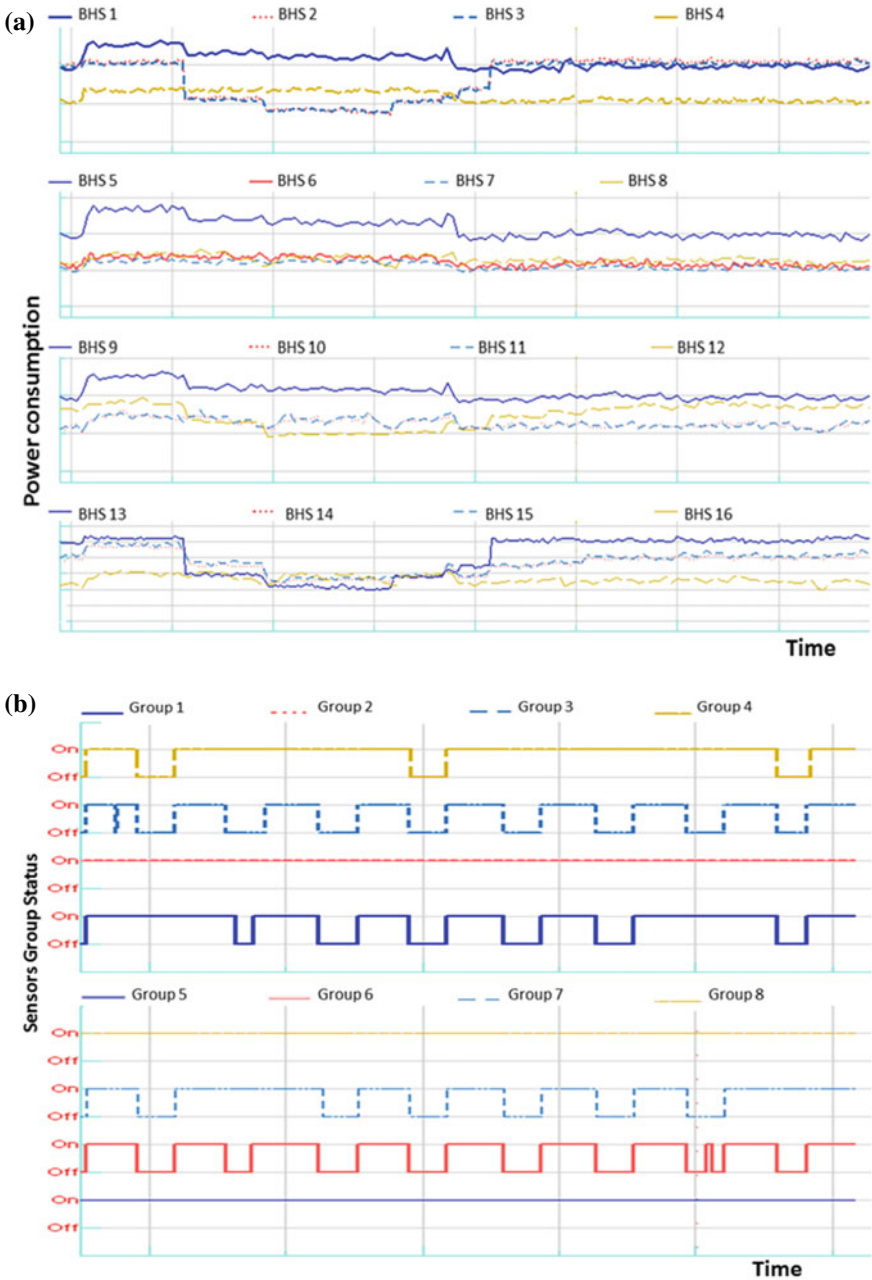
Reduce task = no. of pending reduce tasks / no. of currently running reduce task

The new framework is a combination of algorithms to execute the RTBD analytics and to improve the precision of information which the supervisor needs it. The proposed framework, based on these algorithms, enables the supervisor to work on BD and face constraints and discover useful data and decisions.

## 5 The Results

RT forecasting of complex systems is important because it is easy to plan maintenance, minimize problems caused by sudden downtime, and reduce the value of spare parts consumed in maintenance. This proposal uses BD integration and machine learning processes.

In this case, the focus was on the data which come from the monitoring of the status and power consumption of motors and sensors which is a vital part of the operation in baggage handling systems and describes the experiments carried out to estimate the efficiency of the suggested process. Figure 3 presents the groups' status and power consumption motors and sensors.



**Fig. 3** a Phase of RT information processing and detect faults in BHS. b Phase of RT information processing and detect faults in sensors group

The tool was able to perform real-time processes on gram amounts of data, exploiting the benefits of working on unstructured data in fault detection and decision making to decrease solving problem time.

The experimental results showed a rapid increase in the amount of information that has been processed, in addition to the increase in the amount of data that did not affect the implementation times.

Using this tool provides also a fault ranking estimation, and this is particularly useful for this application on BHS where, almost always, an unbalance fault shows up as well, also if another error is the root cause of the anomaly. Examining the assignment features is then a true benefit of the tool. The balance between the amount of information being worked on and the number of devices must be considered to achieve the best performance. The present technique is found to be more efficient than the existing ones in the literature. Some of the reasons for that are as follows:

1. Most existing techniques have the inherent problems of analyzing unstructured data in real time, and the operations of these techniques depend on collected data for systems especially industrial systems. In our proposal, the algorithm capable of executing operations on huge quantities of unorganized information in real time and increasing the efficiency of fault detection reduces the extra cost if the system has failed down during executing important jobs and improvements of the efficiency of BD operations.
2. The tool reduces maintenance time according to the status, saving of running tasks time when detecting the failure and faults prediction if any data come from sensor contrary to the normal situation.
3. The tool enables data-based decision making because it creates a value of information previously neglected by handling the problems quickly and gives a better understanding and sufficient detail to these troubles.

## 6 Conclusions

This paper provided an overview of processing BD in RT and focused on its challenges. An algorithm is suggested to handle big data by integration with machine learning operations in a multi-user environment optimization. This reduced the maintenance costs and resulted in a better speed of fault detector and provided common operations necessary to process unstructured data. The network transmission speed and the quality of service (QoS) factors have been taken into account. Thus, we make sure that the system works efficiently and make sure that the data processing process is completed efficiently.

## References

1. Kshetri N (2014) Big data s impact on privacy, security and consumer welfare. *Telecommun Policy* 38(11):1134–1145
2. Kitchin R (2014) The real-time city? Big data and smart urbanism. *Geo J* 1–14
3. Chang RM, Kauffman RJ, Kwon YO (2014) Understanding the paradigm shift to computational social science in the presence of big data. *Decis Support Syst* 63:67–80
4. Chen J, Chen Y, Du X et al (2013) Big data challenge: a data management perspective. *Front Comput Sci* 7(2):157–164
5. Zhao JM, Wang WS, Liu X et al (2014) Big data benchmark big DS. *Advancing Big Data Benchmarks*. Springer International Publishing, pp 49–57
6. Wang CC, Chen CL, Hou ZY et al (2015) A 60 V tolerance transceiver with ESD protection for FlexRay-based communication systems. *IEEE Trans Circ Syst I: Regul Pap* 62(3):752–760
7. Hwang K, Chen M (2017) Big data analytics for cloud/IoT and cognitive learning. Wiley, UK
8. Hwang K, Chen M, Wu J (2016) Mobile big data management and innovative applications (editorial). *IEEE Trans Serv Comput* 9(5):784–785
9. Bende S, Shedge R (2016) Dealing with small files problem in hadoop distributed file system. *Procedia Comput Sci* 79:1001–1012
10. Hadoop Distributed File System [online]. <http://hadoop.apache.org/hdfs>
11. Cheng D, Zhou X, Lama P, Wu J, Jiang C (2017) Cross-platform resource scheduling for spark and MapReduce on YARN. *IEEE Trans Comput* 66:1341
12. Wang B, Jiang J, Wu Y, Yang G, Li K (2016) Accelerating MapReduce on commodity clusters: an SSD-empowered approach. In: *IEEE transactions on big data*, IEEE, 2016
13. Y. Liu, M. Qiu, C. Liu, et al., Big data challenges in ocean observation: a survey, *Personal Ubiquitous Comput*. 2017
14. Yildiz O, Ibrahim S, Antoniu G (2017) Enabling fast failure recovery in shared Hadoopclusters: towards failure-aware scheduling. *Future Gener Comput Syst* 74:208–219
15. Mavridis I, Karatza H (2017) Performance evaluation of cloud-based log file analysis with Apache Hadoop and Apache Spark. *J Syst Softw* 125:133–151
16. Xu H, Lau WC (2017) Optimization for speculative execution in big data processing clusters. *IEEE Trans Parallel Distrib Syst* 28(2):530–545
17. Gani A, Siddiqa A, Shamshirband S, Hanum F (2016) A survey on indexing techniques for big data: taxonomy and performance evaluation. *Knowl Inf Syst* 46(2):241–284
18. Gil D, Song I-Y (2016) Modeling and management of big data: challenges and opportunities. *Future Gener Comput Syst* 63:96–99
19. Xu H, Lau WC (2017) Optimization for speculative execution in big data processing clusters. *IEEE Trans Parallel Distrib Syst* 28(2):530–545
20. Sivarajah Uthayasankar (2017) Muhammad Mustafa Kamal, Zahir Irani, Vishanth Weerakkody, Critical analysis of Big Data challenges and analytical methods. *J Bus Res* 70:263–286
21. Zhou L, Pan S, Wang J, Vasilakos AV (2017) Machine learning on big data: opportunities and challenges. *Neurocomputing* 237(10 May):350–361
22. <https://storm.apache.org/>. Accessed July 2019
23. <https://spark.apache.org/>. Accessed July 2019
24. Yokota R, Wu W (eds) (2018) Supercomputing frontiers. In: 4th Asian conference, SCFA 2018 Singapore, 26–29 March 2018
25. Feldman D, Schmidt M, Sohler C (2013) Turning big data into tiny data: constant-size coresets for k-means, PCA and projective clustering. In: *SODA*, 2013
26. Jiang F, Leung CK (2015) A data analytic algorithm for managing, querying, and processing uncertain big data in cloud environments. *Algorithms* 8:1175–1194

27. Cuzzocrea A, Cosulschi M, De Virgilio R (2016) An effective and efficient MapReduce algorithm for computing BFS-based traversals of large-scale RDF graphs. *Algorithms*
28. Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. *Science* 349:255–260
29. Zhou L, Pan S, Wang J, Vasilakos AV (2017) Machine learning on big data: opportunities and challenges. *Neurocomputing* 237(10):350–361

# Integrating Chatbot Application with Qlik Sense Business Intelligence (BI) Tool Using Natural Language Processing (NLP)



Vipul Vashisht and Pankaj Dharia

**Abstract** So far, various approaches have been proposed for implementing chatbots with business intelligence tools, yet most of these approaches have limited adoption in practice. The objective of this paper is to offer an easy to understand approach by the business users that could be quickly adopted for integrating the chatbot with a business intelligence tool. This communication describes a process of applying artificial intelligence technologies; in particular, natural language processing (NLP) for conversation using chatbot. The chatbot implementation and integration with the BI tool has given encouraging results.

**Keywords** Chatbot · NLP · BI · Qlik Sense

## 1 Introduction to Chatbot

Most businesses nowadays are focusing on implementing cost optimization initiatives and quick to market strategies to overcome the ever-increasing competition. This has resulted in widespread consideration of smart and innovative automation options for improving business efficiency. Integration business analytics for gaining quicker insight and single source of truth is gaining popularity among the organizations. In the recent years, there has been multiple innovations that have taken place in field of speech-powered applications. Today many travel portals, hotels, banks, insurance companies have been using NLP-based applications to ensure that user queries are answered at any time, irrespective of availability of human agent. We have seen applications like Microsoft Cortana, which can provide human like responses to the queries.

Our study is focused on integrating the chatbot with a business intelligence tool. This would enable the users to get the intelligent response that include both facts and charts, from data spread across multiple sources in an organization. The chatbot

---

V. Vashisht (✉) · P. Dharia

IRT Digital Analytics Solutions Private Limited, Noida, India  
e-mail: [vipulvashisht@gmail.com](mailto:vipulvashisht@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_69](https://doi.org/10.1007/978-981-15-2329-8_69)



essentially takes input from a user, processes it, and responds to the user. The definition includes intelligence because there are a whole set of artificial intelligence (AI) services available through many other third parties. In most cases, chatbots applications are leveraging some form of AI. People communicate with chatbots via text and possibly even voice. Conversation can be text and/or voice. In most of the cases, the design of a chatbot includes setting up of rules using NLP and allowing interaction with the business users using an easy to use text-based interface.

### ***1.1 Benefits of Chatbot***

- **Conversation:** The primary interface of a chatbot is a conversation window with the end user. The application engages with the end user via CUI.
- **Easy to Use:** The execution just requires to post a query to the chatbot and based on its intelligence, it would provide a quick response and then conversation can start.
- **Availability:** Chatbots applications are available on multiple devices like desktop, laptop, mobile phones, iPads.

### ***1.2 Business Intelligence (BI)***

It was during 1989, when Howard Dresner conceptualized the idea of BI. It was defined as “concepts and methods to improve business decision making by using fact-based support systems.”

With the advent of continuous increase in the size of organizational data, multiple data sources and complexity, the role of BI has been gaining continuous attention. There have been surges in development of BI tools and their marketing among business organizations. Some of the popular BI tools available in market are Qlik Sense, QlikView, Microsoft PowerBI, Tableau, and ThoughtSpot. BI tools are known to provide single source of truth across the complete data landscape in the organization. Earlier, the usage of these tools was limited to the management layer in organizations for quick decision making. With time and availability of BO tools on mobile, the usage of the tools has been extended to the wider section of business users including managers and executives. Some organizations, specifically in manufacturing sector, the BI tools have been extended to the distributors and third party members [1, 2].

Benefits of BI tools include:

- Helping C-level executives, controllers, and operational managers improve equipment efficiency throughout the organization.
- Enabling customer to identify trends sooner and go much deeper in its analysis to generate useful insights.

- Saving crucial time and efforts of business users for defect free deliveries.
- Eliminating data manipulation need by business users.
- Increasing adoption of reports usage in organization.
- Enabling slice/dice approach for detailed analysis.
- Improving response time resulting in accelerated time to market.

In this paper, we have used Qlik Sense BI Tool ([www.qlik.com](http://www.qlik.com)) for showcasing integration with chatbot. Next, we provide an overview of the NLP.

### ***1.3 Natural Language Processing (NLP)***

As per research report conducted by Gartner, there would be a considerable rise in use of speech or NLP for fetching in a BI tool. To see the benefit of the conversational interface, this paper discusses a branch of artificial intelligence known as natural language processing. The existence of NLP does not mean that a chatbot will be able to understand anything a user says. Rather, the developer needs to train the NLP model to understand very specific tasks belonging to a chatbot's domain. In the world of chatbots, there are plenty of opportunities for buttons, cards, and quick commands, but the essential aspect of chatbots is that they have conversational interfaces. With our proposed approach, it would be possible to target multiple channels, where text conversation is the default standard. It has been observed that most of the applications can naturally fit into this conversational world and often yield a superior solution [3, 4].

Next, we provide insights from the literature review of past research work done in the subject area.

## **2 Literature Review**

Chatbot has remained an active and popular field among researchers. There have been many papers written on chatbot technology but there has been little work done in field of integration of BI with chatbot. Some of the relevant work studied during the literature review is provided below:

- Cho and Lee [5] designed an application prototype for use in construction industry. The application provided a report which was based on the conversation among the construction staff team members.
- Shawar and Atwell [6] trained the chatbot using the NLP technology, to new languages.
- Deshpande et al. [7] provided a detail history of chatbots from initial days to current intelligent response providing system.
- Thomas [8] proposed an AI-based approach, where a dataset based on FAQs was used to train the chatbot for providing immediate responses.

- Mondal et al. [9] designed a chatbot in a form of telegram Bot that assists in answering questions.
- Rosruen and Samanchuen [10] proposed a chatbot application for use of services in system provided for medical consultants.

### 3 Chatbot Design Approach

The objective of research paper is to develop a chatbot that can be integrated with a business intelligence solution, in particular Qlik Sense with respect to this paper.

Following components are used in designing a chatbot framework:

1. **Bots** are telegram accounts that can send replies for received messages. These can be easily integrated to work with other programs [11].
2. **Dialogflow** is development suite from Google, which is used for building conversational interfaces for Web sites, mobile-based applications, messaging platforms, and IoT [12].
3. **Client application as telegram:** Telegram is a cloud-based service that has been used to connect and communicate with the chatbot. Telegram provides Bot father service, which is used to create and register BOTs in telegram [13, 14].
4. **.NET Bot service:** .NET Bot service is the centralized service created using console application in .NET. It is responsible for communicating between Qlik Sense, telegram service, and NLP. First, it takes a message from telegram app, sends it to NLP service, gets the required details, and forms the query to retrieve data from Qlik Sense. Once the result is received from Qlik Sense, it will generate the result for telegram.
5. **Qlik Sense as a business intelligence tool:** Qlik Sense is a data visualization and data discovery tool, which helps the user to gain insight from data in a very intuitive way. Qlik Sense works on the associative engine, which creates a data model in which any selection made on data point will slice and dice the entire data set.
6. **NLP:** This is the heart of every chatbot system that help the machine to make sense from a user query. NLP identifies the intent and entity from a user query. Intents are verbs in sentence and entity is parameter from a sentence. For NLP, we are using Google' Dialogflow. Dialogflow is a tool developed by Google used for interpreting the human language based on natural language processing technology [8].
7. **Narratives:** Narratives are used to describe the charts in Qlik Sense as shown in (Fig. 1).

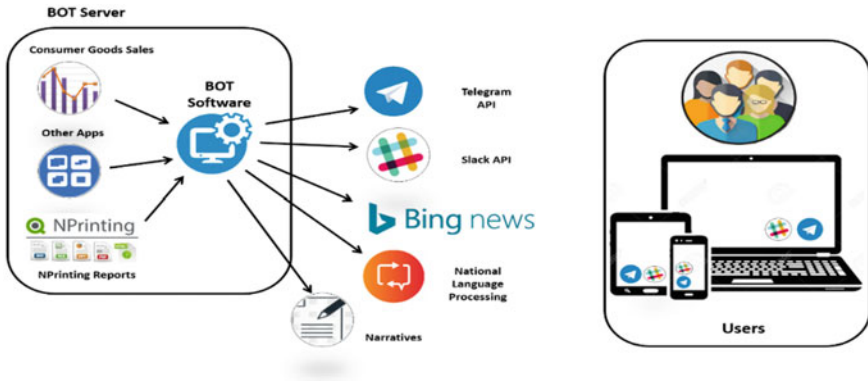


Fig. 1 High-level architecture of chatbot

### 3.1 Implementation Steps

Following are the steps to implement BI Chatbot:

**Configure the Qlik Sense server for Chatbot:** Qlik Sense server can be accessed using virtual proxies and header authentication. Virtual proxies help to connect to Qlik Sense engine and header authentication is a way to implement security to authenticate a user for accessing the virtual proxies. Following are details need to be filled while creating virtual proxies in Qlik Sense [15].

```

Description: Telegram-Bot
Prefix: Telegram-app
Timeout: 50
Session cookie header name: X-Qlik-Session-telegram-bot
Anonymous access mode: No anonymous user
Authentication method: Header authentication static
user directory
Header authentication header name: XXXXXXXXXX
Header authentication static user directory: Sample

```

Header authentication header name is used to make an authentic connection to Qlik Sense. It is recommended to check virtual proxies whether the connection is working and running successfully.

**Configure and register Chatbot in Telegram:** Telegram provides Bot father service to create and manage chatbot in telegram [15]. Once chatbot is configured in Bot father, a service node is created with a token, which will be used to retrieve and send messages to registered telegram Bot.

**Create an NLP model in Dialogflow:** Natural language processing is done using a Dialogflow model. In Dialogflow model, an agent is created, which is trained with



Fig. 2 Example of adding training utterance in Dialogflow model

some scenarios or utterances. As an example, as shown in Fig. 3, as a part of model training, an utterance “Show category” is trained marking the word “category” as dimension entity. Using the same approach, model is trained in Dialogflow with other scenarios (Fig. 2).

### 3.2 Creating a .Net-Based Bot Service

- Next step is to create a .NET console application in visual studio in C#. Initially, a method is created to initialize the telegram Bot and make a connection to the same. As shown in Fig. 3, the Bot information and token is used to make a connection to telegram chatbot using C#.
- OnReceivedMessage method will be called when a message is received by telegram Bot. The information acquired from the OnReceivedMessage will be sent to ProcessMsg method.
- Now ProcessMsg will connect to the Dialogflow API with message received and based on the created model, Dialogflow will return the relevant intents and entities in JSON format.

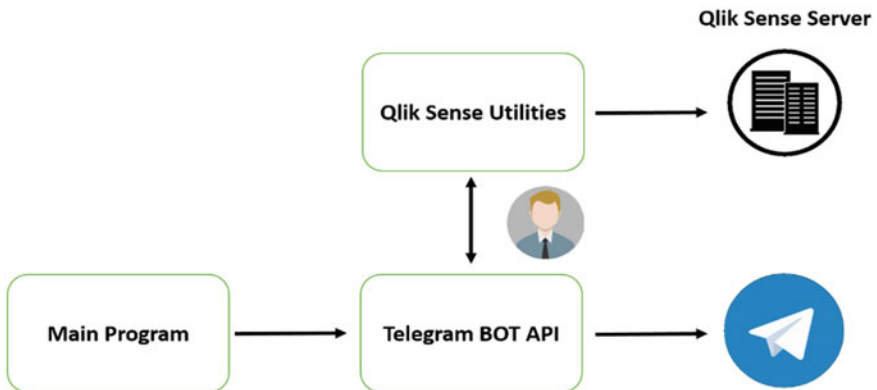


Fig. 3 Service connection

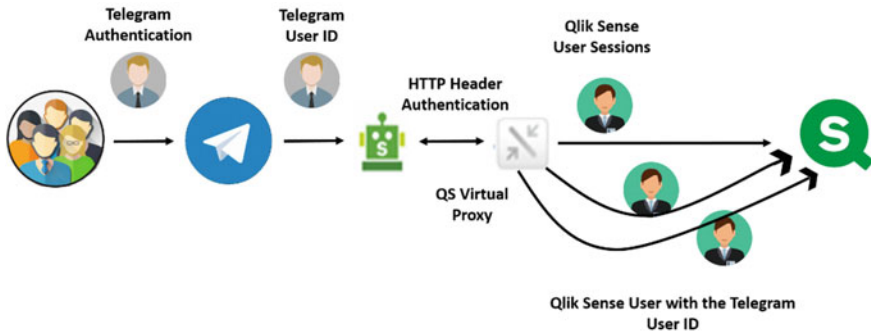


Fig. 4 CheckTheUser () flow for user access authentication

- Based on intent and entities, Qlik Sense query is created to retrieve the relevant data, which will be sent back to the telegram. As shown in Fig. 4, in order to check the authenticity of the user, CheckTheUser () method has been used [11].

## 4 Results and Discussion

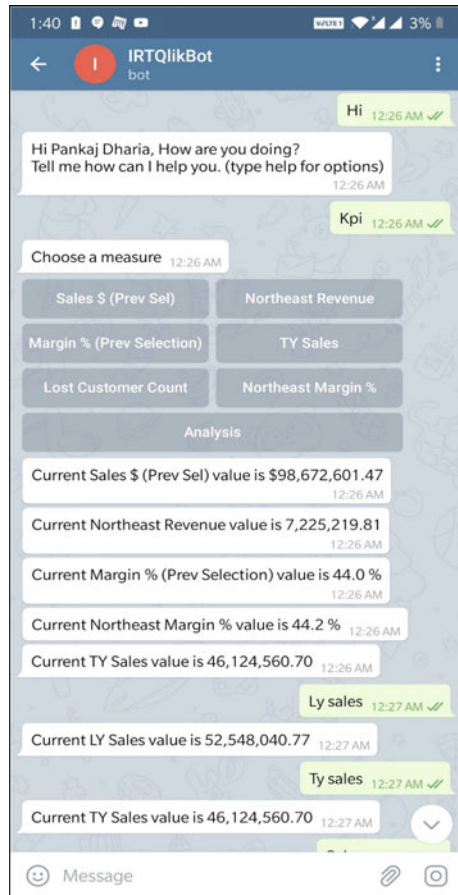
The chatbot implementation and integration with the BI tool have given very encouraging output. Following are the benefits:

- **An intuitive way to access data**—Qlik Sense-based chatbot provides a different way to get intelligent insights into data, as compared to traditional dashboard approach. User can retrieve the insights intuitively by sending different queries or keywords to a chatbot. For example, if a user wants to see current year sales then he/she is just required to send a keyword to like “CY Sales.”
- **Reduced load time**—Users observed reduced time to send a query to a chatbot and get the key KPIs instead of opening the entire dashboard and getting the required key KPIs. Sometimes it may happen that the required KPIs are available in different sheets of an application. For example, user wants to check current year sales and last year sales and these KPIs are present in CY sheet and LY sheet, respectively, in Qlik Sense dashboard. Then, in this case, the user is required to go to CY sheet first and then go to LY sheet for checking the respective sales KPI, but in a chatbot, this can be achieved in a very time-effective way just by typing “CY vs LY sales” queries.
- **Effective end-user experience**—Human language is the most effective way of interaction between humans and computer. The human language communication using technology is a way to connect users and business. Due to this, the use of chatbot apps has become so effective that even the non-technical persons are expert at chatting. One of the best advantages of chatbots is the basic NLP features, which enables the end customer to quickly get the required information.

- **Better customer satisfaction**—Accessing KPIs or data through a chatbot provides more satisfaction to user as compared to accessing it via Web dashboard. Mainly the natural language processing component which allow the user to access data using natural English language which is found more convenient and stratifying to user.

Figure 5 shows the snapshot of telegram BI Bot with some queries. One such example of a query is “KPI” which returns the key KPI’s available in master measures of Qlik Sense application (Table 1).

Fig. 5 Snapshot of a chat window with Qlik BI BOT



**Table 1** Comparison of results of chatbot with Qlik Sense BI dashboard

Parameter	Qlik Sense BI dashboard	Chatbot
Compatibility	Qlik Sense on Web, Qlik Sense iPhone app	Telegram on Web, telegram app for windows, telegram android application, telegram iPhone app
Steps to get KPIs (data)	Multiple steps and a longer time to get KPIs data	Just need to open telegram app and type a keyword to get the KPI data
Interactive	Less interactive as it does not understand natural language	More interactive in terms of understanding nature language of user and communicating back in natural language
Speed to access data	Slower as compared to bot, require more number of clicks and redirection to multiple objects	Fast, only ask a question and get required data
Data representation	Better in dashboard, as it represents the data in terms of different charts	Average as it represents data in text format only

## 5 Future Work

Outputs from our implementation recommends that the current telegram chatbot implementation based on Dialogflow NLP approach have good properties such as speed, accessibility, compatibility, and interactivity over traditional BI dashboard.

Since the chatbot provides response in terms of text to user query, in future communication, there is a scope of adding features to show response (data or KPI) in terms of different charts. Our results are based on study of BI projects using NLP. It is anticipated that the current implementation could be further customized to suit integration with other business intelligence tools. The investigation in those respects will be described in subsequent communications.

## References

1. Vera-Baquero A, Colomo-Palacios R, Molloy O (2013) Business process analytics using a big data approach. *IT Prof* 15(6):29–35
2. Seufert A, Schiefer J (2005) Enhanced business intelligence—supporting business processes with real-time business analytics. In: Proceedings of the 16th international workshop on database and expert systems applications (DEXA'05)
3. Programming the Microsoft Bot framework: a multiplatform approach to building chatbots
4. Magic Quadrant for Analytics and Business Intelligence Platforms. <https://www.gartner.com/en/documents/3900992/magic-quadrant-for-analytics-and-business-intelligence-p>
5. Cho J, Lee G (2019) A chatbot system for construction daily report information management. In: Proceedings of the 36th international symposium on automation and robotics in construction (ISARC). Available at <http://dx.doi.org/10.22260/isarc2019/0058>



6. Shawar BA, Atwell E (2003) A corpus-based approach to generalising a chatbot system. *Procesamiento del lenguaje natural* 31
7. Deshpande A, Shahane A, Gadre D, Deshpande M, Joshi PM (2017) A survey of various chatbot implementation techniques. *Int J Comput Eng Appl* 11
8. Thomas NT (2016) An e-business chatbot using AIML and LSA. In: 2016 International conference on advances in computing, communications and informatics (ICACCI). IEEE, pp 2740–2742
9. Mondal A, Dey M, Das D, Nagpal S, Garda K (2018) Chatbot: an automated conversation system for the educational domain. In: 2018 International joint symposium on artificial intelligence and natural language processing (iSAI-NLP). IEEE, pp 1–5
10. Rosruen N, Samanchuen T (2018) Chatbot utilization for medical consultant system. In: 2018 3rd technology innovation management and engineering science international conference (TIMES-iCON), pp 1–5. IEEE Telegram software definition. [https://en.wikipedia.org/wiki/Telegram\\_\(software\)](https://en.wikipedia.org/wiki/Telegram_(software))
11. Telegram software definition. [https://en.wikipedia.org/wiki/Telegram\\_\(software\)](https://en.wikipedia.org/wiki/Telegram_(software))
12. Dialog flow definition. <https://en.wikipedia.org/wiki/Dialogflow>
13. Qlik BotNet. <https://community.qlik.com/t5/Qlik-Bot>
14. Let make this BOT. <https://community.qlik.com/t5/Qlik-Bots/Step-2b-Let-s-make-this-bot-chat/gpm-p/1469037>
15. Register Chatbot in Telegram. <https://community.qlik.com/t5/Qlik-Bots/Step-2a-Register-a-Chatbot-in-Telegram/gpm-p/1476379>

# Area Efficient Multilayer Designs of XOR Gate Using Quantum Dot Cellular Automata



Rupali Singh and Devendra Kumar Sharma

**Abstract** Quantum dot cellular automata is a well known technology which is a prospective paradigm for quantum computing. It is evident that QCA is going to be an alternative for CMOS technology for future circuits due to its property of low power, high speed and high density. Numerous digital circuits employ exclusive OR functions for executing arithmetic, error detecting or correcting operations. Thus, design of XOR gate is crucial with regard to cost efficiency and modular design competency. In QCA circuits, improved cost efficiency can be attained by minimizing one of the important parameters i.e. area of the QCA layout. Multilayer topology works successfully to reduce area and enhance the density of large circuits. This paper targets the design of multilayer XOR gate with improved cost function. Four possible QCA structures of XOR gate are proposed here, using multilayer topology. Multilayer topology assures area efficient structures. Moreover, the multiplexer and half adder circuits are presented here, using the most efficient proposed XOR design.

**Keywords** QCA · XOR · Multilayer · Adder · Multiplexer

## 1 Introduction

The number of transistors on chip is increasing constantly which in turn is restricting the performance of complementary metal oxide semiconductor (CMOS) circuits [1]. Power dissipation, leakage currents and complexity are unavoidable instances that occur as the size of the transistor shrinks. Thus, researchers are working to find substitute technology which can alter CMOS circuits and discard its shortcomings. The most appropriate way out is QCA which works on all the key parameters such as area, size and speed.

---

R. Singh (✉) · D. K. Sharma  
Department of Electronic and Communication Engineering, SRM Institute of Science and Technology, NCR Campus, Ghaziabad, India  
e-mail: [rupal.rishi@gmail.com](mailto:rupal.rishi@gmail.com)

D. K. Sharma  
e-mail: [d\\_k\\_s1970@yahoo.co.in](mailto:d_k_s1970@yahoo.co.in)

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_70](https://doi.org/10.1007/978-981-15-2329-8_70)

In QCA, there does not exist a movement of electrons or flow of current like CMOS. But, the QCA cell changes its state of polarization affecting the adjacent cells and thus, passing information from one end to other [2]. The change in polarization state occurs due to Coulombic interaction between the electrons [3]. This feature of QCA makes it suitable for ultra low power applications. There are some realistic designs of QCA circuit fabricated using semiconductor, magnetic and molecular technology operating at room temperature. Yet many traits of QCA need to be explored.

Some important arithmetic circuits such as adders [4–7], ripple carry adders [8–10] subtractors [11–13], multiplexers [14–16], encoders [17, 18] are designed using QCA. Many other sequential circuits such as latches and flip flops [10, 19–21], shift registers [22–24], counters [25, 26] and many more are implemented in QCA. In most of the circuits, XOR gate is a crucial element and in most of the papers the coplanar QCA structures are used for implementation. While designing the large circuit, crossovers are required to establish interface between two modules. The crossovers can be of two types, coplanar or multilayer. The coplanar crossovers are prone to crosstalk between the adjacent wires; on the other hand, multilayer crossovers are immune. Moreover, active components can be placed on different layers in multilayer QCA unlike CMOS [27]. The multilayer structures of QCA are not much addressed due to complexity involved in its design procedure. This paper targets the multilayer designs of XOR gate optimized in area and QCA cell count. The paper is organized into five sections. Section 2 gives review of QCA and multilayer circuits. Section 3 describes multilayer XOR designs in QCA with cost analysis and Sect. 4 presents design of adder and comparator using efficient XOR gate. Section 5 concludes the paper.

## 2 Review of QCA

The basic element of QCA is a cell which is a square shaped formation of four quantum dots situated at four corners. Two diagonal cells possess electrons which tunnel between the dots when barrier potentials are lowered. Depending on the position of electrons, two polarization states are defined as binary logic 1 and logic 0 as given in Fig. 1. These QCA cells are arranged in particular order to form various QCA devices such as majority gate, inverter, logic gates and interconnects. QCA wire is an arrangement of QCA cells placed adjacent to each other as given in Fig. 1b. Majority voter is three inputs and one output gate with the output equal to the majority of inputs. Majority gate can function as AND or OR gate by fixing one of its input to logic 1 or 0, respectively.

When the QCA cells at output terminal are placed at an angle of  $45^\circ$  to the input cells, the inverted logic is obtained. All the basic QCA structures are shown in Fig. 1. The information flow in QCA is achieved using a specific clock signal [28]. This is an adiabatic clock which ensures the directed flow of logic state from input cell

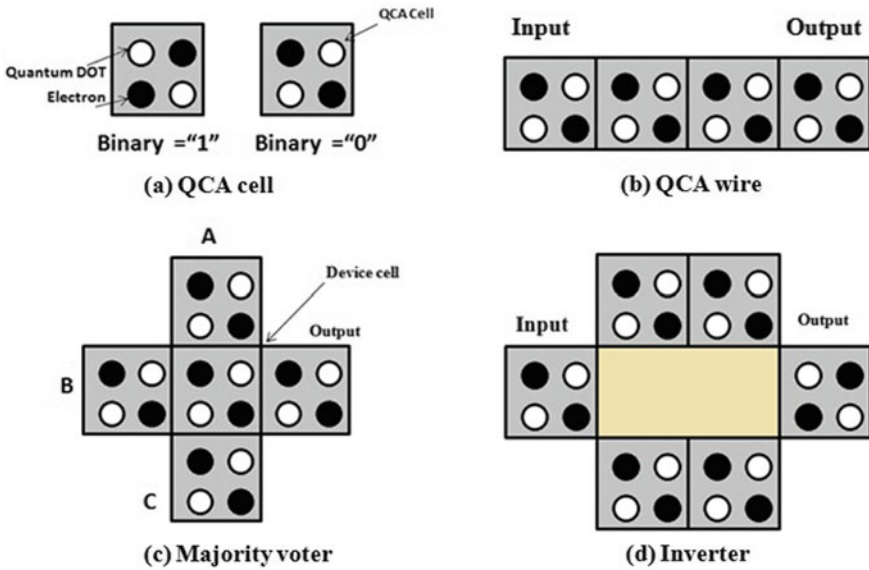


Fig. 1 Elementary QCA structures

to output QCA cell. It involves four clocking zones and each zone has four phases: switch, hold, release, relax as shown in Fig. 2.

Further, multilayer QCA structures are formed when QCA cells are arranged over different layers. If the cells are positioned adjacent to each other, the polarization of each cell becomes same according to the input cell. But if the cells are facing each other at different layers then their polarization gets reversed as shown in Fig. 3. Thus, edge wise aligned cells show similar polarization while facing cells show opposite polarizations.

The possibility of alternate arrangement of QCA cells was explored in [29]. Figure 3 illustrates the majority voter and its representation at different layers. Each cell is placed at different cell and accordingly the equation of majority voter changes. It shows that active components can be placed at different layers to form diverse structures which can attain numerous logical functions.

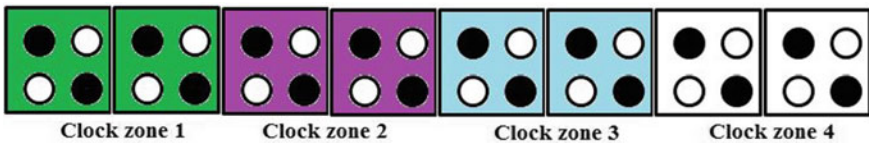


Fig. 2 Clocking zones in QCA

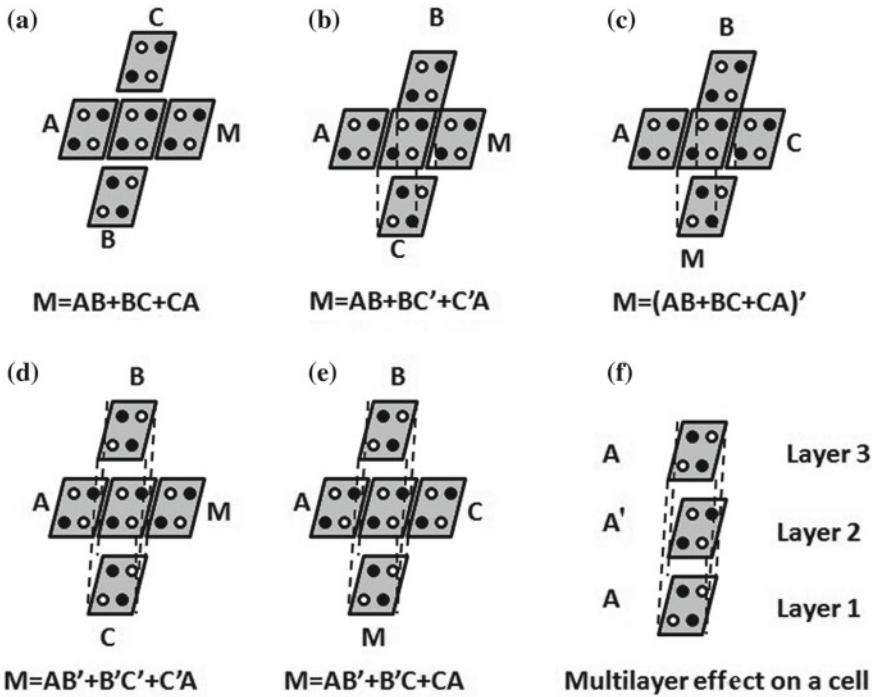


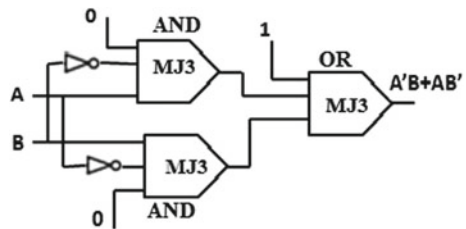
Fig. 3 Multilayer QCA structures

### 3 Proposed Multilayer XOR Structures

Exclusive OR gate is one of the crucial logic gate which is required in designing of many combinational and sequential circuits. This section gives the design of proposed multilayer, area efficient XOR structures. Figure 4 shows the schematic of XOR function realized using three majority voters. All the structures are realized using two AND and one OR gates as shown in Fig. 4. The output function of XOR gate is given by Eq. (1).

$$Z = AB' + A'B \tag{1}$$

Fig. 4 Schematic of XOR gate using 3 input majority gates



The QCA implementation of this circuit using multilayer paradigm is shown in Fig. 5. It requires three layers. Layer 1 is carrying two majority AND gates with inputs *A* and *B*, layer 2 is depicting via connects between layer 1 and layer 3 and layer 3 is showing majority OR gate. The output *Z* is obtained at layer 1 with the delay of 0.75 clocks (3 clock zones).

Figure 6 is another structure using different analogy in placement of the majority voters. Here, one majority AND gate with input *A* is placed in layer 1 and another with input *B* is placed in layer 2. OR gate with the output *Z* is accommodated in layer 1 itself and layer 3 carrying constant inputs. The latency of the circuit is 0.75.

Another three-layer XOR structure is proposed in this paper as shown in Fig. 7. This configuration of XOR gate shows different cell arrangement of majority voters at layer 1 and layer 3. Layer 2 has via connects between the two layers. The latency of the circuit is 0.75 clocks.

The proposed circuits are designed and simulated on QCA Designer 2.0.3. The simulation waveform of XOR gate is given in Fig. 8. The bistable simulation engine

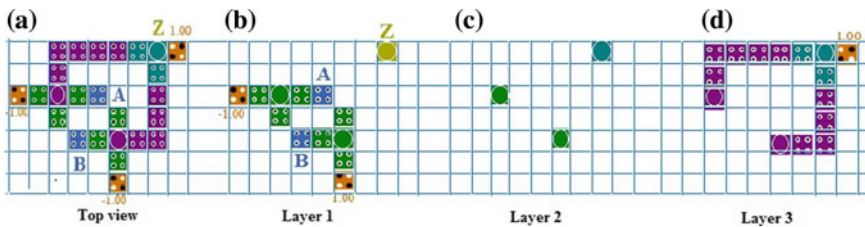


Fig. 5 Multilayer XOR1 design. a Top view. b Layer 1. c Layer 2. d Layer 3

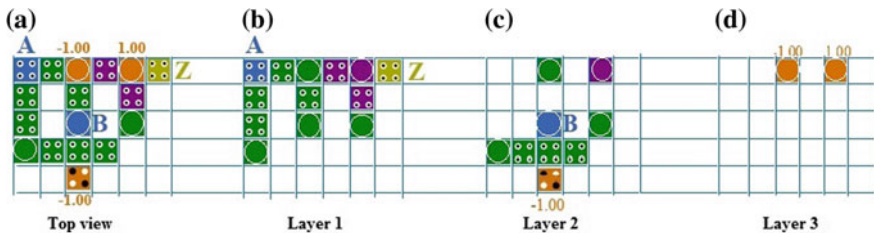


Fig. 6 Multilayer XOR2 design. a Top view. b Layer 1. c Layer 2. d Layer 3

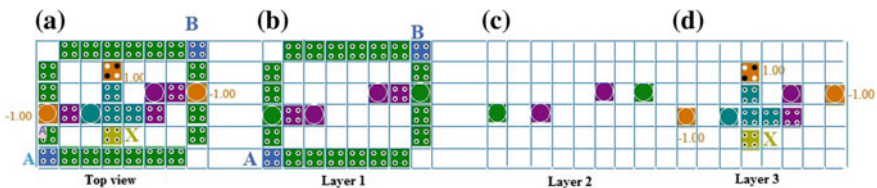


Fig. 7 Multilayer XOR3 design. a Top view. b Layer 1. c Layer 2. d Layer 3

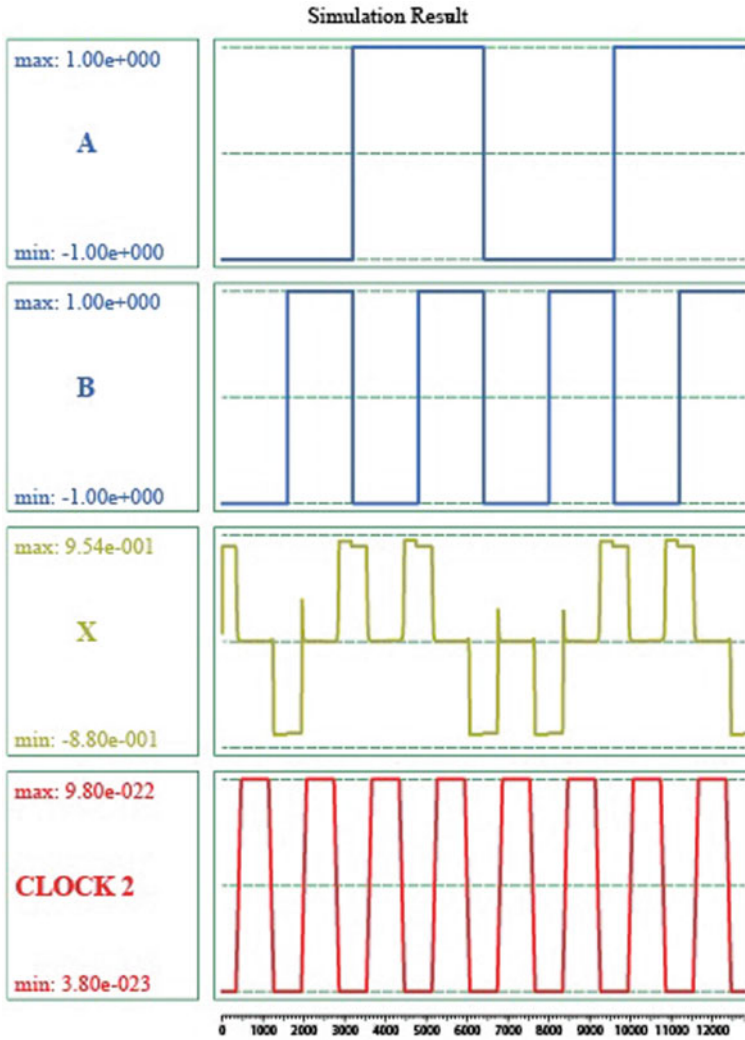


Fig. 8 Simulation waveform of multilayer XOR2

is used to obtain the simulations of proposed circuits. In all the above circuits, three input majority gate is used as an elementary component of XOR gate. The XOR function can be realized using five input majority gate also. The schematic for XOR gate using five input majority voter is shown in Fig. 9 [30]. The output function of five input majority voter is given by Eq. (2).

$$\begin{aligned}
 M_j(A, B, C, D, E) = & ABC + ACD + ADE + BCD + BCE \\
 & + BDE + CDE + ACE + ABD + ABE \quad (2)
 \end{aligned}$$

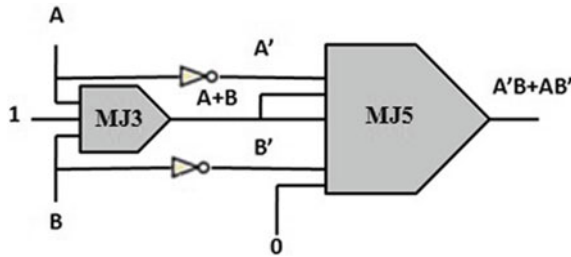


Fig. 9 Schematic of XOR gate using five input majority gate

$$X = Mj(A', A + B, A + B, B', 0) = A'B + AB' \tag{3}$$

Equation (3) shows that five input majority function can be used as XOR using specified inputs. The QCA layout for the schematic in Fig. 9 can be realized as illustrated in Fig. 10. The proposed circuit utilizes five layers to realize the XOR function. The circuit utilizes five input majority gate and one three input majority gate. Latency of the circuit is 0.75 clocks. The simulation waveform of the proposed QCA layout is shown in Fig. 11.

All the XOR structures presented here have unique cell arrangement utilizing lesser area. The QCA analysis of the proposed circuit is shown in Table 1. To scrutinize the performance of the proposed circuits, cost function is evaluated using Eq. (4).

$$\text{Cost function} = \text{cellcount} \times \text{area} \times \text{latency} \tag{4}$$

It is evident in Table 1 that all the proposed XOR structures have comparatively improved cost as compared to many existing XOR gates. Particularly, the proposed XOR2 design has shown the superior performance in terms of cell count, area and cost function as compared to the other existing designs in the literature.

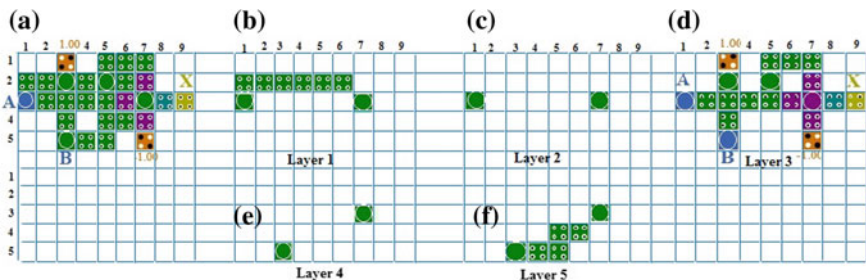


Fig. 10 Multilayer XOR4 design using five input majority voter. a Top view. b Layer 1. c Layer 2. d Layer 3. e Layer 4. f Layer 5



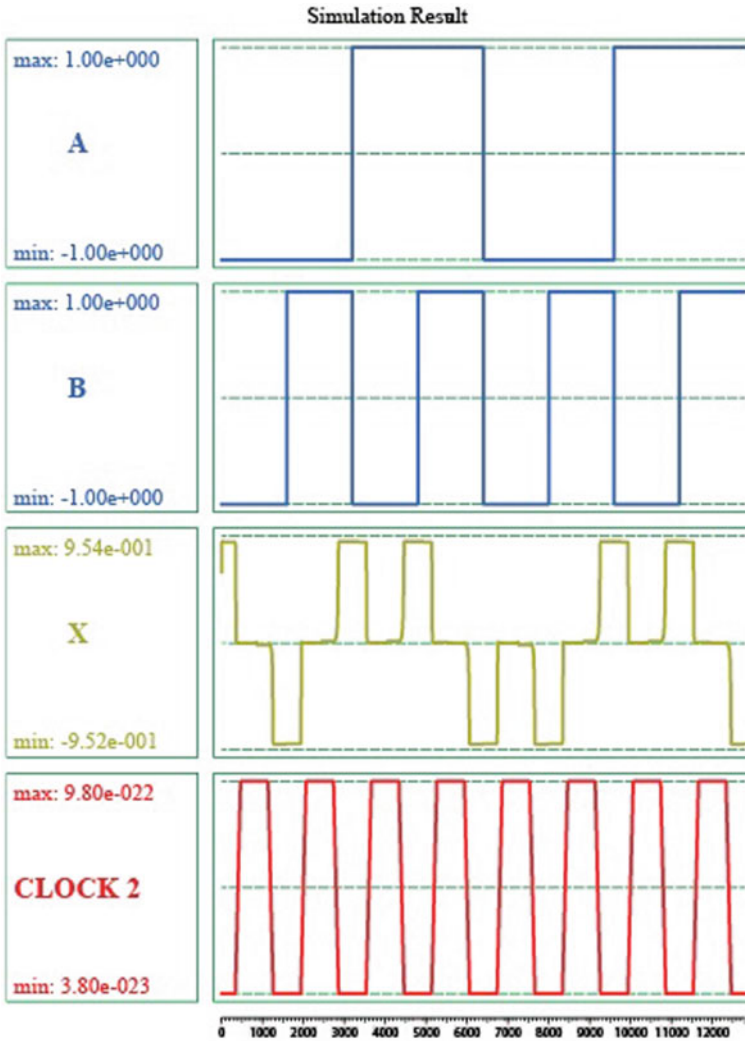


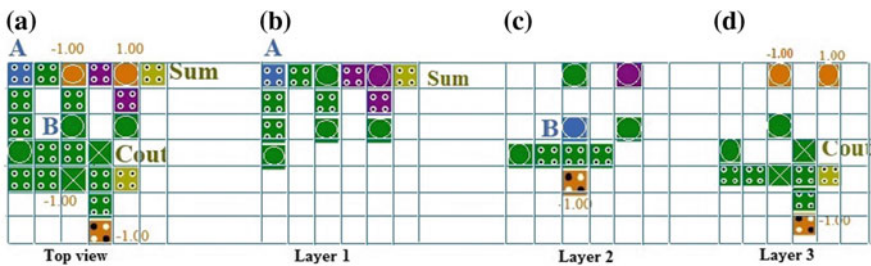
Fig. 11 Simulation waveform of multilayer XOR4 design

### 4 Design of Combinational Circuits

This section presents the design of few combinational circuits such as multiplexer and half adder. These circuits are fundamental modules which can be used to form large circuits. Figure 12 illustrates the proposed circuit for multilayer half adder which is designed using XOR2 gate and Fig. 13 gives its simulation waveform.

**Table 1** Comparative analysis of proposed XOR gate designs

XOR designs	Cell count	Area	Latency	Cost function	Type
Walus [27]	85	0.078	1	6.63	Multilayer (only crossover)
Angizi [31]	67	0.06	1.25	5.03	Single layer
Suresh [9]	45	0.03	0.75	1.01	Single layer
Beigh design 1 [32]	54	0.07	1	3.78	Single layer
Beigh design 2 [32]	48	0.06	0.5	1.44	Single layer
Beigh design 3 [32]	42	0.05	0.5	1.05	Single layer
Kianpour [13]	49	0.05	1	2.45	Single layer
Mohammadi [33]	39	0.03	0.75	0.87	Single layer
Poorhosseini design 1 [34]	45	0.05	1	2.25	Single layer
Poorhosseini design 1 [34]	37	0.03	1	1.11	Single layer
Ahmad [35]	32	0.03	0.75	0.72	Multilayer (only crossover)
Proposed XOR1	31	0.04	0.75	0.93	Multilayer
Proposed XOR2	24	0.02	0.75	0.36	Multilayer
Proposed XOR3	40	0.04	0.75	1.2	Multilayer
Propose XOR4 with MJ5	38	0.03	0.75	0.85	Multilayer



**Fig. 12** Multilayer half adder. **a** Top view. **b** Layer 1. **c** Layer 2. **d** Layer 3

In the proposed half adder circuit,  $A$  and  $B$  are inputs while  $Sum$  and  $C_{out}$  are outputs with the latency of 0.75 and 0.5 clock, respectively. The proposed QCA layout uses 34 QCA cells with  $0.03 \mu m^2$  area. The output functions for  $Sum$  and  $C_{out}$  are given by Eqs. (5) and (6).

$$Sum = AB' + A'B \tag{5}$$

$$C_{out} = AB \tag{6}$$

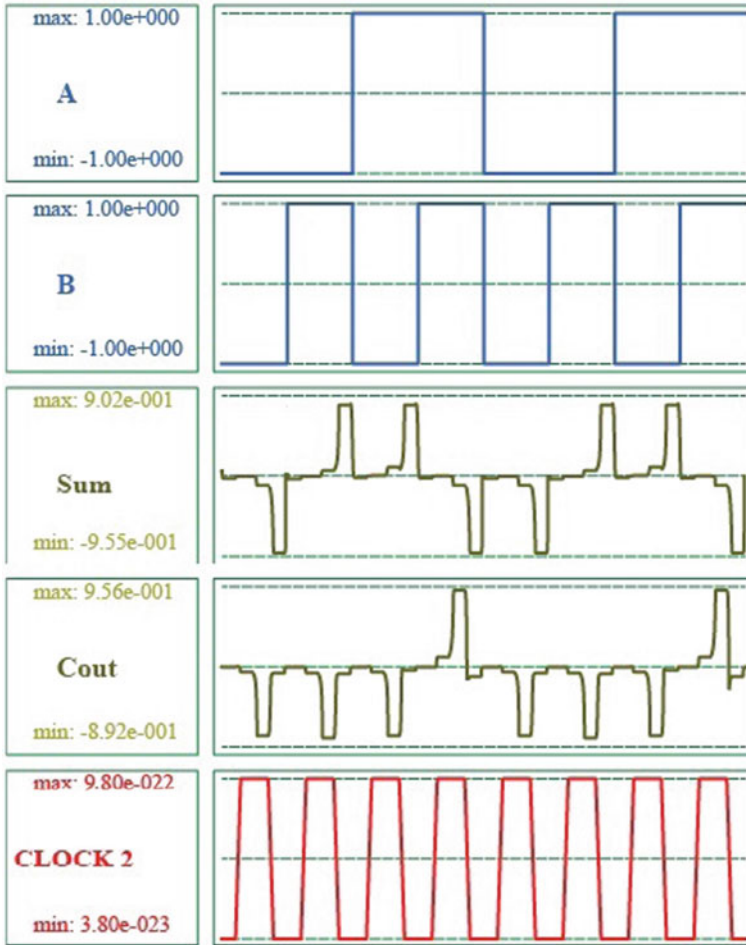


Fig. 13 Simulation waveform for multilayer half adder

Further, the proposed multilayer 2:1 multiplexer using XOR2 is shown in Fig. 14 and its simulation waveform is depicted in Fig. 15. A and C are input terminals while B is selection line in the proposed circuit. The proposed QCA layout of multilayer 2:1 multiplexer utilizes 21 QCA cells with the area of  $0.02 \mu\text{m}^2$ . The output function is given by Eq. (7).

$$Z = AB' + BC \tag{7}$$

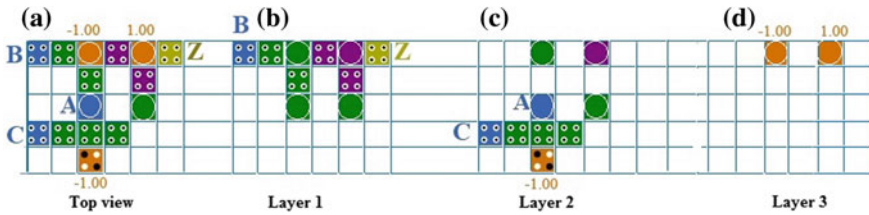


Fig. 14 Multilayer 2:1 multiplexer. a Top view. b Layer 1. c Layer 2. d Layer 3

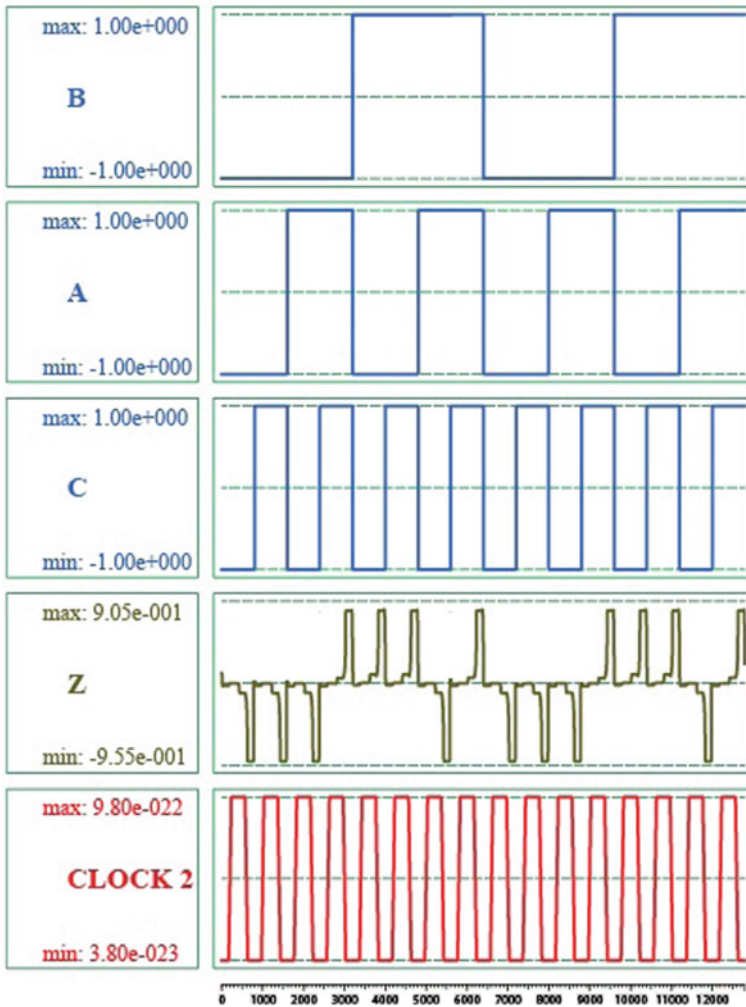


Fig. 15 Simulation waveform for multilayer 2:1 multiplexer

## 5 Conclusion

This paper presents the design of XOR gate using multilayer paradigm of QCA. Four designs of XOR gate are proposed here. Three designs are using 3 input majority gates while fourth design makes use of 5 input majority voter for the formation of XOR gate QCA layout. The proposed designs are area as well as cost efficient and have shown improved performance. The proposed XOR2 design has shown an improvement of 22.58% in cell count, 33.33% in area and 50% in cost function as compared to the best XOR design reported earlier. This XOR design is further utilized to design multilayer half adder and 2:1 multiplexer structures. Thus, the proposed multilayer XOR designs can be efficiently used to design any combinational or sequential circuits. It is apparent from the results that multilayer paradigm can be used to design area efficient QCA circuits with improved resistance towards signal interference.

## References

1. Compagno R, Molenkamp L, Paul DJ (2001) Technology roadmap for nanoelectronics. European Commission IST Programme, vol 5, 2nd edn, pp 1–107
2. Lent CS, Taugaw PD, Porod W, Bernstein GH (1993) Quantum cellular automata. *Nanotechnol* 4(1):49–57
3. Lent CS, Tougaw PD (1997) A device architecture for computing with quantum dots. *Proc IEEE* 85(4):541–557
4. Kassa SR, Nagaria RK (2016) An innovative low power full adder design in nano technology based quantum dot cellular automata. *J Low Power Electron* 12(2):1–6
5. Sasamal TN, Kumar A, Mohan A (2018) An efficient design of quantum-dot cellular automata based 5-input majority gate with power analysis. *Microprocess Microsyst* 59:103–117
6. Roshany HR, Rezai A (2019) Novel efficient circuit design for multilayer QCA RCA. *Int J Theor Phys* 58(6):1745–1757
7. Wang L, Xie G (2019) A power-efficient single layer full adder design in field-coupled QCA nanocomputing. *Int J Theor Phys* 58(7):2303–2319
8. Wang L, Xie G (2018) Novel designs of full adder in quantum-dot cellular automata technology. *J Supercomput* 74(9):4798–4816
9. Suresh K, Ghosh B (2013) Ripple carry adder using two XOR gates in QCA. *Appl Mech Mater* 467:531–535
10. Roshan MG, Gholami M (2018) Novel D latches and D flip-flops with set and reset ability in QCA nanotechnology using minimum cells and area. *Int J Theor Phys* 57(10):3223–3241
11. Abdullah-Al-Shafi M, Bahar AN (2018) An architecture of 2-dimensional 4-dot 2-electron QCA full adder and subtractor with energy dissipation study. *Acta Passiva Electron Compon*
12. Ahmad F, Ahmed S, Kakkar V, Bhat GM, Newaz A, Wani S (2018) Modular design of ultra-efficient reversible full adder-subtractor in QCA with power dissipation analysis. *Int J Theor Phys* 57(9):2863–2880
13. Kianpour M, Sabbaghi-Nadooshan R, Navi K (2014) A novel design of 8-bit adder/subtractor by quantum-dot cellular automata. *J Comput Syst Sci* 80(7):1404–1414
14. Mosleh SAM (2018) A novel fault-tolerant multiplexer in quantum-dot cellular. *J Supercomput* 74(9):4696–4716
15. Rashidi H, Rezai A, Soltany S (2016) High-performance multiplexer architecture for quantum-dot cellular automata. *J Comput Electron* 15(3):968–981

16. Ahmad F (2017) An optimal design of QCA based  $2n:1/1:2n$  multiplexer/demultiplexer and its efficient digital logic realization. *Microprocess Microsyst* 56:64–75
17. Debnath B, Das JC, De D (2017) Reversible logic-based image steganography using quantum dot cellular automata for secure nanocommunication. *IET Circ Devices Syst* 11(1):58–67
18. Newaz A, Billah M, Maksudur M, Bhuiyan R (2018) Ultra-efficient convolution encoder design in quantum-dot cellular automata with power dissipation analysis. *Alexandria Eng J* 57(4):3881–3888
19. Boi B, Misra NK, Pradhan M (2017) Design and evaluation of an efficient parity-preserving reversible QCA gate with online testability. *Cogent Eng* 4(1):1–18
20. Seyedi S, Darbandi M, Navimipour NJ (2019) Designing an efficient fault tolerance D-latch based on quantum-dot cellular automata nanotechnology. *Opt–Int Light Electron Opt* 185:827–837
21. Bahar AN, Laajimi R, Ahmed AK (2018) Toward efficient design of flip-flops in quantum-dot cellular automata with power dissipation analysis. *Int J Theor Phys* 57(11):3419–3428
22. Chandra J, Debashis D (2017) Operational efficiency of novel SISO shift register under thermal randomness in quantum-dot cellular automata design. *Microsyst Technol* 23(9):4155–4168
23. Sabbaghi-Nadooshan R, Kianpour M (2014) A novel QCA implementation of MUX-based universal shift register. *J Comput Electron* 13(1):198–210
24. Niknezhad M, Abdalhossein D, Asghar R (2018) Towards multilayer QCA SISO shift register based on efficient D-FF circuits. *Int J Theor Phys*
25. Niknezhad Divshali M, Rezaei A, Falahieh Hamidpour SS, (2019) Design of novel coplanar counter circuit in quantum dot cellular automata technology. *Int J Theor Phys*:2677–2691
26. Newaz A et al (2017) Designing single layer counter in quantum-dot cellular automata with energy dissipation analysis. *Ain Shams Eng J*
27. Walus K, Schulhof G, Jullien GA (2004) High level exploration of quantum-dot cellular automata (QCA). In: *Thirty-Eighth Asilomar conference on signals, systems and computers*. IEEE, pp 30–33
28. Hennessy K, Lent CS (2001) Clocking of molecular quantum-dot cellular automata. *J Vac Sci Technol B Microelectron Nanom Struct* 19(5):1752
29. Gin A, Tougaw PD, Williams S (1999) An alternative geometry for quantum-dot cellular automata. *J Appl Phys* 85(12):8281–8286
30. Singh G, Sarin RK, Raj B (2016) A novel robust exclusive-OR function implementation in QCA nanotechnology with energy dissipation analysis. *J Comput Electron* 15(2):455–465
31. Angizi S, Alkaldy E, Bagherzadeh N, Navi K (2014) Novel robust single layer wire crossing approach for exclusive OR sum of products logic design with quantum-dot cellular automata. *J Low Power Electron* 10(2):259–271
32. Beigh MR, Mustafa M, Ahmad F (2013) Performance evaluation of efficient XOR structures in quantum-dot cellular automata (QCA). *Circ Syst* 4(April):147–156
33. Mohammadi H, Navi K (2018) Energy-efficient single-layer QCA logical circuits based on a novel XOR gate. *J Circ Syst Comput Art*
34. Poorhosseini M, Hejazi AR (2018) A fault-tolerant and efficient XOR structure for modular design of complex QCA circuits. *J Circ Syst Comput* 27(07):1850115
35. Ahmad F, Din Bhat GM (2015) Design of novel inverter and buffer in quantum-dot cellular automata (QCA). In: *International conference on computing for sustainable global development, INDIACom 2015*, pp 67–72

# Improved Design of Digital IIR Second-Order Differentiator Using Genetic Algorithm



Amit Bohra, Rohit Sharma and Vibhav Kumar Sachan

**Abstract** In this paper, a new design of digital differentiator of the second order using genetic algorithm is presented. By the use of genetic algorithm, transfer function differentiator of second order is derived. Then compare the results with the ideal second-order differentiator. Result is also compared with digital IIR differentiator second order using backward difference formula which is already exist.

## 1 Introduction

Digital differentiator is extremely helpful to approximate and determine the time differentiation of given signals. For example, Laplacian operator is used to detect the edge of the image in image processing [1]. In radars, the boost can be calculated from the place measurements using second-order differentiator [2]. In medical engineering, it is a must to obtain the higher-order differentiation of medical data [3]. At present, there are many methods available to design differentiator of second order like eigenfilter [4] method and limit computation method [5]. The response of frequency of an ideal differentiator of second order is given by

$$D(\omega) = (j\omega)^2 \quad (1)$$

Now we want to design second-order digital differentiator for minimum relative error with the ideal as in (1). When the difference formulas are used, then accuracy is not good. So to get more accurate result, we use the genetic algorithm. Genetic algorithm is a very good optimization technique.

---

A. Bohra (✉) · R. Sharma  
SRMIST, NCR, Ghaziabad, India  
e-mail: [amitpbohra@gmail.com](mailto:amitpbohra@gmail.com)

R. Sharma  
e-mail: [rohitapece@gmail.com](mailto:rohitapece@gmail.com)

A. Bohra · V. K. Sachan  
KIET Group of Institutions, Ghaziabad, India

© Springer Nature Singapore Pte Ltd. 2020  
D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_71](https://doi.org/10.1007/978-981-15-2329-8_71)

In this paper, Section I covers the theoretical background to design digital differentiator of second order, and design methodology is given in Section II. Finally, the simulated results are discussed in the last section.

## 2 Theory

By using backward difference formula, the second-order differentiator [6] is presented by Eq. (2)

$$A_0(e^{j\omega}, \alpha) = \frac{(1 - 2e^{-j\alpha\omega} + e^{-2j\alpha\omega})}{\alpha^2} \quad (2)$$

By the help of Taylor series expansion of exponential function, the frequency response can be rearranged as

$$A_0(e^{j\omega}, \alpha) = \frac{\left(1 - 2 \sum_{k=0}^{\infty} \frac{(-j\alpha\omega)^k}{1k} + \sum_{k=0}^{\infty} \frac{(-2j\alpha\omega)^k}{1k}\right)}{\alpha^2} \quad (3)$$

$$A_0(e^{j\omega}, \alpha) = D(\omega) + \sum_{k=1}^{\infty} a_k \alpha^k \quad (4)$$

$$A_0(e^{j\omega}, \alpha) = D(\omega) + 0(\alpha) \quad (5)$$

where  $0(\alpha)$  indicates the error term which depends upon  $\alpha$ . As fast as  $\alpha$  decay, the error term also decay. If the parameter  $\alpha$  tends to zero, we have the given result

$$\lim_{\alpha \rightarrow 0} A_0(e^{j\omega}, \alpha) = D(\omega) \quad (6)$$

From using Richardson extrapolation [6], we can find Eq. (7)

$$A_k(Z, \alpha) = \frac{2^k A_{k-1}(Z, \alpha) - A_{k-1}(Z, \alpha)}{2^k - 1} \quad (7)$$

The frequency response error is given by Eq. (8)

$$E(\omega) = 20 \log_{10}(|A_k(e^{j\omega}, \alpha) - D(\omega)|) \quad (8)$$



### 3 Design Methodology

By the use of genetic algorithm, the error term minimized is given by

$$E_{\text{diff}} = \int_0^{\pi} (\omega^2 - |H(e^{j\omega})|)^2 d\omega \quad (9)$$

This error term  $E_{\text{diff}}$  is basically the difference between the second-order differentiator's energy and the proposed second-order differentiator's energy taken for the period  $[0, \pi]$ . Writing the equation in the form

$$H_{\text{diff}} = \frac{s(1) + s(2)Z^{-1} + s(3)Z^{-2} + s(4)Z^{-3} + s(5)Z^{-4} + s(6)Z^{-5} + s(7)Z^{-6}}{1 + s(8)Z^{-1} + s(9)Z^{-2} + s(10)Z^{-3} + s(11)Z^{-4} + s(12)Z^{-5} + s(13)Z^{-6}} \quad (10)$$

The selection of genetic algorithm (GA) parameters (selection mechanism, crossover and mutation rate) is problem-dependent. Generally, GA practitioners preferred tournament selection. The values of crossover and mutation rates are set to 0.9 and 0.1. These values decide the trade-off between exploration and exploitation of solutions during evolutionary process. GA uses fixed-length string (chromosome/individual) for encoding solution of problem. You can decide the size of the individual (chromosome) based on the problem you are trying to solve. However, as the size of individual increases, it adversely affects the convergence rate of GA.

Thus, the values of 13 coefficients are calculated with the help of genetic algorithm and observation as shown in Table 1.

All these differentiators' magnitude response is not plotted in the given table because few of these overlap intermittently and the graph looks very complicated.

From the results, we can say easily that the result obtained using genetic algorithm is best than the backward difference formula using Richardson Extrapolation.

### 4 Simulated Results

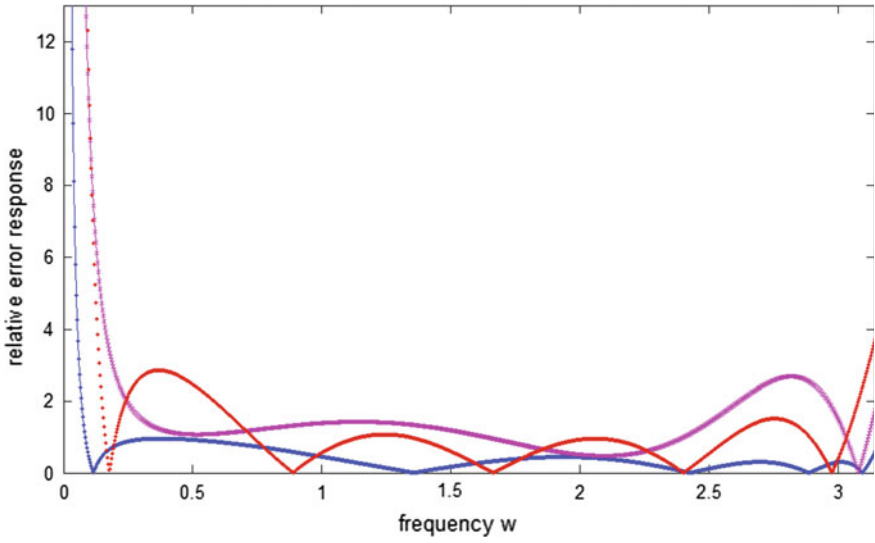
See Figs. 1, 2 and 3.

### 5 Conclusion

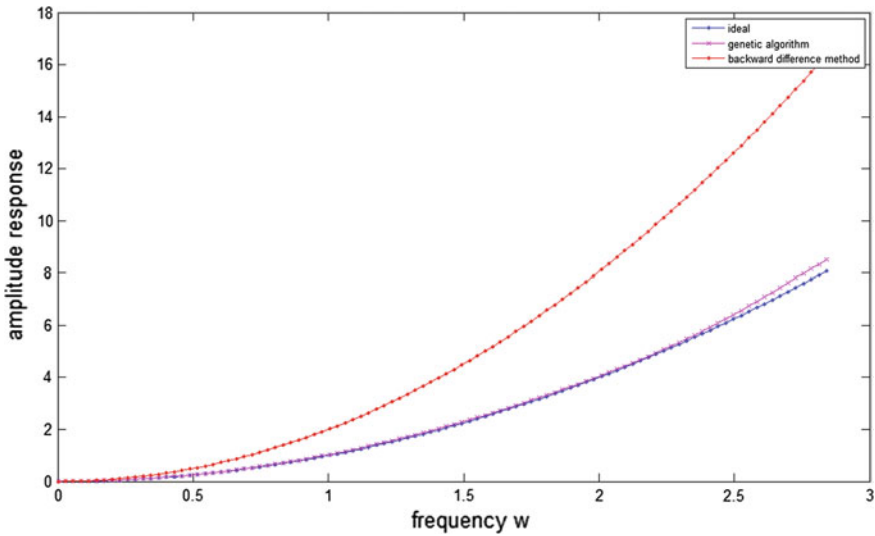
The improved design of differentiator of second order using genetic algorithm has been given in this paper. First, the transfer function of differentiator of second order finds by using genetic algorithm. Then, compare its relative error with the error of

**Table 1** Values of 13 coefficients calculated with the help of genetic algorithm and observation

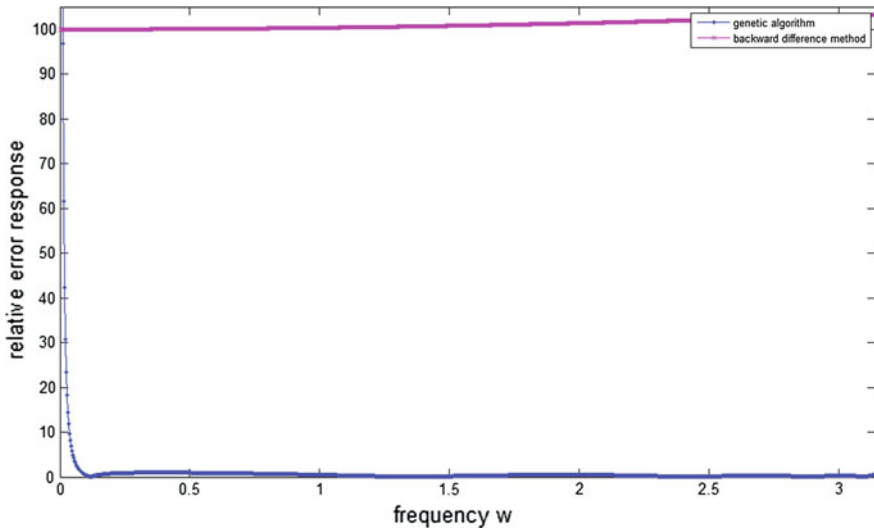
S. No.	S(1)	S(2)	S(3)	S(4)	S(5)	S(6)	S(7)	S(8)	S(9)	S(10)	S(11)	S(12)	S(13)
1	5.479	-3.933	1.2711	1.401	0.2351	-0.0810	0.0149	1.8737	3.3395	-0.3172	0.0358	-0.0005	-0.000012
2	4.385	-2.984	2.6529	1.0051	1.276	-1.102	0.0274	2.1083	4.0018	-2.917	0.0381	-0.0019	-0.000053
3	6.562	-4.764	0.6509	-2.498	-0.0671	0.7910	-1.291	0.8721	2.9193	-4.075	0.1932	-0.0027	0.000927
4	2.876	-1.654	-1.634	2.7619	0.5817	0.4612	0.00543	0.7287	-1.028	-2.017	0.671	-0.1001	-0.00005



**Fig. 1** Relative error response  $E_{diff} = \int_0^{\pi} (\omega^2 - |H(e^{j\omega})|)^2 d\omega$  for few curves using genetic algorithm



**Fig. 2** Amplitude response for ideal, genetic algorithm and backward difference formula second order differentiator



**Fig. 3** Relative error response curve for genetic algorithm second-order differentiator and backward difference formula second-order differentiator

the differentiator of second-order transfer function which is derived from backward difference formula. And get the best results. Then, compare the amplitude response of both differentiator of second order with the amplitude response of the ideal differentiator second order, and from these results, we can say that we improved the design of differentiator of second order.

## References

1. Gonzluez RC, Woods RE (2002) Digital image processing, 2nd edn, Prentice Hall
2. Skolnik MI (1980) Introduction to radar systems. McGraw-Hill, New York
3. Usui S, Amidror I (1982) Digital low-pass differentiation for biological signal processing. *IEEE Trans Biomed Eng* 29:686–693
4. Pei SC, Shyu JJ (1989) Eigenfilter design of higher-order digital differentiators. *IEEE Trans Acoust. Speech Signal Process* 37(April):505–511
5. Tseng CC (2002) Digital differentiator design using fractional delay filter and limit computation. *IEEE Trans Circuits Syst I* 52:2248–2259
6. Tseng C-C, Lee S-L (2008) Digital IIR Integrator using Richardson Extrapolation and fractional delay. *IEEE Trans Circuits Syst I* 55(8):2300–2309

# Easy Synthesis of Nanostructures of ZnO and ZnS for Efficient UV Photodetectors



Vipin Kumar, Ishpal Rawal and Vinod Kumar

**Abstract** Here, we report the facile synthesis of oxide (ZnO) and sulfide (ZnS) of zinc in nanorods and nanotubular structures by simple hydrothermal reduction method. The structural and phase analysis of the prepared samples has been done through X-ray diffraction, whereas morphological investigations have been done through transmission electron microscopy studies. The optical band gaps of the synthesized materials have been examined through UV-Vis spectroscopy studies. Photoconductivity measurements have been performed of both the ZnO nanorods and ZnS nanotubes samples in UV-illumination ( $\lambda \approx 365$  nm) and at an illumination intensity of  $\sim 3.3$  mW/cm<sup>2</sup>. The prepared ZnO nanorods are found to have greater photoresponse of  $\sim 30\%$  as compared to ZnS nanotubes  $\sim 25\%$ . The adhesion and removal of oxygen molecules on the prepared samples' surface are considered to be the mechanism of photodetection.

**Keywords** UV photodetection · Nanorods · Nanotubes

## 1 Introduction

In recent years, considerable efforts have been made on synthesis of semiconducting nanostructures due to their potential applications in the fabrication electronic and photonic nanodevices such as nanolasers, nanosensors, field-effect transistors and nanocantilevers [1–3]. The oxides and sulfides of the different metals are considered to be the most promising materials for these applications due to their wide optical band gaps, high environmental stability and high device performance for a longer time. Out of these metal oxides or sulfides, zinc oxide (ZnO) and zinc sulfide (ZnS) stand out due to their exceptional device efficiency, easy synthesis process, better

---

V. Kumar · V. Kumar (✉)

Department of Electronics and Communication Engineering, SRM-Institute of Science and Technology, Gaziabad, UP, India  
e-mail: [vinodkur1@srmist.edu.in](mailto:vinodkur1@srmist.edu.in)

I. Rawal (✉)

Department of Physics, Kirori Mal College, University of Delhi, Delhi 110007, India

© Springer Nature Singapore Pte Ltd. 2020

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering*, Lecture Notes in Networks and Systems 106,  
[https://doi.org/10.1007/978-981-15-2329-8\\_72](https://doi.org/10.1007/978-981-15-2329-8_72)

environmental stability in harsh conditions. Moreover, the efficiency of these devices can be improved through their nanostructural forms, and the properties of nanostructured materials are quite different from their bulk counterpart due to their size and quantum confinement effects. Therefore, the emerging nanostructure industry has been focused their intensive research in building blocks for nanoelectronic and photonic systems [4–7]. The high optical band gap of ZnO (3.37 eV) and ZnS (3.72 eV for zinc blende and 3.77 eV for ZnS-wurtzite structure) and high excitation energy (~60 meV for ZnO and 40 meV for ZnS) make them suitable candidate for the visible blind photodetectors generally employed in UV region for the different civilian and military applications [8–10]. Therefore, in the present study, the nanostructures of ZnO and ZnS are prepared by chemical reduction method and employed for the fabrication of UV photodetectors.

## 2 Experimental Details

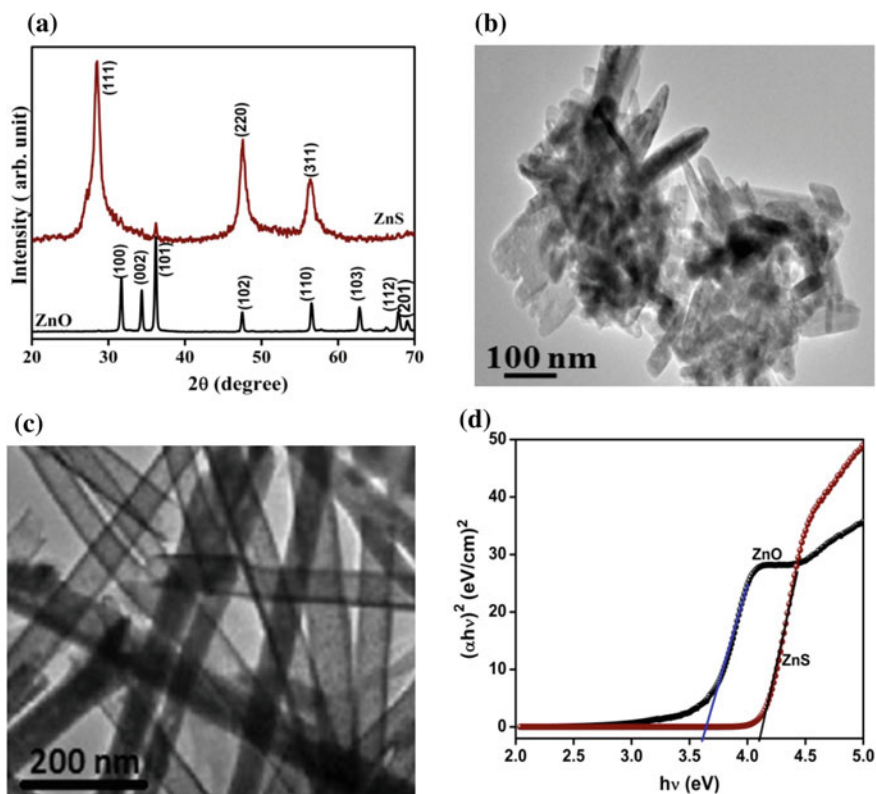
For the preparation of ZnO nanorods, 0.5 M solution of zinc acetate dehydrate ( $\text{Zn}(\text{CH}_3\text{COO})_2 \cdot 2\text{H}_2\text{O}$ ) was prepared in deionized water and mixed with separately prepared 0.5 M solution of KOH and 1 M ammonia for 10 min and then the solution was autoclaved for 6 h at 100 °C in an autoclave with Teflon cavity. The mixture was then allowed to cool naturally and the precipitates were filtered out and parched at 60 °C. The obtained sample was further annealed at 500 °C for 3 h. Similar method has also been adopted for the synthesis of ZnS nanotubes using 0.5 M solution of thioacetamide (TAA) in place of 0.5 M solution of KOH.

The structural properties and phase analysis of the prepared samples have been done through X-ray diffraction (XRD) performed at the ASX-Bruker made X-ray diffractometer (D8 Discover Advance), whereas morphological analysis of the prepared materials has been done through FEI made high-resolution transmission electron microscope (HRTEM) (model- Tecnai G2F30-STWIN). The optical characteristics of the synthesized materials were investigated through UV-Vis-NIR spectrophotometer made by Varian Netherlands, model no. Cary-5000. The performance of the prepared UV photodetectors is examined under UV source of wavelength ~365 nm at an intensity of ~3.3 mW/cm<sup>2</sup>. To test the device performance, two parallel contacts of silver are painted over the top surface of the sample pellets and the change in transient electric current has been noted through Keithley 2410 instrument connected with computer.

### 3 Results and Discussion

#### 3.1 Structural, Morphological and Optical Properties

Structural properties and phase analysis of the prepared samples have been studied through XRD technique. Figure 1a elucidates the XRD patterns of the prepared ZnO nanorods, and ZnS nanotubes. It is observed that the XRD pattern of ZnO exhibits the all the characteristic XRD peaks of the ZnO located at  $\sim 68.72^\circ$ ,  $67.56^\circ$ ,  $62.5^\circ$ ,  $56.24^\circ$ ,  $47.14^\circ$ ,  $35.84^\circ$ ,  $34.00^\circ$  and  $31.34^\circ$  which are indexed to the (hkl) planes (201), (112), (103), (110), (102), (101), (002) and (100), respectively [8, 9, 11] and well corroborated with JCPDF file no. 06-2151, 14, whereas the XRD pattern of ZnS nanotubes contains three major peaks observed at  $28.8^\circ$ ,  $48.14^\circ$  and  $57.02^\circ$  corresponding to the (hkl) planes (111), (220) and (311) of zinc blende [10] and well matched with JCPDS data files no. 80-5520. No extra peak of any impurity element



**Fig. 1** a XRD patterns of ZnS nanotubes and ZnO nanorods, TEM images of **b** ZnO nanorods **c** ZnS nanotubes and **d** Tauc's plot  $(\alpha h\nu)^2$  versus  $h\nu$  for ZnO nanorods and ZnS nanotubes

was observed in both XRD patterns, which indicates the high quality and purity of the prepared samples.

Figure 1b, c demonstrate the TEM micrographs of ZnO nanorods and ZnS nanotubes, exhibiting the rod-like morphology of the ZnO samples and tubular structure of ZnS samples, i.e., both the sample exhibiting elongated one-dimensional structures distributed throughout the samples. It is found that the average linear dimension of the ZnO nanorods is 150 nm and width is 20 nm (Fig. 1b), whereas the tubular structures of ZnS having average length of 250 nm with 50 nm width (Fig. 1c). The little agglomeration in ZnO nanorods could be due to the polar surface nature of ZnO while ZnS nanotubes are quite uniform in nature and well dispersed.

The UV–Vis absorption spectra of ZnS nanotubes and ZnO nanorods have been used to determine the optical band gaps of prepared nanostructures using well-known Tauc's relation. Figure 1d illustrates the  $(\alpha h\nu)^2$  versus  $h\nu$  plot for both the samples used for estimation of the band gap of prepared nanostructural forms, by extending the linear portion to the energy axis. It is found that the optical band gaps of synthesized samples increase from 3.62 (ZnO) to 4.11 eV (ZnS) with change in phase. A large band shift toward blue edge in case of ZnS with respect to ZnO nanorods also suggests the augmentation in optical band gap from ZnO to ZnS samples. Thus, the qualitative information about the band gap evaluated is well corroborated with the quantitative results. However, the increase in the band gap is a well-known phenomenon of quantum confinement. Since the aspect ratio in case of the nanotubes is maximum because of the availability of the two surfaces. Hence, the effect of the quantum confinement is maximum, and thus, the band gap will be maximum for nanotubular structure [12].

### 3.2 Photoconductivity Measurements

The photoconduction behavior of prepared ZnO nanorods and ZnS nanotubes has been recorded in the presence of UV light of power density of  $\sim 3.3$  mW/cm<sup>2</sup> and wavelength 365 nm for three cycles of ON/OFF to check the repeatability of the samples. The samples are illuminated for 60 s and allowed to recover for 60 s at a fixed applied voltage of 5 V. The transient current flowing through the sample is found to increase with exposure time due to generation of electron/hole pairs in the synthesized materials and their transport to the electrodes. The photoresponse in the present samples is calculated using the relative variation in electric current using the expression [8, 9]

$$\text{Response}(\%) = \frac{I_p - I_d}{I_d} \times 100 \quad (1)$$

where

$I_p$  is the photocurrent



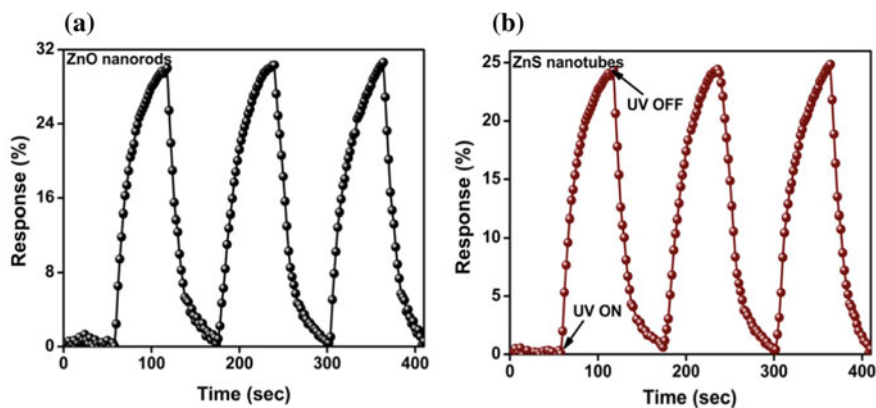
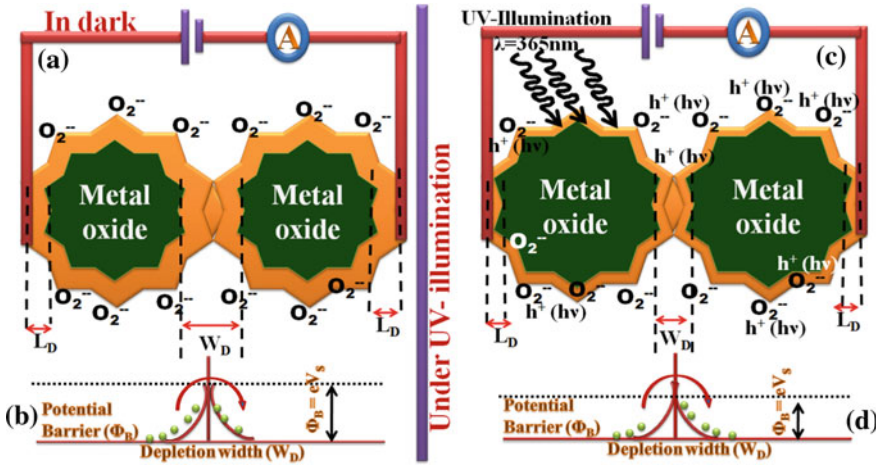


Fig. 2 Transient photoresponse curves for a ZnO nanorods b ZnS nanotubes

$I_d$  is the dark current.

Figure 2a, b show the photoresponses of the prepared ZnO nanorods and ZnS nanotubes with varying time. It is found that the photoresponse of ZnO nanorods sample is greater than the ZnS nanotubes. The transient photoresponse in metal oxides and metal sulfide materials is generally governed by the occurrence of trap states in the band gap region and the level of adsorbed oxygen molecules at surface of the prepared material. The rise in photoresponse is determined the generation of electron-hole pairs and annihilation of these charge carriers at the recombination centers and by the surface accumulation oxygen molecules and finally removal of these molecules from the prepared samples' surface. The photoresponses of the prepared ZnO nanorods and ZnS nanotubes are found to 30% and 25%, respectively.

It is observed that once the UV source is switch off, both the samples regain their initial states of dark currents and are found to recover with in 60 s. The greater photoresponse of the ZnO nanorods can be endorsed to the adsorption of greater number of  $O_2$  molecules on the ZnO nanorods due to its intrinsic character and presence of large number of defects associated with the oxygen and zinc vacancies or the interstitial and substitutional defects as compared to ZnS nanotubes. The mechanism of photodetection in the present nanorods and nanotubular samples can be determined through the accumulation and removal of  $O_2$  molecules on the sample's surface, as follows [8, 9]. Under the dark condition,  $O_2$  molecules are adhered at the surface of ZnO nanorods and ZnS nanotubes and captured the electrons from the interior/bulk of the samples and convert into the oxygen ions ( $O_2(\text{ads}) + e^- \leftrightarrow O_2^-(\text{ads})$ ). The capturing of these electrons from the bulk of the sample resulting into the formation of lower conductive region called the depletion region (Fig. 3a). The presence and the extent of this depletion layer result into the lower electrical conductivity of the sample and thus generally metal oxide and sulfides are of lower conductivity and have high optical band gaps. The formation of this depletion region causes an upward bend bending of valence and conduction bands leading to origination of potential



**Fig. 3** Schematic representation of **a** development of depletion layer at the grains and grain boundaries in ambient environment **b** development of potential barrier at boundaries of grains of height ( $\Phi = eV_s$ ) **c** diminishing of extent of depletion layer on exposure to UV light and **d** lowering of potential barrier height on UV exposure

barrier of height equal to the extent of band bending ( $\Phi = eV_s$ ) as shown in Fig. 3b. Here  $V_s$  is the surface potential.

This potential barrier control the current flowing through the sample as [8, 9]

$$I = I_0 e^{\left(\frac{-eV_s}{kT}\right)} = I_0 e^{\left(\frac{-\Phi_B}{kT}\right)} \tag{2}$$

where

$V_s$  be the potential developed at the grains’ surface due to the presence of depletion layer

$K$  is the Boltzmann’s constant

$T$  is the temperature

On exposure to UV light, the electron-hole pairs are produced at the sample’s surface  $h\nu \rightarrow e^- + h^+$ . The photo-generated holes are interacted with the surface oxygen ions present at sample surface [ $h^+(h\nu) + O_2^- \rightarrow O_2(\text{ads})$ ] and desorbed from the surface of the prepared ZnO nanorods/ZnS tubes. Electrons are released to interior of the sample in this process and cause a decrease in the width of depletion layer at the grains and grains boundaries (Fig. 3c). The decrement in the width of depletion layer results into the diminishing of potential barrier height (Fig. 3d) and correspondingly the increase in electric current or photoresponse, which can be expressed as [8, 9]

$$\text{Response} = e^{\left(\frac{-\Delta\Phi_B}{kT}\right)} = e^{\left(\frac{-\Delta V_s}{kT}\right)} \tag{3}$$

where

$\Delta V_s$  is the variation in surface potential on exposure to the UV light

$\Delta \Phi_B$  is the corresponding change in the potential barrier height.

Furthermore, the moment, the UV source is off, adsorption of  $O_2$  molecules again takes place on the surface of the sample, resulting in the recovery of the samples.

## 4 Conclusion

The structural properties and phase analysis of the samples have been done through X-ray diffraction studies. XRD patterns of both the samples contain well-defined reflection, which suggest the nanocrystalline nature of both the samples. HRTEM studies reveal the formation of nanorod-like structure in case of ZnO, while nanotubular structures are formed in case of ZnS. UV-Vis spectroscopy studies are employed to compute the band gaps of the samples and observed that the band gaps for ZnO nanorods and ZnS nanotubes are 3.62 and 4.11 eV, respectively. The prepared nanostructured materials are utilized for the UV photodetector. It is found that the ZnO nanorods having the greater photoresponse (~30%) as compared to ZnS nanotubes (~25%) due to accumulation of oxygen molecules and presence of large surface and bulk defect states in ZnO nanorods. The accumulation and removal of  $O_2$  molecules on the sample's surface are the expected mechanisms of UV photodetection in the present samples.

## References

1. Wen F, Zhang WQ, Wei GW, Wang Y, Zhang JZ, Zhang MC, Shi LQ (2008) Synthesis of noble metal nanoparticles embedded in the shell layer of core-shell Poly(styrene-co-4-vinylpyridine) microspheres and their application in catalysis. *Chem Mater* 20:2144–2150
2. Berndt I, Popescu C, Wortmann FJ, Richtering W (2006) Mechanics versus thermodynamics: swelling in multiple-temperature-sensitive core-shell microgels. *Angew Chem Int Ed* 45:1081–1085
3. Shao HF, Qian XF, Zhu ZK (2005) The synthesis of ZnS hollow nanospheres with nanoporous shell. *J Solid State Chem* 178:3522–3528
4. Shao HF, Qian XF, Huang BC (2007) Synthesis of single crystal ZnO nanorods and ZnS nanotubes through a simple ultrasonic chemical solution method. *Mater Lett* 61:3639–3643
5. Datta A, Panda SK, Chaudhuri S (2007) Synthesis and optical and electrical properties of CdS/ZnS core/shell nanorods. *J Phys Chem C* 111:17260–17264
6. Alivisatos AP (1996) Perspectives on the physical chemistry of semiconductor nanocrystals. *J Phys Chem* 100:13226–13239
7. Wang D, Chu X, Gong M (2007) Hydrothermal growth of ZnO nanoscrewdrivers and their gas sensing properties. *Nanotechnology* 18:185601
8. Singh K, Berwal N, Rawal I, Dahiya S, Punia R, Dhar R (2018) Determination of valence and conduction band offsets in  $Zn_{0.98}Fe_{0.02}O/ZnO$  hetero-junction thin films grown in oxygen environment by pulsed laser deposition technique: a study of efficient UV photodetectors. *J Alloy Compd* 768:978–990

9. Singh K, Rawal I, Sharma N, Gautam P, Dhar R (2019) Quantum efficient fast UV photodetectors based on nanocrystalline  $Zn_{1-x}P_xO$  ( $x = 0.00, 0.03, 0.07$ ) thin films deposited by pulsed laser deposition technique. *Mater Sci Semicond Process* 95:7–19
10. Fang X, Bando Y, Liao M, Gautam UK, Zhi C, Dierre B, Liu B, Zhai T, Sekiguchi T, Koide Y, Golberg D (2009) Single-crystalline ZnS nanobelts as ultraviolet-light sensors. *Adv Mater* 21:2034–2039
11. Rawal I (2015) Facial synthesis of hexagonal metal oxide nanoparticles for low temperature ammonia gas sensing applications. *RSC Adv* 5:4135–4142
12. Djuricic AB, Choy WCH, Roy VAL, Leung YH, Kwong CY, Cheah KW, GunduRao TK, Chan WK, Lui HF, Surya C (2004) Photoluminescence and electron paramagnetic resonance of ZnO tetrapod structures. *Adv Funct Mater* 14:856–864

# Smart Government E-Services for Indian Railways Using Twitter



Mukta Goyal, Namita Gupta, Ajay Jain and Deepa Kumari

**Abstract** The research area of social media platform for e.g. Twitter is being regarded as most popular social media platform not only to share and disseminate the information but also raising complaints/grievances from citizens. In fact, high velocity, veracity and variety of real time data, it is very hard to analyze the data related to citizen's complaints and problems through manual processing (Agarwal et al. in CoDS-COMAD '18 Proceedings of the ACM India joint international conference on data science and management of data. ACM, pp 67–77, 2018 [1]). So there is a need to filter relevant data with automation which requires some actions by concerned authority as a part of Smart Governance. Smart Governance demands distinguishing the appreciations, praises, issues, problems and grievances posted at social platform by civilians to inform government authorities. It also includes facilitating public agencies to respond to these problems, issues, complaints so that citizen's services can be improved without delay. Thus this paper proposes intelligent techniques to mining public citizens' complaints and grievances from user-generated contents. Since Twitter considered as most popular social media platform which increases the

---

The original version of this chapter was revised: Figure 1 is replaced with a revised figure and the heading of Section 2.2 "Micro Post Enrichment Algorithm" is replaced with "Extraction of Tweets". The correction to this chapter is available at [https://doi.org/10.1007/978-981-15-2329-8\\_74](https://doi.org/10.1007/978-981-15-2329-8_74)

---

M. Goyal

Department of Computer Science and IT, Jaypee Institute of Information Technology, Noida, Uttar Pradesh, India

e-mail: [mukta.goyal@jiit.ac.in](mailto:mukta.goyal@jiit.ac.in)

N. Gupta · D. Kumari

School of Business Studies, Sharda University, Greater Noida 201306, India

e-mail: [go2namita@gmail.com](mailto:go2namita@gmail.com)

D. Kumari

e-mail: [deepakumari@hotmail.com](mailto:deepakumari@hotmail.com)

A. Jain (✉)

Department of Management, SRM Institute of Science and Technology, Modinagar, Uttar Pradesh 201204, India

e-mail: [ajayjain.srm@gmail.com](mailto:ajayjain.srm@gmail.com)

© Springer Nature Singapore Pte Ltd. 2020, corrected publication 2021

D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication*

*Engineering*, Lecture Notes in Networks and Systems 106,

[https://doi.org/10.1007/978-981-15-2329-8\\_73](https://doi.org/10.1007/978-981-15-2329-8_73)

chances of immediate action and fast processing of their request and grievances by the relevant government department or authorities excellent.

**Keywords** E-governance · E-services · Railway complaints · Twitter

## 1 Introduction

Nowadays, it has been observed that social media has become mostly used platforms by civilians as well as government agencies and their usage trend is increasing drastically. These platforms are being used for spreading as well as acquiring information. This is also called Crow sourcing. Due to velocity, veracity and variety of data, it is very impractical to identify complaints and grievances with manual process posted online by public. It is need of time to automate the identification of only relevant information. Thus, government is very keen to have social media platform account which can help to address public issues and grievances posted online using Twitter or Facebook etc.

As twitter and Facebook have wide reach ability across country/world, so each government has planned or planning to have account on such social platform to make direct connection to public/citizens/civilians. Indian Government offices such as Ministry of Railways has twitter account named as (@railminindia), Ministry of road and transport has account named as (@morthindia), traffic police named as (@dtptraffic) and income tax department named as (@incometaxindia). Analysis shows that 50% Government Twitter accounts' tweets posted in an hour are of complaints and grievances reported from various regions of India [2]. The application of applying the intelligence techniques to mine, extract, finding actionable data, information which are useful in sights [3]. Kumar et al. [4] proposed a Twitter application that will help to find out road hazards by applying language models on data of twitter's user online messages. Government has also taken steps to frame policies and branches to handle such issues, problems, grievances, raised by citizens using social media platforms. The objectives of setting up such branches in different department of governments (TwitterSeva, MociSeva, Cybercell, DOTSeva) is to filter issues raised and forward to the concerned department to take action immediately within given time frame. This paper is proposed to automate the inspection of tweets without manual intervention. Twitter has limit of 140 characters so due to short length, gives birth to slang languages and abbreviations. Twitter allows freely typed text and there is no defined structure of language hence the number of spelling and grammatical errors have been unwisely increased. Twitter also supports multilingual text and data. So it is a big challenge in front of developer to identify linguistic terminology of data posted through tweeter account. The further challenge is to identify complaints tweets from other types of tweets.

However, Mining complaints on Indian Railways that contain information about poor service of railways and discomfort to the citizens has become important as the volume of complaints through tweets is very large and impossible to be done by

human inspection. Hence, this paper aims to define and develop a method/procedure that will automate the process of identifying issues raised, complaints and grievances posted using tweeter account, without human intervention and classify those tweets into complaint or non-complaints so that instant action can be taken to redress the grievance. This paper uses Twitters’ REST API to mine tweets from Twitter and made use of the server site scripting language—Python to make requests at Twitter API and results are stored in CSV format that can be easily read by the system. The purpose of this paper is to present a detailed description on mining complaints on the Indian Railways and to classify them into complaints and non-complaints. Further, non-complaints into Appreciation, Informational and Promotional (AISP) tweets using machine learning algorithms and compare different classifiers in order to improve the prediction accuracy.

## 2 Methodology

Architecture shows the Functionality of mentioned components in figure: tweets extraction from public agencies’ account, enrichment and enhancement of raw micro posts (tweets) and learning the features of non-complaint and complaint report tweets [5] Complaints, public issues, problems and grievances are inspected on Twitter’s user’s account forwarded to concerned authorities (Figs. 1 and 2). Figure 1 shows the architecture of Smart E-governance for Railway complaints whereas Fig. 2 shows the complaints in tweet from.

### 2.1 Data Set Collection

There Dataset collected in real-time using Twitter from Indian Government. @RailMinIndia. It is the official account of Ministry of railways in India. Al Twitter

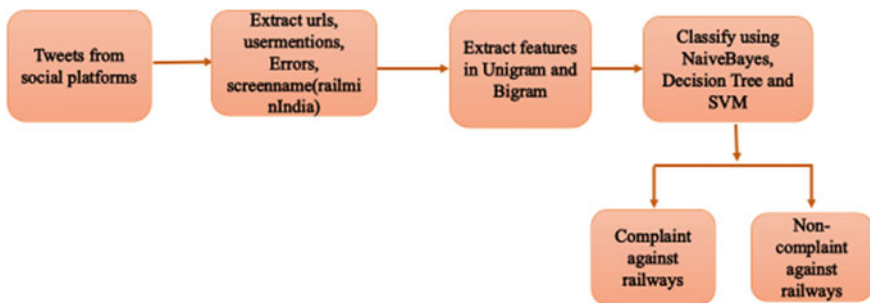


Fig. 1 Architecture for Smart E-governance



Fig. 2 Sample Tweets

REST API is used. Using Twitter API, a collection of total of 16,074 tweets has only 11,292 English tweets. This paper presents this work (Fig. 3).

## 2.2 Extraction of Tweets

Considering only the English tweets, a Micropost Enhancement and Enrichment techniques to clean the tweets and make the tweets apt for applying various classifiers in the future. Following image shows the tokenization of tweets after micro-post algorithms (Fig. 4).

The algorithm followed for Micropost Enhancement and Enrichment is a multistep iterative process that takes a raw tweet as an input and provides a syntactic and semantically enriched tweet. The proposed algorithm primarily consisting of phases: Sentence Segmentation (SSN), Stop Words Removal, Hashtag Expansion (HTE), Padding Space Correction (PSC), Spelling Error Correction (SEC), Acronyms & Slang Treatment (AST), and @Username Mentioned Expansion (UME).

**Sentence Segmentation:** Sentence segmentation involves removing all the URLs and filler terms (such as haha hmm, aah, ohh) posted from tweets [3]. It also considers for replacement of special characters occurring consecutively with one character.

**Stop Words Removal:** Stop words are natural language words which have very little meaning, such as “and”, “the”, “a”, “an”, and similar words.



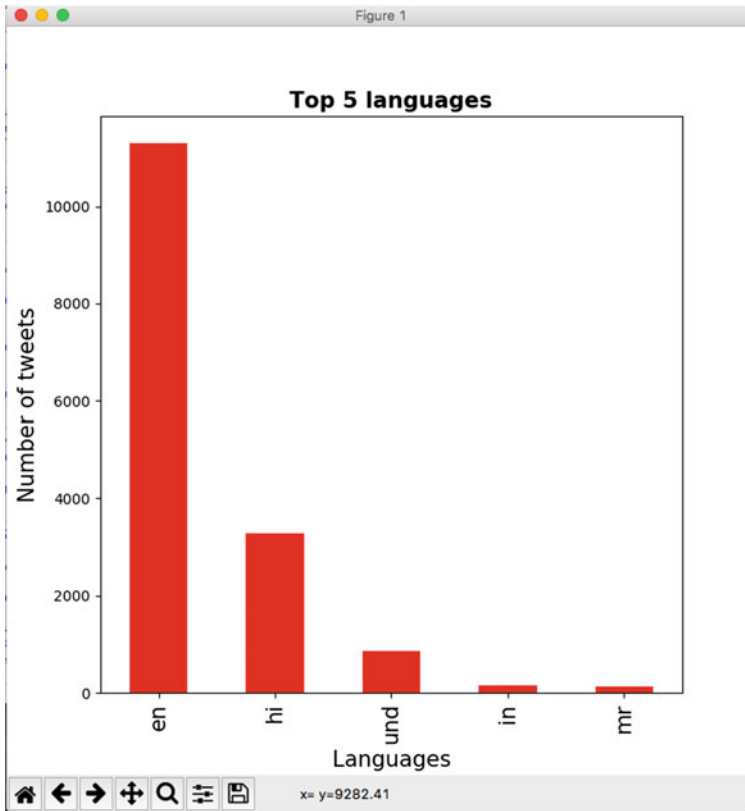


Fig. 3 Statistics between English tweets and other languages

**Hashtag Expansion:** Hashtag expansion is method of splitting it around common separators. The process of splitting a hashtag is done by maintaining sequence of uppercase together till the last upper case in a sentence (lower case is prefixed by acronym).

**Padding Space Correction:** Prefix the whitespace with respect to all special characters such as comma, period, question mark, colon, semicolon, underscore and exclamatory marks.

**Spelling Error Correction:** The spelling error correction is achieved by applying n-gram model.

- **Acronyms and Slang Treatment:** Expansion of web Slangs and normalized contents used in tweets are written in ‘SMS’ linguistic strings in three stages: slang related to domain, standard slang, and user slang [3].
- **Username Expansion:** The expansion of direct mentions is achieved by replacing the Twitter @screenname with the user profile name which helps in making them easily recognizable by the named entity recognizers [3].



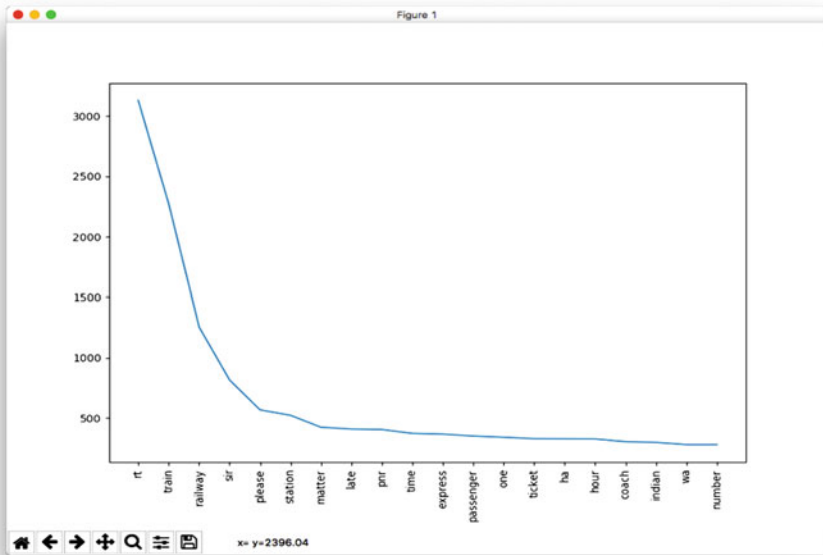


Fig. 5 Frequency distribution of words in sentences

### 2.3.2 Bigrams

Bigrams are word pairs in the data set which occur in succession in the corpus. These features are a good way to model negation in the natural language like in the phrase—This is not good (Fig. 6).

## 2.4 Classifiers

Different classifiers are used to classify the complaints and non-complaints.

### 2.4.1 Naïve Bayes

Naïve Bayes is a simple model which can be used for text classification. In this model, the class  $\hat{c}$  represents tweet  $t$ , where

$$\hat{c} = \arg \max P(c|t)c$$

$$nP(c|t) \propto P(c)P(f_i|c)$$

$$i = 1$$

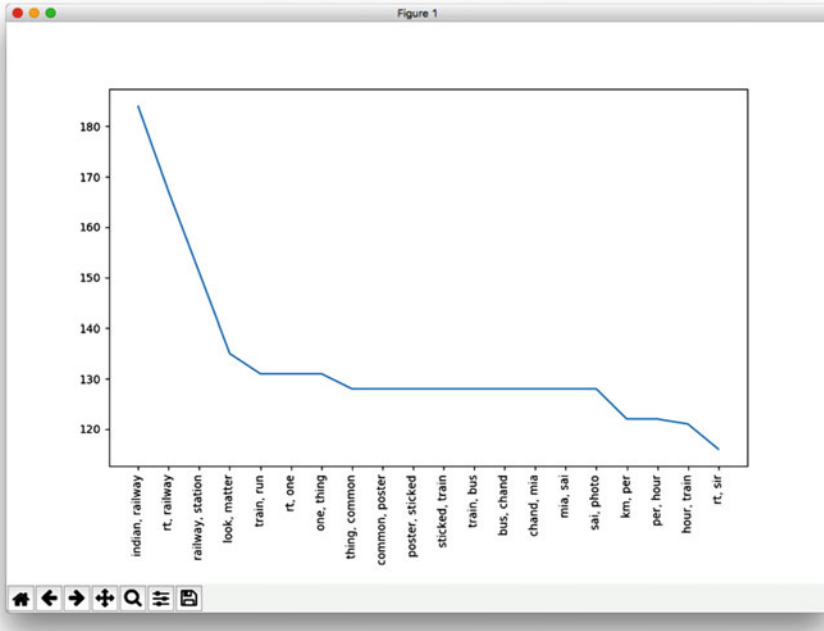


Fig. 6 Bigrams

Formula above,  $f_i$  represents  $i$ th feature from total available  $n$  features.  $P(c)$  and  $P(f_i | c)$  is calculated with respect to maximum likelihood estimates (MLE).

### 2.4.2 Decision Tree

Decision tree classifier models comprised of internal nodes and leaf nodes. The test on data set of attributes is shown as node and their children are outcomes. The final classes of data are shown in leaf node. The model is based on supervised classifier model that forms the tree with known labels and this model applied to test data. The best test condition for each node is taken into consideration [6].

A GINI factor is used to decide the best split. For a given node  $t$ ,  $GINI(t) = 1 - j[p(j|t)]^2$ , where  $p(j|t)$  is the relative frequency of class  $j$  at node  $t$ , and  $GINI_{split} = \sum_{i=1}^k niGINI(i)$  ( $ni$  = count of records at child  $i$ ,  $n$  = count of records on  $i = 1$  to  $n$  node  $p$ ) indicates quality of the split. We choose a split that minimizes the GINI factor.

### 2.4.3 Support Vector Machine

SVM (support vector machines), is a binary linear classifier based on non-probabilistic theory. For a training set of points  $(x_i, y_i)$  where  $x$  is the feature vector and  $y$

is the class, we want to find the maximum-margin hyper plane that divides the points with  $y_i = 1$  and  $y_i = - 1$ . The equation of the hyper plane is as follow

$$w \cdot x - b = 0 \tag{1}$$

To maximize the margin, denoted by  $\gamma$ , as follows

$$\max \gamma, \text{ s.t. } \forall i, \gamma \leq y_i(w \cdot x_i + b) \tag{2}$$

$w, \gamma$  in order to separate the points well.

### 3 Results and Discussion

Comparing the accuracy of Naïve Bayes, Decision Tree and SVM both with unigrams and bigrams, It is found that the best accuracy is achieved using Naïve Bayes which is 83.05%. Decision Tree is 82.94% and SVM is 78.92% accurate. The results showed that SVM has the highest accuracy followed by Naïve Bayes and Decision Tree. The accuracy of the classifiers using a Confusion Matrix are shown in Figs. 7, 8, 9 and 10.

		Predicted		Total
		Complaint	Non-Complaint	
Actual	Complaint	146	57	203
	Non-Complaint	134	560	694
Total		280	617	

Fig. 7 Confusion matrix for decision tree

		Predicted		Total
		Complaint	Non-Complaint	
Actual	Complaint	162	58	220
	Non-Complaint	85	592	677
Total		247	650	

Fig. 8 Confusion matrix for SVM

		Predicted		Total
		Complaint	Non-Complaint	
Actual	Complaint	151	41	192
	Non-Complaint	113	592	705
Total		264	633	

Fig. 9 Confusion matrix for Naïve Bayes

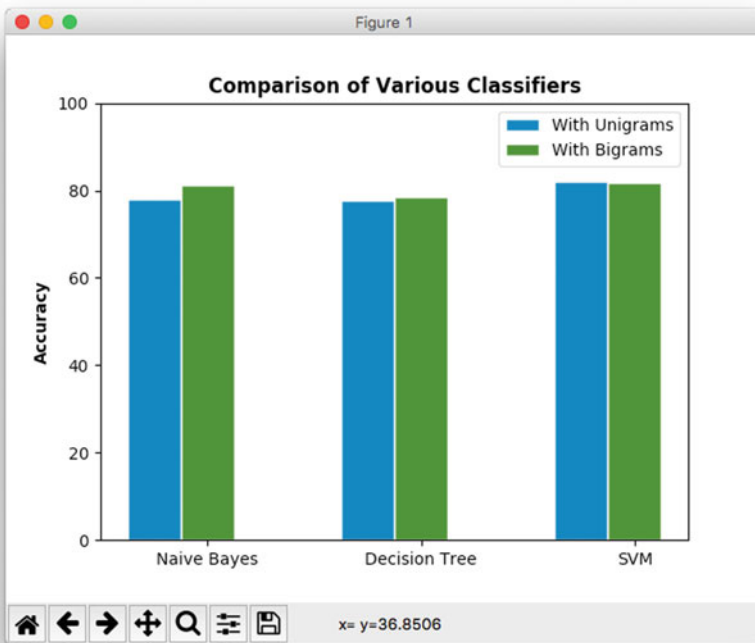


Fig. 10 Comparison of various classifiers

It has been noted that highest accuracy is achieved by Naïve Bayes's algorithm.

## 4 Conclusion

Due to the immense popularity of Twitter, it is seen that civilians are in habit of using Twitter to report their problems, complaints and grievances on various issues actively. However, free text form nature of social media platform and rapid velocity of various types of contents posting, automation of these reports is a technically challenged task. The results show that a significant percentage of complaints posted on Twitter are incomplete and lack the major components in the report making them less likely to get addressed and resolved. Limitation of this approach is that Twitter user can only post 140 characters in one tweet at a time which leads to increase the chances of usage of slang and abbreviations that likely creates grammar and spelling errors. This creates an almost impossible task to automatically classify tweets as complaint and non-complaint tweets. Unavailability of geographical location with the tweet makes it difficult to find the region the complaint has come from. Re-tweets contain identical key that creates difficulty in classification of tweets. Due to the free-form nature of social media text and high velocity of data, automatic identification of these reports is a technically challenging crunch.

## References

1. Agarwal S, Mittal N, Sureka A (2018) Potholes and bad road conditions- mining Twitter to extract information on killer roads. In: CoDS-COMAD '18 Proceedings of the ACM India joint international conference on data science and management of data. ACM, pp 67–77
2. Agarwal S, Sureka A (2017) Investigating the role of Twitter in E-governance by extracting information on citizen complaints and grievances reports. In: Reddy P, Sureka A, Chakravarthy S, Bhalla S (eds) Big data analytics. BDA 2017. Lecture Notes in Computer Science, vol 10721. Springer, Cham
3. Agarwal S, Sureka A, Goyal V (2015) Open source social media analytics for intelligence and security informatics applications. In: International conference on big data analytics. Springer, Cham, pp 21–37
4. Kumar A, Jiang M, Fang Y (2014) Where not to go?: detecting road hazards using Twitter. In: Proceedings of the 37th international ACM SIGIR conference on research & development in information retrieval. ACM, pp 1223–1226
5. Mittal N, Agarwal S, Sureka A (2016) Got a complaint?-keep calm and tweet it!. In: International conference on advanced data mining and applications. Springer, Cham, pp 619–635
6. Benny A, Philip M (2015) Keyword based tweet extraction and detection of related topics. *Procedia Comput Sci* 46:364–371

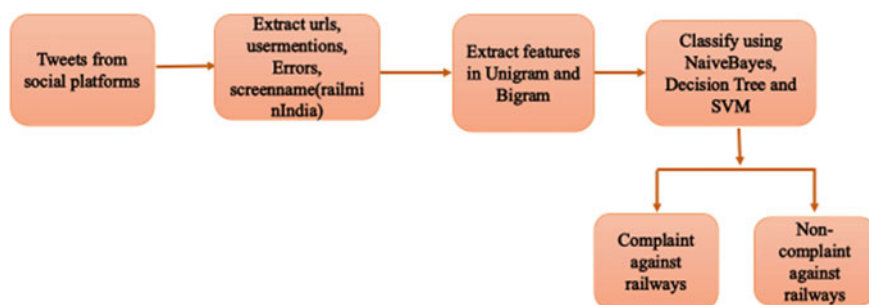
# Correction to: Smart Government E-Services for Indian Railways Using Twitter



Mukta Goyal, Namita Gupta, Ajay Jain and Deepa Kumari

**Correction to:**  
**Chapter “Smart Government E-Services for Indian Railways Using Twitter” in: D. K. Sharma et al. (eds.), *Micro-Electronics and Telecommunication Engineering, Lecture Notes in Networks and Systems 106*, [https://doi.org/10.1007/978-981-15-2329-8\\_73](https://doi.org/10.1007/978-981-15-2329-8_73)**

In the original version of the book, in Chapter 73, Figure 1 is replaced with a revised figure as given below and the heading of Section 2.2 “Micro Post Enrichment Algorithm” is now replaced with the heading “Extraction of Tweets”. The chapter and book have been updated with the change.



**Fig. 1** Architecture for Smart E-governance

The updated version of this chapter can be found at [https://doi.org/10.1007/978-981-15-2329-8\\_73](https://doi.org/10.1007/978-981-15-2329-8_73)

C1



# Author Index

## A

Abbas, Ali, 289  
Afify, Farag M., 669  
Agarwal, Abhishek, 11  
Agarwal, Aditya, 593  
Agarwal, Aparna, 289  
Aggarwal, Garima, 143  
Amannah, Constance Izuchukwu, 459  
Ansari, M. A., 41, 51, 75, 351  
Awasthy, Neeta, 521

## B

Bajaj, Aditya, 365  
Bajaj, Vasvi, 601  
Balyan, Vipin, 1, 255  
Banerjee, Ananyo, 317  
Banerjee, Rahul, 403  
Banerjee, Subhasish, 531  
Barik, Rahul Kumar, 659  
Bekele, Gashaw, 265  
Bharadwaj, Rajath S., 403  
Bhatt, Ishan, 223  
Bohra, Amit, 707  
Buji Babu, P., 439

## C

Chai, Soo See, 117, 125  
Chaudhary, Neelu, 333  
Chaujar, Rishu, 197  
Chourasiya, Shikha, 641  
Chunka, Chukhu, 531  
Cook, Henricus Augustus, 1

## D

Damodaran, D., 307  
Dhall, Ishika, 143, 649  
Dharia, Pankaj, 683  
Dixit, Mrudul, 19  
Dubey, Sanjay Kumar, 387, 395

## G

Ganesh Babu, R., 97  
Garg, Aditya, 365  
Garg, Lalit, 601  
Gautam, Megha, 117, 125  
Gautam, Neha, 117, 125  
Goel, Lipika, 307  
Goel, Nikita, 611  
Goswami, Rajat Subhra, 531  
Goyal, Amit Kumar, 197  
Goyal, Mukta, 721  
Gupta, Bharat, 601, 611  
Gupta, Namita, 611, 721  
Gupta, Neha, 197  
Gupta, Prinima, 153  
Gupta, Saptarshi, 521  
Gupta, Shaloo, 275  
Gupta, Ujjawal, 11

## H

Haripriya, D., 659  
Hasan, Raza, 289

## I

Idris, Mohd Yazid, 447

**J**

Jain, Ajay, 721  
 Jain, Dhruv, 611  
 Joshi, Manoj, 325

**K**

Kahn, Mohamed Tariq Ekeramodien, 1, 255  
 Kakde, Sandeep, 483  
 Kapileswar, N., 659  
 Karan, Aritra, 659  
 Karthika, P., 97, 265  
 Karthik, P. A., 97  
 Kaur, Amanpreet, 299  
 Kaur, Harsupreet, 245, 631  
 Kaur, Komalpreet, 299  
 Khan, Alina, 509  
 Khatri, Sunil Kumar, 307  
 Konda Babu, A., 431  
 Kumar, Ajay, 197  
 Kumar, Hemant, 275  
 Kumari, Deepa, 721  
 Kumar, P., 575, 583  
 Kumar, Shashi Bhushan, 135  
 Kumar, Suresh, 351  
 Kumar, Vinod, 567, 713  
 Kumar, Vipin, 713  
 Kumar, Yogesh, 413

**L**

Lakra, Sachin, 185, 493  
 Lal, Roshan, 365, 499  
 Le, Lu Minh, 205, 213  
 Le, Tien-Thinh, 205, 213  
 Le, Vuong Minh, 205, 213  
 Ly, Hai-Bang, 205, 213

**M**

Maarof, Mohd Aizaini, 447  
 Maheshwari, Kushaagra, 621  
 Maheshwari, Sudhanshu, 621  
 Mahima, 11  
 Mahmood, Salman, 289  
 Mallick, P. S., 317  
 Mani, Bhavika, 275  
 Mani, Prashant, 483, 593  
 Mehta, Hema, 245  
 Mohanty, Sovan, 509  
 Mondal, Bhaskar, 545  
 More, Gayatri R., 65  
 Mukherji, Prachi, 471  
 Muley, Krutika, 19

**N**

Nag, Soumyajit, 531  
 Naidu, Vikas Rao, 289  
 Nandan, Durgesh, 431, 439  
 Nazir, Amril, 447  
 Negi, Poorvika Singh, 499  
 Nhlengethwa, N. L., 583  
 Niekerk van, Bruce Alistair, 255  
 Nirmala, J. S., 403  
 Nlerum, Promise, 459

**O**

Ohatkar, Sharada N., 65

**P**

Palaniappan, Sellappan, 289  
 Pal, Nidhi Singh, 41, 51, 75  
 Panda, Rajat Bhusan, 601  
 Pandey, Priyanka, 631  
 Pandey, Shivam, 351, 659  
 Pannu, Preeti, 341  
 Patel, Harsh, 375  
 Patidar, Yatindra, 11  
 Pawar, Riya, 499  
 Pham, Binh Thai, 205, 213  
 Phani Kumar, P., 659  
 Pilia, Urmila, 153  
 Prasad, Balkrishna, 471  
 Prathap, M., 265

**R**

Rahouma, Kamel H., 669  
 Rai, Garima, 395  
 Rajasree, M. S., 235  
 Raman, Ashish, 413  
 Ramaswamy, Kartik, 223  
 Ramesh, S., 555  
 Ranga, Virender, 421  
 Ranjan, Ashish, 325  
 Ranjan, Ravi, 413  
 Rao, Roshni, 375  
 Rawal, Ishpal, 713  
 Rohini, 41  
 Roy, Manan, 197

**S**

Sabitha, S., 235  
 Sachan, Vibhav Kumar, 707  
 Sagar, Ravi, 75  
 Sai Sravani, E., 431

Sampath, Meghna, 107  
 Sanega, A., 575  
 Saraswat, Shipra, 649  
 Sarin, R. K., 413  
 Sarker, Kamal Uddin, 289  
 Sattar, Mian Usman, 289  
 Sethi, Anita, 31  
 Shah, Manan, 375  
 Shah, Priyam, 375  
 Shaikh, Saniya, 19  
 Shamsudim, Nelufer, 107  
 Sharma, Devendra Kumar, 341, 693  
 Sharma, Kajal, 387  
 Sharma, Mayank, 307  
 Sharma, Ritu, 19  
 Sharma, R. L., 521, 593  
 Sharma, Rohit, 707  
 Sharma, Sudhir Kumar, 175  
 Sharma, Yashika, 185, 493  
 Singhal, P. K., 135  
 Singh, Baldev, 289  
 Singh, Kushall Pal, 11  
 Singh, Mukul, 351  
 Singh, Nivedita, 51, 75  
 Singh, Rupali, 693  
 Sowmya, K., 439  
 Sreehitha, A. V., 431  
 Srivastava, Kavita, 175  
 Srivastava, Sakshi, 107

**T**

Taneja, Nikita, 163  
 Thakare, Rajesh, 483  
 Thakur, Hardeo Kumar, 163, 333

Thirunavukkarasu, Usharani, 85  
 Tiwari, Satyam, 51  
 Tripathi, MM, 197  
 Tripathi, Pragati, 351  
 Tsegaye, Melkamu, 265

**U**

Umopathy, Snehalatha, 85, 107  
 Uma Sathyakam, P., 317

**V**

Vashisht, Vipul, 683  
 Vashisht, Shubham, 143, 649  
 Vedpathak, Madhavi, 471  
 Vidyarthi, Vikas Kumar, 641  
 Vijay, Sandip, 31  
 Vinoth Kumar, M., 567  
 Vydeki, D., 555

**W**

Waseem, Quadri, 447

**Y**

Yadav, Dilip, 51, 75  
 Yadav, Piyush, 621  
 Yadav, Sunil Kumar, 325

**Z**

Zear, Aditi, 421